Understanding coronavirus fusion through structure

Alexandra Corynn Walls

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

David Veesler, Chair

Alexey Merz

W. Conrad Liles

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Understanding coronavirus fusion through structure

Alexandra Corynn Walls

Chair of the Supervisory Committee:

David Veesler

Department of Biochemistry

Two recent coronavirus epidemics highlight the need for vaccine development since no therapeutic currently exists. There are six known human-infecting coronaviruses and many others that can infect livestock and household pets. Viral surveys in bat populations suggest many coronaviruses are poised to cross the species barrier, suggesting future outbreaks are likely. Coronaviruses are enveloped viruses and infection is mediated by a trimeric spike glycoprotein present on the virion surface. The spike proteins are class I fusion proteins, similar to HIV envelope or influenza hemagglutinin, and are the main target of neutralizing antibodies during infection. The viral tropism and ability for the virus to enter cells is determined by the spike protein which undergoes a large conformational change from its pre-fusion state on the viral surface to the post-fusion state following viral and host membrane fusion. The structure and organization of the coronavirus spike protein had not previously been established. In this dissertation, I begin by investigating the architecture of the coronavirus spike protein from various

different genus's and infection hosts. Using state of the art single particle cryo-electron microscopy we were able to solve two near atomic resolution structures of the first β-coronavirus and α-coronavirus spike proteins and elucidate their commonalities and differences. In order to better understand the fusion transition, we also solved the first structure of a β-coronavirus spike protein in the post-fusion state. Moving forward from these snapshots of coronavirus fusion proteins we wanted to start understanding both how fusion is prevented by the human immune system and also how receptor binding can trigger fusion. We solved the structures of two deadly β-coronaviruses in complex with potent, neutralizing human antibodies where we learned the interactions necessary for blocking receptor binding and where we also identified a functional mimicking antibody that triggered protein refolding similarly to the native receptor. Finally, we have also been working to understand how glycan (co-)receptors work with coronaviruses, adding another layer of specificity to the challenges of entry. Through this work and the work done by others in the field, we now have a clear picture of coronavirus spike protein structure and are beginning to uncover the mechanisms of coronavirus entry and membrane fusion.

Table of Contents

Acknowledgements

Writing an acknowledgements section, much like writing a thesis, is a daunting task. It is impossible to thank all of the people who have helped me get to this place in my life personally and professionally. Because of this, I will say a generic and huge thank you to begin my acknowledgements. I will do my best to give gratitude and thanks to the people who have supported me without being asked, laughed with me both when it was easy and hard, taught me things without asking for anything in return, and have been there for me on days I did not know I needed it. From the bottom of my heart, thank you.

I will start by thanking all of my collaborators and silent helper heros. Without all of your had work and dedication none of these projects would have come to fruition! An extra large thank you to Alejandra Tortorici, Xiaoli Xiong, Joost Snijder, Max Sauer, Young-Jun Park, Felix Rey, Joel Quispe, Aria Eshraghi, Joseph Mougous, Brandon Frenz, Frank DiMaio, Irini Topalidou, Wentao Li, and Berend-Jan Bosch for their collaborative spirit. I also want to thank all current, past and present biochemistry graduate students for being a welcoming home for the last five years. I would also like to thank my thesis committee: Justin Kollman, Kelly Lee, W. Conrad Liles, Alex Merz, and Joseph Mougous for their invaluable ideas and questions over the years.

I want to thank the people I surrounded myself with every day for the past five years with the Veesler lab members current and past. Andrew Borst and Young-Jun Park were in the lab with me on day one and without them I would not have survived. Young- being your bay-mate has been a fantastic experience and I will never stop asking you about your weekend plans and experiences! Thank you for allowing us to laugh in lab and sharing your dead-pan jokes with us constantly. Andrew- sharing the experience that was grad school made it a hell of a lot easier. Cookie-o-clock will never die! I have been lucky enough to share a mentor-mentor relationship with many post-doc/staff scientists in the laboratory where I have learned as much from them as they learned from me. Joost Snijder taught me mass-spectrometry, how to both love and respect science, and how to speak Danish. Xiaoli Xiong taught me everything about fusion proteins (mostly HA though…), cell culture, and how to cook Chinese food when extremely hungover. Max Sauer taught me about being careful, raquetball laughs, and putting on a fancy and delicious dinner spread. Alejandra Tortorici taught me how to efficiently think about my project, multi-task like a champion, drop anything to help another person in lab, opened my eyes to my own laughter, and allowed me to sing *at* her incessantly because she knew it brought me joy. Hà Dang: I still cannot believe I convinced you to join this lab, but I am so, so glad I did! You've been my Veesler lab buddy and we've shared art, hiking, dancing classes, cabin weekends, bubble tea, but most of all bread and I could not have done it without your friendship. You were the first rotation student I trained, and I am so excited to see what you achieve. Ollie Acton and Zhaoqian Wang, you both helped keep me sane in the last few months of writing and the existential dread that is graduating!

I need to give an extra special thank you to David Veesler, my boss and mentor for the past five years. He has supported me both when it was easy and when it was difficult. He pushed me to become a better scientist than I could have ever imagined, and I am so grateful to learn directly from him and also alongside him. His first piece of advice to me was to be a used car salesman and be more confident when giving scientific talks

and I have never given a talk without running those words through my head before I begin. I can say *with confidence* that your constant encouragement and suggestions have pushed not only my work and talks but also my self-identity forward in my professional life. David you have become a fantastic boss and also friend and I am so excited to see where you take the lab next. You have given us all such a nurturing place to explore the scientific world and I feel so lucky for that. Thank you for the aforementioned and more, David.

While the lab was a critical component to my graduate school success, there were many people and things outside of the lab that also grounded me and kept me sane when work was trying to do the exact opposite. To my dance friends and everyone who has given me a smile while we dance, I thank you for pushing the creative parts of my brain. To my activity, tea-party, cooking, hiking, and climbing friends: thank you for giving me something concrete to work towards or think about while exploring the world around us. To my friends from high school and college who have supported me near or far: thank you for checking in, taking me to fun places, and being willing to reminisce or do stupid things until the wee hours of the morning. To the ice cream crew who turned into family: thank you for creating the coziest of spaces inside of Frankie and Jos for us all to share and smile together.

Graduate school would not have been the same without my cohort who turned into my second family: Bobby, Domnita, Eleanor, Gulsi, Rachel, Sean, and Sean. I don't come close to having the words to thank you all for being there in every single way. Games, beers, traveling, tears, discussions, hugs, laughter and more. It has been a pleasure to watch you all grow and become even better versions of yourselves than we thought possible. To many more adventures!

My final thank you goes out to my family. Mom, Dad, Danielle, and Jon- you all got me here and saw me through to the end. Seeing you out in the crowd while I defended will remain one of my happiest moments. I am so thankful for all that you have done and continue to do for me, and truly don't know how I became so lucky. I am so happy we all have traveled the world together, enjoy engaging in challenging discussions, and still manage to laugh after all these years. Love you all more than you know.

List of Figures:

Chapter I:

Chapter II:

Chapter V

Chapter I: Introduction

I.I Virology 101

Viruses were identified one century after the first vaccine was developed for small-pox by Edward Jenner in the late 1700's. Edward Jenner and many others found that exposure to a similar, but non-fatal, disease named cowpox was enough to provide protection from the fatal and ghastly small-pox which infected humans. This knowledge spread, but the infectious agent was still not named. This changed just before the turn of the 20th century when Dmitry Ivanovsky discovered a new class of infectious agents that were small enough to be filtered, unlike the long-known bacteria that were visible by a light microscope. This was the Tobacco Mosaic virus which infected tobacco plants. With the invention of the electron microscope in the early 1930's, the first visualization of a virus began, and exploration of virus structure has only sped up since then (Kausche et al. 1939). Early viral classification, distinction, and identification was completed by electron-microscopy until sequencing became the norm, and still has been instrumental in viral identification during new and unknown outbreaks (Chua et al. 2007). Once identified, it was found that viruses contain nucleic acid, either RNA or DNA. The nucleic acid is encapsulated by a shell of protein and sometimes by a lipid bilayer. Viruses are now typically distinguished and segregated into families based on their genome organization. They are obligate parasites, requiring the machinery of a host cell to first be overtaken, used for their own genome reproductive needs, and repeated in subsequent cells and eventually hosts. These parasites are neither alive nor dead but have some of the most striking diversity of any other group of lifeforms due to their typically high mutation rate, short turnover time between entry and production, and a variety of hosts

and infection types. Viruses populate all different shapes, sizes, and environmental niches from humans to *Vibrio cholerae*. Because of the extreme diversity and sheer number of different viruses, they are subject to intense study for a variety of reasons including curiosity, public health, and recently for applications as vast as drug delivery, genome modification, antibiotics, and cancer therapies. The remainder of this document will focus on one family of viruses that until recently had not been highly investigated.

I.II Coronavirus discovery and general architecture

Coronaviruses (CoVs) were first discovered in 1931 with the discovery of a new and deadly respiratory disease in baby chicks by Schalk and Hawn (Schalk et al. 1931). The etiological agent, named infectious bronchitis virus, causing this respiratory disease was first isolated in 1937 by Beaudette and Hudson after it had previously been shown to be a filterable (and therefore non-bacterial) infectious agent by Bushnell and Bradley in 1933. Since then, coronaviruses have been found to infect and circulate in various mammalian and avian species including bats, chickens, pigs, camels, and humans (Licitra et al. 2013; Mole, 2013). They can cause deadly respiratory, neurotropic, and enteric diseases or can be nearly asymptomatic in some hosts. CoVs are enveloped viruses with a large (~30kilobase) single stranded positive sense RNA genome. There are four different genera of coronaviruses, α, β, δ, and γ, separated based on their genome. The large genome encodes for both non-structural and structural proteins including the replicase-transcriptase proteins, envelope protein (E), matrix protein (M), nucleocapsid protein (N), heamglutinin esterase (HE), spike protein (S), and variable accessory and non-structural proteins (Bakkers et al. 2016). The assembled, enveloped virion is between 70-120nm in diameter with M and E situated in the membrane, S protruding outside

creating a ‘crown’ of peplomers, and N wrapped around the genome in the interior of the virion (Neuman et al. 2006). There are approximately 65 S per pleiomorphic virion (Booth et al. 2006). There are four known circulating CoVs that are endemic in humans and cause up to 30% of all mild respiratory infections and in some cases can cause severe disease in infants, the elderly, or the immunocompromised (Hoek et al. 2004; Fouchier et al. 2004). These viruses are HCoV-229E and HCoV-NL63 (α-CoVs), and HCoV-OC43 and HCoV-HKU1 (β-CoVs). CoVs can also have devastating consequences in both livestock and domestic animals, such as with swine and feline CoV’s (Mole, 2013; Licitra et al. 2013; Coleman & Frieman, 2014). Recently, CoVs have fostered a lot of attention due to the emergence of two deadly viruses: severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 which is still an active infection risk. Both SARS-CoV and MERS-CoV cause deadly, atypical pneumonia with fatality rates of ~10 and 35% respectively (Coleman & Frieman, 2014). To date, there are no approved antiviral treatments or vaccines for any human CoVs.

I.III Coronavirus S proteins

The S protein is the major antigen present at the viral surface, and during infection S is the main target of neutralizing antibodies during infection (Zheng, B. et al. 2009). The CoV S protein is a large, homo-trimeric transmembrane protein ranging between 400 and 600kDa with heavy glycosylation (between 20 and 40 N-linked sites predicted per protomer) (Bosch et al. 2003). Other similar proteins include class I fusion proteins such as HIV envelope, influenza hemagglutinin, paramyxo fusion, and Ebola glycoprotein (**Figure I.I**). All of these family members have been (comparatively) intensely studied

biochemically, structurally, and functionally due to their long-known public health risks (Igonet & Rey, 2012; Pancera et al. 2014). Initiation of CoV infection is choreographed by the S protein's ability to bind the host-cell receptor, facilitate membrane fusion, and allow delivery of the viral genome for eventual propagation and subsequent infection of the host cell (Harrison, 2008). On the viral surface, the fusion proteins are poised in their pre-fusion conformation, and upon a trigger, such as low pH or receptor binding, the fusion protein rearranges into the post-fusion conformation and initiates membrane fusion. S belongs to the class I viral fusion protein family, along with HIV envelope, paramyxovirus fusion F protein, and influenza hemagglutinin (Bosch et al. 2003). This class of fusion proteins is characterized by their trimeric nature, compact pre-fusion structures that undergo large scale rearrangements during fusion, and heptad repeat coiled coil post-fusion structures (Harrison, 2008). Two subunits make up the CoV S protein: the receptor binding S_1 subunit and a conserved membrane fusion S_2 subunit. The S_1 subunit comprises four domains named A-D from N-C termini of the domain (Figure I.II). Domains A and B can serve as receptor binding domains (RBDs) and can bind proteins or glycans (Li et al. 2005; Peng et al. 2011; Tortorici et al. 2019). A protease cleavage site is located between the S_1 and S_2 subunit in some CoV S proteins (named the S_2 cleavage site) while all CoV S proteins contain a protease cleavage site located upstream of the putative fusion peptide (named the S_2' cleavage) (Millet & Whittaker, 2015). Cleavage at S_2 occurs during biogenesis and viral assembly in the host cell, at varying locations along the ER to Golgi trafficking pathway depending on the protease utilized. This cleavage is not believed to disrupt the S_1 and S_2 subunit interaction, as they remain non-covalently bound to each other on the viral surface in a meta stable conformation (Burkard et al. 2014). S_2

cleavage may enhance or enable the subsequent S_2' cleavage that occurs preceding membrane fusion. The S_2' cleavage is spatiotemporally controlled by viral tropism and protease location. This cleavage can occur at the host cell surface or during endocytosis (Millet and Whittaker, 2015). The S_1 subunit contains the receptor binding domain (Mou et al. 2013). The S_1 subunit is more diverse amongst CoVs due to the variation in receptor and host cell than the S_2 subunit whose membrane fusion function is conserved. Even the location of the receptor binding domain within the gene varies, from N terminally to centrally located (Peng et al. 2011; Wu et al. 2009). Most human infecting CoVs use transmembrane peptidases as their receptor, while other CoVs can use sugars or sialic acid16. The S_2 subunit contains the fusion machinery, fusion peptide, S_2' cleavage site, heptad repeat regions typical of class I fusion proteins, the transmembrane domain, and a small cytoplasmic moiety. CoVs are zoonotic viruses and both SARS-CoV and MERS-CoV may have originated in bats, with civets serving as an intermediate host for SARS-CoV and camels as a stable source for MERS-CoV (Graham et al. 2013; Sabir et al. 2016). Upon investigation in bats, it was demonstrated that there is a reservoir of CoVs with the potential to cross the species barrier and infect humans (Graham et al. 2013). HKU4-CoV, a bat-infecting CoV, can bind to dipeptidyl peptidase 4 (DPP4), the MERS-CoV human receptor with low affinity. After three amino acid mutations in the S RBD of HKU4-CoV, the HKU4-CoV-RBD to DPP4 affinity is enhanced by two orders of magnitude (Wang et al. 2014). Receptor binding and S processing by host cell proteases are contributing components of viral tropism. Two mutations in the S_2 cleavage site of HKU4-CoV S allowed efficient entry into human cells (Yang et al. 2015). These reports suggest

that the cross-species transmission of CoVs from bats is an imminent and expanding threat to human health.

I.IV Previous Coronavirus Spike Structural Studies

Crystal structures of CoV S postfusion cores and of receptor binding domains in complex with their respective receptor or neutralizing antibodies have been previously reported (accounting for ~25% of the protein) (Gao et al. 2013; Li et al. 2005; Lu et al. 2013; Peng et al. 2011; Supekar et al. 2004; Wu et al. 2009; Xu et al. 2004; Zheng et al. 2006). This accounted for two full domains (A and B) of the S_1 subunit and the heptad repeats present in the S_2 subunit. These did not give any picture to the overall architecture of the spike protein nor the dynamics of the protein. Cryo-electron microscopy of intact SARS-CoV virions also provided low-resolution insights into the shape and coverage of the CoV S protein (Beniac, Andonov, Grudeski, & Booth, 2006). Prior to the work presented in this thesis, no structure of the S protein had been solved, intractable to both crystallography and traditional NMR studies due to its large size, heavy glycosylation, numerous cysteine residues, and metastable fold. Collaborators had tried many times to crystallize the entire S and were unsuccessful (discussions with M. A. Tortorici). The recent advancements in cryo-electron microscopy, the direct detector cameras, and the associated software was one of the many breakthroughs necessary to solve the first CoV S structure to near-atomic resolution.

I.V Advances in Cryo-Electron Microscopy

The Nobel prize has now been awarded many times for electron microscopy: for the invention of the phase contrast method, the electron microscope itself, reconstructing 3-dimensional models from two dimensional images, and for the application of cryo-

electron microscopy (cryo-EM). Cryo-EM has recently undergone a resolution revolution for structural determination of non-crystalline biological samples (Lyumkis, 2019). Proteins are captured in near-native states due to the quick vitrification step which both pauses the motion of the protein and protects the protein from the damage of the electron beam (Cheng et al. 2015). A variety of protein sources, amounts, and purities have now been shown to be amenable to structural determination by CryoEM (Verbeke et al. 2019). Through routine imaging and data processing techniques the field now can routinely achieve 3-5Å resolution with a direct detector camera and either 200 or 300kV instrument with a stable stage such as a Glacios or Krios and a small personal GPU box or cluster with use of freely available softwares. These technical advancements have allowed previously intractable structures by crystallography or NMR to be solved, especially dynamic proteins and membrane proteins, and the field is progressing towards drug design and new innovations such as time resolved structural determination (Kaledhonkar et al. 2019).

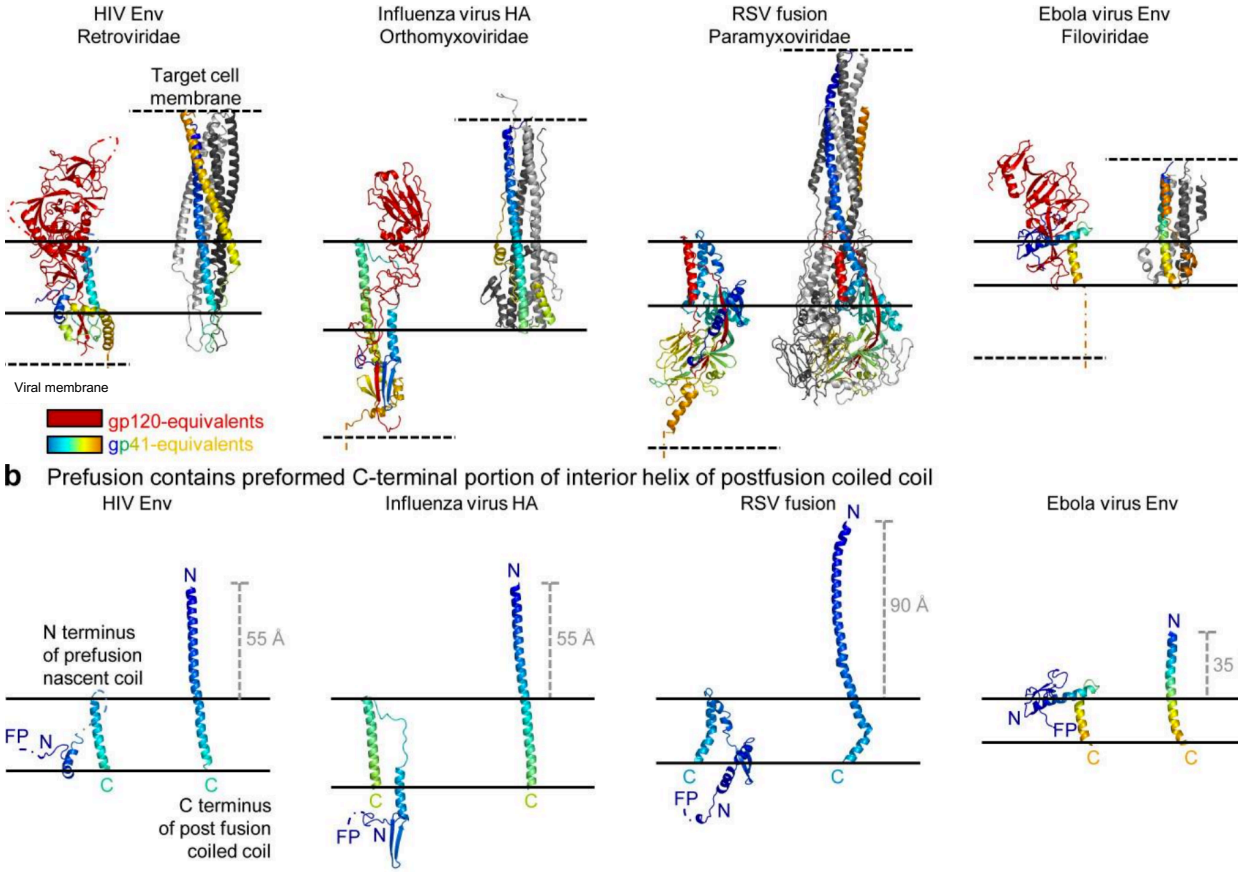
A purified, ideally homogenous and freely tumbling sample is vitrified in a thin layer of non-crystalline ice. This sample is placed one or more times (Snijder et al. 2017) either over a holey carbon or gold support (Russo & Passmore, 2014) or over a layer of thin amorphous carbon (Walls et al. 2019). The frozen grid is then imaged, typically amassing ~1,000 images (Cheng et al. 2015; Lyumkis, 2019). The individual images are a sum of movie frames that are first aligned (Campbell et al. 2012; Li, X. et al. 2013) and then particles are picked from these (Lander et al. 2009). These particles are averaged into 2D, and then aligned into 3D to ultimately determine the final structure using maximum likelihood or stochastic descent methods to properly assign angles and shifts to the

particles (Punjani et al. 2017; Zivanov et al. 2018). A structure is then built and refined using existing modeling tools for X-ray crystallography and some designed specifically for CryoEM methods (Brown et al. 2015; DiMaio et al. 2015; Frenz et al. 2019).

While CoV S proteins had not been previously structurally solved, the precedent had been set by other class I fusion protein structures solved by CryoEM (Lyumkis et al. 2013a). At the time the project was started, proteins of similar size had been structurally characterized (Wu, J. et al. 2015) and the biochemistry of the CoV S protein had previously been worked out for crystallization trials (communication with B.J. Bosch and M.A. Tortorici). This meant that the project was primed for initial structural determination with the quickly maturing field of CryoEM.

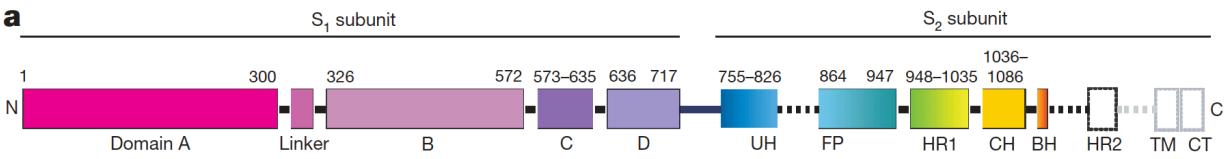
Figures Section I: I.I-I.II

Figure I.I: Prefusion HIV-1 gp120-gp41 structure shares conserved structural and topological features with other type I fusion machines. Adapted from Pancera et al. 2014.



a, Prefusion (left) and postfusion (right) structures. The prefusion structures are shown for a single protomer in ribbon-representation with gp120-equivalent subunits in red, and gp41-equivalent subunits in rainbow (blue to orange). The trimeric postfusion structures are shown with one subunit in rainbow (blue to orange), and the other in light and dark gray. **b**, The C-terminal portion of the preformed interior helix of postfusion coiled-coil from **a** is shown, with fusion peptides (FP) and N and C terminal residues of postfusion coiled-coils labeled, and the distance the inner coiled-coil extends between prefusion and postfusion conformations indicated.

Figure I.II: Organization of the coronavirus spike protein. Adapted from Walls et al. 2016.



a. Schematic diagram of the Mouse Hepatitis Virus 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 (β49–β50); CH, central helix; CT, cytoplasmic tail; FP, fusion peptide; HR1/HR2, heptad-repeats; TM, transmembrane domain; UH, upstream helix.

Chapter II: Structure of Pre-Fusion Coronavirus Spike Proteins

Prior to this dissertations work, there were no structures of coronavirus spike proteins in their pre-fusion conformation. The tremendous pandemic potential of coronaviruses was demonstrated twice recently by two global outbreaks of deadly pneumonia (Drosten et al. 2003; Zaki et al. 2012). The trimeric coronavirus spike (S) glycoprotein contains the receptor-binding domains and the fusion machinery responsible for delivery of the viral genome into the host cell, allowing for subsequent infection. In spite of the threat they represent for human health, the mechanisms used by coronaviruses to initiate infections are poorly understood due to a paucity of structural data (Zheng et al. 2009) .This knowledge gap was hindering the understanding of coronavirus infection, entry, and vaccine design. To solve the identified problem, David Veesler started and I continued the single particle CryoEM project of coronavirus spike proteins once I joined his laboratory, taking advantage of recent technological developments in single particle cryoEM. We attribute this success to three main factors which are the design of a pre-fusion stabilized construct, the strategy employed for cryoEM data collection and processing, and the availability of a recently developed *de novo* model building algorithm using Rosetta (DiMaio et al. 2015; Song et al. 2013; Wang et al. 2015). In this section, I will discuss the structure of mouse hepatitis virus spike (MHV S), the most biochemically studied β-coronavirus, and HCoV-NL63 is an α-coronavirus that can cause severe lower respiratory tract infections requiring hospitalization. The human coronavirus NL63 (HCoV-NL63) is an α-coronavirus, which is genetically distinct from β-coronaviruses mouse hepatitis virus (MHV, the prototypical coronavirus), MERS-CoV and SARS-CoV, and was first isolated from a 7-month-old patient with a respiratory

tract infection (Fouchier et al. 2004; Hoek et al. 2004). Further studies revealed HCoV-NL63 infections appear to be common in childhood, and most adult sera contain antibodies that neutralize the virus (Hoek et al. 2004; Hofmann et al. 2005). HCoV-NL63 is a major cause of bronchiolitis and pneumonia in newborns worldwide and can cause severe lower respiratory tract infections requiring hospitalization, especially of young children, the elderly, and immunocompromised adults (Chiu et al. 2005). HCoV-NL63 infections have been reported in 12 countries across Europe, Asia and North America, indicating that it is circulating among the human population worldwide. Other α-coronaviruses related to the human respiratory pathogen HCoV-229E have recently been identified in camels co-infected with MERS-CoV (Sabir et al. 2016), further underscoring the importance of characterizing this coronavirus genus. Additionally, the emergence of the highly lethal porcine epidemic diarrhea coronavirus (PEDV, α-genus) had devastating consequences for the US swine industry recently (Mole, 2013) .

Adapted from:

Walls, A.C.*, Tortorici, M.A.*, Bosch, B.J.*, Frenz, B., Rottier, P.J.M., DiMaio, F. Rey, F.A., Veesler, D. Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. Protein Science. 2016.

Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., DiMaio, F., Bosch, B.J., Veesler, D. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nature Structural Molecular Biology. 2016.

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

II.I Characterization of Mouse Hepatitis Virus Spike Protein

We produced an MHV S (Walls et al. 2017a; Walls et al. 2016) ectodomain trimer with enhanced stability by mutating the S_2 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 (**Figure II.I**). We used state-of-the art CryoEM (Cheng, 2015) to determine the structure of the MHV S ectodomain trimer at 4.0 Å resolution (**Figure II.II, II.III, II.IV**). We fit the crystal structures of two S_1 domains (Peng et al. 2011; Peng et al. 2012) and built de novo the rest of the polypeptide chain using Coot (Brown et al. 2015) and Rosetta (DiMaio et al. 2015; Wang et al. 2015) (**Figure II.II, II.V**). The final model includes residues 15 to 1118, with an internal break corresponding to a loop immediately upstream from the S_2' cleavage site (residues 827–863). The region connecting the S_1 and S_2 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.

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. The structure comprises two functional subunits: a distal moiety constituted by the S_1 subunits; and a central stem connecting to the viral membrane formed by the S_2 subunits (**Figure I.II, II.VI**). The S_1 subunit has a 'V' shape contributing to the overall triangular appearance of the S trimer (**Figure II.VII**). The S_1 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 S_1 C-terminal half folds as three spatially distinct β-rich domains, termed B, C and D (**Figure II.VI**).

The S_2 subunit connects to the viral membrane and is characterized by the presence of long α-helices (**Figure II.VI, Figure II.VIII**). A central helix (α30) stretches 75 Å along the three-fold molecular axis towards the viral membrane (**Figure II.VIII**). It is located immediately downstream of the HR1 motif, which folds as four consecutive α-helices (α26–α29; **Figure II.VIII, II.IX**), in sharp contrast to the 120-Å long HR1 helix observed in the post-fusion S structures (Duquerroy et al. 2005; Gao et al. 2013; Xu et al. 2004) (**Figure II.IX**). The 55-Å-long upstream helix (α20), so named because it is located immediately upstream of the S_2' cleavage site, runs parallel to and is zipped against the central helix via hydrophobic contacts largely following a heptad-repeat pattern. A core antiparallel β-sheet (β46–β49–β50) is present at the viral membrane proximal end and is assembled from an N-terminal β-strand (β46), preceding the upstream helix, and a C-terminal β-hairpin (β49–β50), 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 (**Figure II.VIII**). Importantly, these motifs were shown to remain invariant in the pre- and post-fusion F structures (McLellan et al. 2006). 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 β50, 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 (**Figure II.X**). 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 (**Figure II.VIII**). 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 $Å^2$). Many of these contacts involve the HR1 helices and the fusion peptide region. These polypeptide segments undergo major refolding during the fusogenic conformational changes (**Figure II.IX**), 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 α28 or of the conserved leucine 1062 by phenylalanine in the central helix were shown to attenuate fusogenicity (Krueger et al. 2001; Taguchi & Matsuyama, 2002). Our structure suggests that the former substitution would strengthen hydrophobic packing against the core β-sheet **(Figure II.X**), and that the later substitution could reinforce molecular stapling of the central helices (**Figure II.X**). The expected modification of the energy landscape between pre-fusion and post-fusion conformations would explain the reduction in fusion activity of these mutants21,22.

The predicted fusion peptide includes the C-terminal half of helix α21 and extends up to the N-terminal half of α22 (refs 6 and 23) (**Figure II.III**). α21 is an amphipathic helix located at the periphery of the S trimer, burying hydrophobic side chains towards the S_2 centre and exposing charged residues to solvent (**Figure II.III, II.X**). In the case of porcine

epidemic diarrhoea coronavirus, trypsin processing at the S_2' site can only occur after host cell attachment (Wicht et al. 2014). This indicates that receptor binding could allosterically increase the accessibility of the S_2' site, which is located within helix α21. The acidic pH of the endolysosomes could also contribute to exposing the S_2' cleavage site for coronaviruses requiring cleavage in this compartment. The fact that helix α21 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 S_2' site, as it frees the fusion peptide from the S_2 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 prefusion structures. The notable accessibility of the fusion peptide and its sequence conservation among coronaviruses (Burkard et al. 2013; Madu et al. 2009) suggest that it would be an ideal target for epitope-focused vaccinology initiatives aimed at raising broadly neutralizing antibodies against S glycoproteins (**Figure II.XI**). Major 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 (Daniel et al. 1993; Zhang et al. 2004). 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 F[28]. 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 (Peng et al. 2011), whereas MERS-CoV and SARS-CoV rely on the β-motif protruding from domain B to bind to DPP4 (Lu et al. 2013) or ACE2 (Li et al. 2005; Wu et al. 2009), respectively (**Figure II.VII**). The poor sequence conservation of the B domain β-motif among coronavirus S proteins, its considerable length variation among MHV strains 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 S_1 and S_2 subunits respectively share ~25% and ~40% sequence similarity with many other coronavirus S proteins. 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 MHV (Peng et al. 2011) and bovine coronavirus (Peng et al. 2012) A domains; (2) the MHV, MERS-CoV (Lu et al. 2013), SARS-CoV (Li et al. 2005) and HKU4 (Wang et al. 2014) B domains; (3) the post-fusion cores of MHV (Xu et al. 2004), SARS-CoV (Duquerroy et al. 2005; Supekar et al. 2004), and MERS-CoV (Gao et al. 2013) ; and (4) the isolation of infectious coronaviruses featuring a deletion of the A domain and using domain B as the receptor-binding domain (Reguera et al. 2012). Our results now provide a framework to understand coronavirus entry and suggest ways for preventing or treating future coronavirus outbreaks.

II.II Methods for characterization of Mouse Hepatitis Virus Spike Protein

a) Construct Design:

Viral fusion proteins adopt a metastable pre-fusion conformation at the virus surface until triggered to rearrange into a more stable postfusion conformation which promotes merger of viral and host membranes (Harrison, 2008). The significant magnitude of the conformational changes taking place during the fusion reaction could result in masking of epitopes initially accessible in the prefusion state and exposure of new epitopes specific to the post-fusion state. As a result, vaccine design initiatives aim at targeting the pre-fusion state of viral fusion proteins, which correspond to the conformation that could be detected by the immune system before infection. The intrinsic metastability of viral fusion proteins is usually associated with challenges to preserve the pre-fusion state during purification. This is illustrated by the case of the respiratory syncytial virus (paramyxovirus) F protein which required co-expression of the ectodomain (fused to a T4 foldon motif) with a pre-fusion specific Fab to enable isolation of this conformation (McLellan et al. 2011; McLellan et al. 2013a; Swanson et al. 2011).

During biogenesis, the MHV S protein is often naturally cleaved at the S_1-S_2 junction (S_2 cleavage site) by Golgi-resident furin(-like) proteases (Bosch et al. 2003; Burkard, et al. 2014) resulting in an increase in its fusogenic propensity. After cleavage, the S_1 and S_2 subunits remain non-covalently associated in the metastable pre-fusion S trimer. In the case of SARS-CoV and MERS-CoV, S_2 processing has also been suggested to promote subsequent cleavage at a second site located just upstream of the fusion peptide (S_2' cleavage site) to allow the fusion reaction to proceed upon virion uptake by a target cell (Millet & Whittaker, 2014; Millet & Whittaker, 2015). We engineered a

construct featuring a single amino acid substitution in the S_2 cleavage site to prevent furin processing and enhance the stability of the MHV S ectodomain pre-fusion structure. Substitution of an arginine residue present at position 717 by a serine residue at the site of cleavage (from RAHR↓ to RAHS) resulted in the purification of a homogeneous uncleaved protein product as confirmed by SDS-PAGE analysis.

Although MHV S is known to oligomerize into homo-trimers upon translation *in vivo*, expression of the ectodomain yielded predominantly monomers, indicating that the transmembrane domain is required for trimerization and/or trimer stabilization. To promote oligomerization, an engineered trimerization motif based on the transcription factor GCN4 (Eckert et al. 1998; Yin et al. 2006) was C-terminally fused to the MHV S ectodomain in frame with the heptad repeat 2 (HR2) motif helix. Biophysical analyses using analytical size exclusion chromatography coupled online to multi-angle light scattering (Veesler et al. 2009) (SEC-MALS) as well as native mass spectrometry confirmed the trimeric organization of the GCN4 stabilized MHV S ectodomain. Proper folding of the purified MHV S ectodomain was confirmed by analyzing its binding affinity to the CEACAM1a ectodomain (the viral receptor) using microscale thermophoresis. We determined a dissociation equilibrium constant of 48.5 ± 3.8 nM which is in good agreement with the reported value of 21.4 ± 4.2 nM (Peng et al. 2011) for the isolated receptor-binding domain. Imaging of this sample using negative staining EM further confirmed the homogeneity of the purified protein and suitability for high-resolution studies.

Technical Outline:

A human codon-optimized gene encoding the MHV spike gene (UniProt: P11224) was synthesized with a R717S amino acid mutation to abolish the furin cleavage site at the S_1-S_2 junction (S_2 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 (IKRMKQIEDKIEEIESKQKKIENEIARIKKIK)(Eckert et al. 1998; Yin et al. 2006), 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 murine CEACAM1a (residues 35-142; gb NP_001034274.1) was amplified by PCR and cloned into a mammalian expression plasmid (Zeng et al. 2008), 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.

293-F cells were grown in suspension using FreeStyle™ 293 Expression Medium (Life technologies) at 37 °C in a humidified 5% CO_2 incubator on a Celltron shaker platform (Infors HT) rotating at 130 rpm (for 1 L culture flasks). Twenty-four hours before transfection, cell density was adjusted at 1.5×10^6 cells/ml, and culture grown overnight in the same conditions as mentioned above to reach ~ 2.5×10^6 cells/ml the day of transfection. Cells were harvested by centrifugation at 1250 rpm 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.

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 5 min on shaker in the culture incubator before adding 9 µg/ml of Linear polyethylenimine (PEI) solution (25 kDa, Polysciences). Twenty-four hours post-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 10-fold using Vivaflow tangential filtration cassettes (Sartorius, 10 kDa cutoff) 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 $CaCl_2$. The reaction mixture was incubated at 25°C overnight and re-loaded in a Protein A column to remove uncleaved protein and the F_C 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.

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 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 of $CdCl_2$ at a density of approximately 10^7 cells per mL. After a week at 28 °C clarified cell supernatants were concentrated 40-fold using Vivaflow tangential filtration cassettes (Sartorius, 10 kDa cutoff) 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.

b) Biophysics and CryoEM Data Collection and Processing

Ice thickness has a strong influence on the final achievable resolution of single particle reconstructions. Ideally, the vitreous ice should be as thin as possible to still accommodate the particles of interest while maximizing Thon ring intensity at high spatial frequencies (Wu et al. 2016). Imaging was completed on a Titan Krios 300 kV microscope equipped with a Gatan K2 Summit direct electron detector operated in counting mode (Li et al. 2013). Similarly to our previous work on the *Thermoplasma acidophilum* 20S proteasome, we initially sought to acquire data from holes having the thinnest possible vitreous ice (Campbell et al. 2012). However, the MHV S protein clearly showed signs of denaturation when images where acquired in such conditions. We interpret this observation as resulting from the surface tension exerted on the S trimers in thin vitreous ice. Hence, we targeted holes with slightly thicker ice than desired in which we could

observe compact well-folded MHV S trimers, similar to what was observed using negative staining EM. We collected a large dataset (1,600 micrographs) at high defocus (2.0–5.0 μm) to maximize the low-resolution contrast and our ability to align the particle images during subsequent processing. This example illustrates that although it is not always possible to acquire data in the thinnest possible areas of ice (especially for fragile protein complexes), near-atomic resolution reconstructions can still be obtained by tailoring the imaging conditions appropriately.

One of the major challenges encountered during processing of cryoEM data is the presence of multiple 3D structures in a given dataset. These differences can result from different conformations of the same protein, different chemical compositions due to loss of one or several subunits of a protein complex, or (partial) denaturation of a fraction of the particles during purification or vitrification. If left untreated, this heterogeneity can limit the resolution and compromise the quality of the final map. 3D classification has emerged as an extraordinarily powerful tool to deal with structural heterogeneity in allowing to computationally isolate homogeneous subsets of the data (Bai et al. 2015; Lyumkis et al. 2013; Scheres et al. 2007).

We relied on extensive 2D and 3D classification using the Relion software (Scheres et al. 2012a; Scheres et al. 2012b) to deal with the marked structural heterogeneity of the MHV S ectodomain trimer dataset. We ran a first round of 3D classification without imposing symmetry to improve separation of “good” and “compromised” particle images. Although looking at the classified maps did reveal differences between the different classes, looking at the aforementioned slices further confirmed the structural heterogeneity present in this dataset at a glance, as previously suggested by Scheres et

al. (Bai et al. 2013) At the resolution of our analysis, we could not identify distinct conformations of the MHV S trimer and postulated that the particles contributing to less-well resolved classes could be partially denatured.

Comparison of the results of projection-matching refinements (using C3 symmetry) run before and after the aforementioned 3D classification step suggested that both reconstructions had similar resolution (4.4 Å) according to the gold standard Fourier shell correlation ($FSC_{0.143}$) criterion. The quality of the two maps, however, differed significantly as only the reconstruction computed after 3D classification showed features compatible with the resolution estimate. This case study highlights that gold standard FSC measures internal consistency between two halves of the data (Egelman, 2014; Subramaniam et al. 2016) not resolution, and that the quality of the final map should always be in agreement with any numerical estimates of resolution. Starting from 1,200,000 particle images, we significantly reduced the size of the data set to 82,000 particles using 2D and 3D classification to generate the final 3D reconstruction at 4Å resolution showing well resolved α-helices, β-strands and amino acid side chains for a large part of the map.

Technical Outline:

For size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis, samples (0.2 mL at 1mg/mL) were loaded onto a Superdex 200 10/300 GL column (GE Life Sciences, 0.4 mL/min 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 analyzed using Astra 6 software (Wyatt Technology Corp., CA, USA).

Solution MST binding studies were performed using standard protocols on a Monolith NT.115 (Nanotemper Technologies). Briefly, recombinant CEACAM1a ectodomain protein was labeled 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 labeled 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 1000 units at 40% LED power. The samples were loaded into standard-treated Monolith TM capillaries and were measured by standard protocols using a Monolith NT.115, NanoTemper (Munich, Germany). The changes of 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.

Three microliters of MHV spike at 1.85 mg/mL were applied to a 1.2/1.3 C-flat grid (Protochips), which had been glow-discharged for 30s at 20mA. Thereafter, grids were plunge-frozen in liquid ethane using a Gatan CP3 and a blotting time of 3.5 s. Data was 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 software (Glaeser et al. 2011). Automated data collection was carried out using Leginon (Suloway et al. 2005) 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/physical pixel/s which corresponds to ~12 electrons/physical pixels/s (when accounting for coincidence loss (Li et al. 2013)). Each movie had a total accumulated

exposure of 53 e/Å^2 fractionated in 38 frames of 200 ms (yielding movies of 7.6 s). A dataset of ~1600 micrographs was acquired in a single session using a defocus range comprised between 2.0 and 5.0 µm.

Whole frame alignment was carried out using the software developed by Li *et al* (Li et al. 2013), which is integrated into the Appion pipeline (Lander et al. 2009), to account for stage drift and beam-induced motion. The parameters of the microscope contrast transfer function were estimated for each micrograph using ctffind3(Mindell & Grigorieff, 2003). 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 DogPicker (Voss et al. 2009). Extraction of particle images was performed using Relion 1.4 with a box size of 320 pixels^2 and applying a windowing operation in Fourier space to yield a final box size of 288 pixels^2 (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 RELION (Scheres, 2012). An initial 3D model was generated using OPTIMOD (Lyumkis et al. 2013b) 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 (~500,000 particles). These ~500,000 particles were then subjected to RELION 3D classification with 4 classes (using c1 symmetry) starting with our initial model low-pass filtered to 40Å resolution. We subsequently used the ~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 utilized the particle polishing procedure in RELION 1.4 to correct for

individual particle movement and radiation damage (Bai et al. 2013; Scheres et al. 2014). 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 $Å^2$ using Relion post processing. Reported resolutions are based on the gold-standard FSC=0.143 criterion (Scheres & Chen, 2012) and Fourier shell correction curves were corrected for the effects of soft masking by high-resolution noise substitution (Chen et al. 2013). 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 resolved (Beniac et al. 2006).

c) Model Building

Obtaining an atomic model of the MHV S glycoprotein required a hybrid approach combining docking of available crystal structures, *de novo* modeling using Rosetta (DiMaio et al. 2011; DiMaio et al. 2013; DiMaio et al. 2015; Wang et al. 2015) and Coot (Brown, et al. 2015; Emsley et al. 2010) and density-guided homology modeling using RosettaCM (Song et al. 2013). The C-terminal S_2 subunit, which is the fusion machinery, is best defined in the density and was built using a combination of hand tracing with Coot and Rosetta *de novo* building (Wang et al. 2015). The observation of large, bulky side chain densities, several disulfide bonds resolved in the map and of density putatively

corresponding to glycans for several asparagine residues belonging to N-glycosylation sequons were used as internal controls during model building.

The density corresponding to the N-terminal receptor-binding S_1 subunit is not as well resolved as for the fusion machinery and features various levels of resolution in the reconstruction. The availability of two crystal structures for domain A (Peng et al. 2011; Peng et al. 2012) (including a structure of the MHV domain A) and of several crystal structures for domain B (Li et al. 2005; Lu et al. 2013) was of tremendous assistance and allowed us to directly dock these models into the reconstruction. RosettaCM was then used to rebuild the core β-sheet of domain B and to derive a putative model (using density-guided homology modeling) for the disordered extension corresponding to the receptor-binding motifs in MERS-CoV and SARS-CoV. The quality of the map corresponding to domains C and D hampered manual sequence assignment for this region of the protein. Rosetta *de novo* (Wang et al. 2015) successfully identified a ~30 residue-long fragment allowing to anchor the sequence register for domains C and D. The placement of several bulky side chains accounted for by the density and the identification of putative N-linked glycans suggested correct assignment, and allowed completion of the model. The density for the linker connecting the S_1 and S_2 subunits is poorly resolved and Rosetta *de novo* was used to generate a putative model of this region of the protein which should be analyzed cautiously, as suggested by the high B-factors associated with it.

Technical Outline:

Fitting of atomic models into cryoEM maps was performed using UCSF Chimera (Goddard et al. 2007) and Coot (Brown et al. 2015; Emsley et al. 2010). We initially

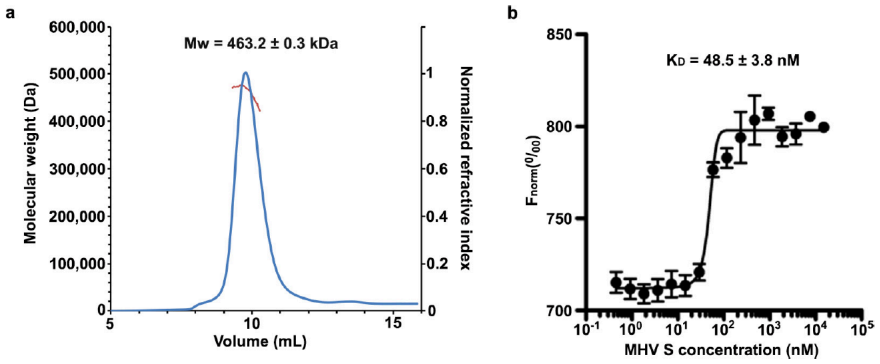
docked the MHV domain A structure (PDB 3R4D) and used a crystal structure of a bovine coronavirus domain A (PDB 4H14) to model the three-stranded β-sheet and the α-helix present on the viral membrane proximal side of the galectin-like domain. Next, the MERS-CoV domain B crystal structure (PDB 4KQZ) was also fit into the density and rebuilt and refined using RosettaCM (Song et al. 2013). 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. Then, 1000 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. Due 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-modeling 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 (*Wang et al. 2015). 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, RosettaCM (Song et al. 2013) 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 Rosetta (DiMaio et al. 2015). 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 overfitting (DiMaio et al. 2013). The quality of the final model was analyzed with Molprobity (Chen et al. 2010). Structure analysis was assisted by the PISA (Krissinel & Henrick, 2007) and DALI (Holm & Rosenström, 2010) servers. The sequence alignment was generated using MultAlin (Corpet, 1988) and colored with ESPript (Robert & Gouet, 2014). All figures were generated with UCSF Chimera (Goddard et al. 2007) .

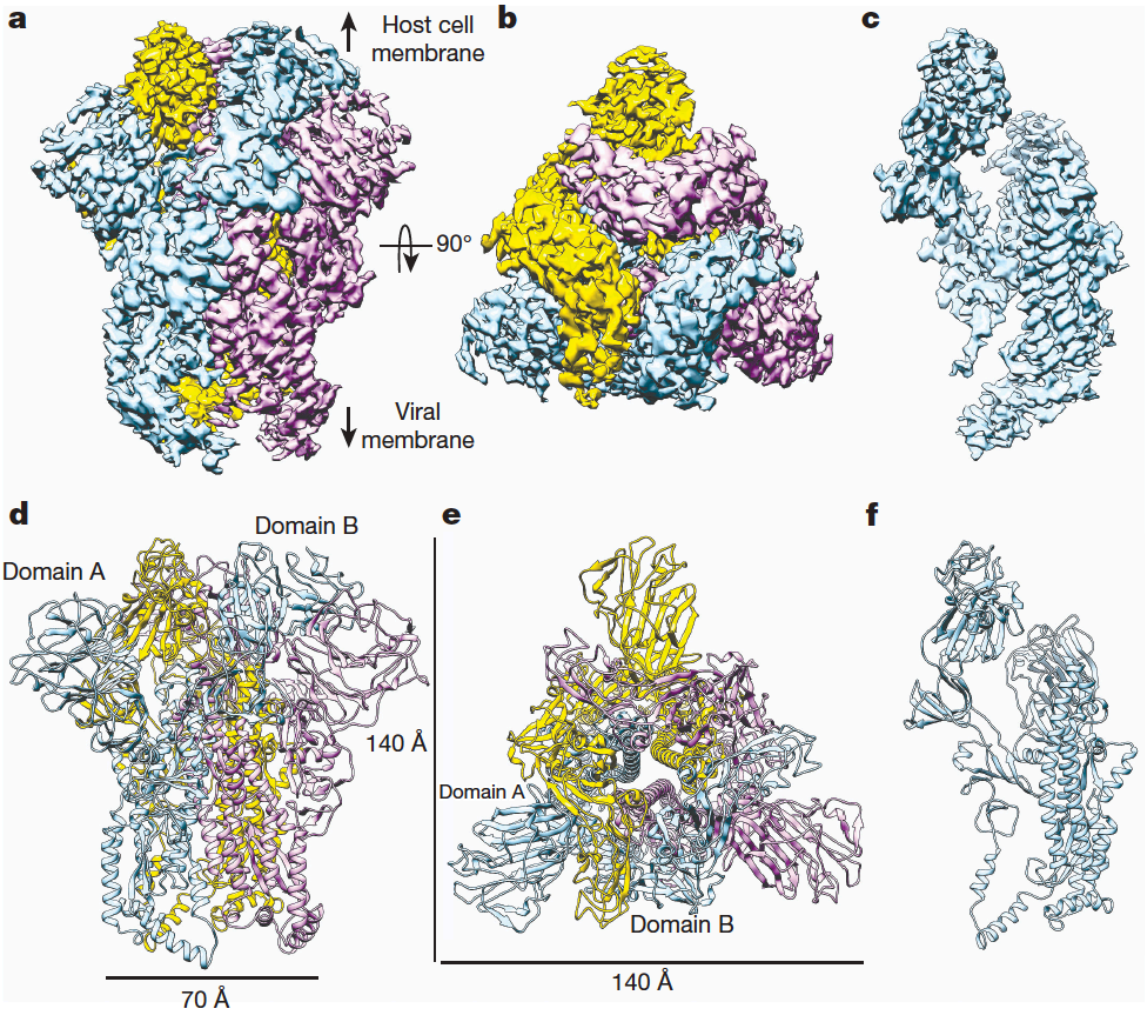
Figures Chapter II II.I-II.XI: Adapted from Walls et al. 2016

Figure II.I: Biophysical characterization of the MHV S ectodomain.



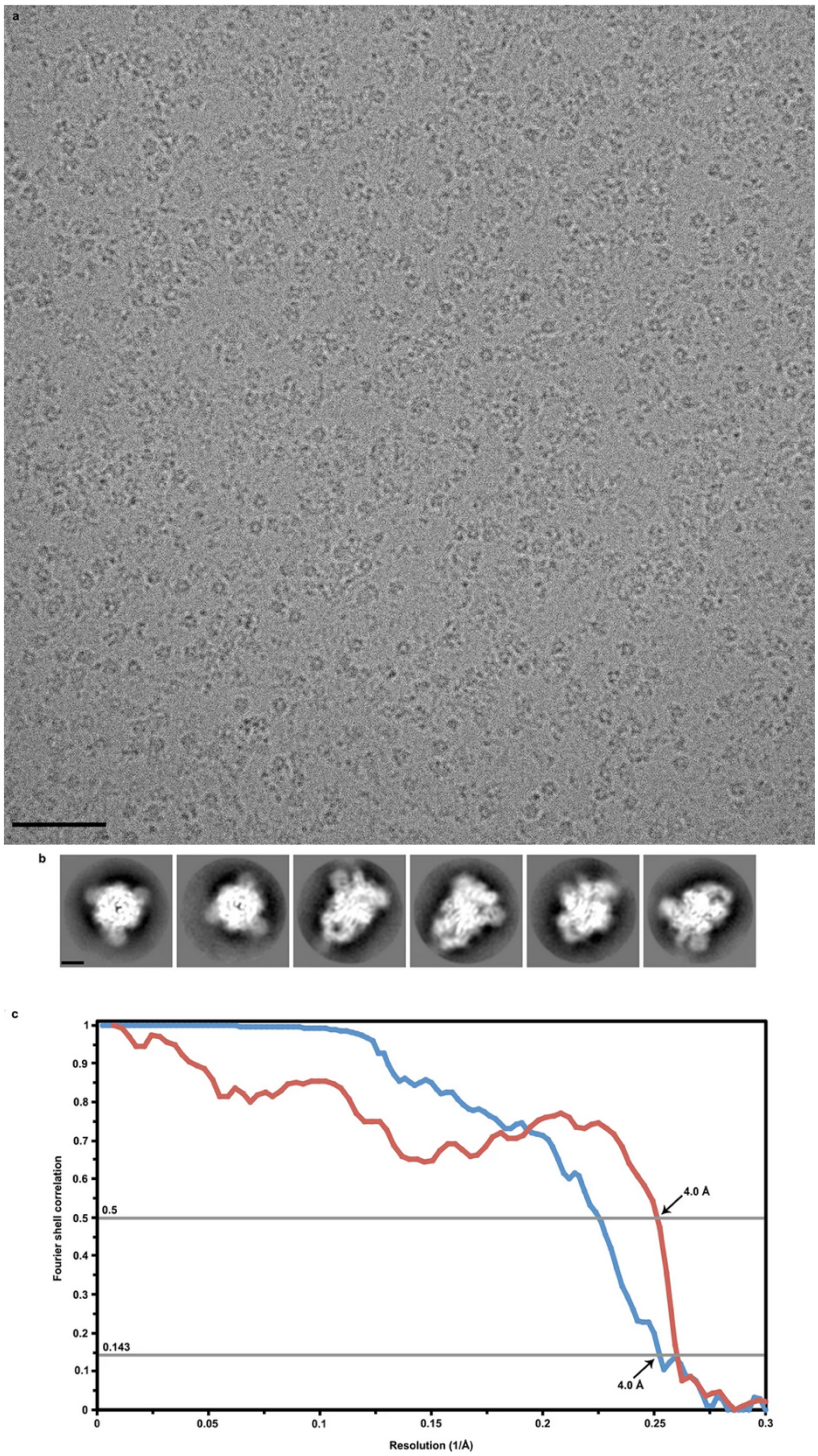
a, The MHV S molecular mass was determined to be 463.2 +/- 0.3 kDa (mean +/ s.e.m.) (corresponding to a trimer) using size-exclusion chromatography coupled in-line with multi-angle light scattering and refractometry. The blue line represents the normalized refractive index (right ordinate axis) and the 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 (Kd) 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.

Figure II.II: 3D reconstruction of the MHV S trimer determined by single-particle cryoEM.



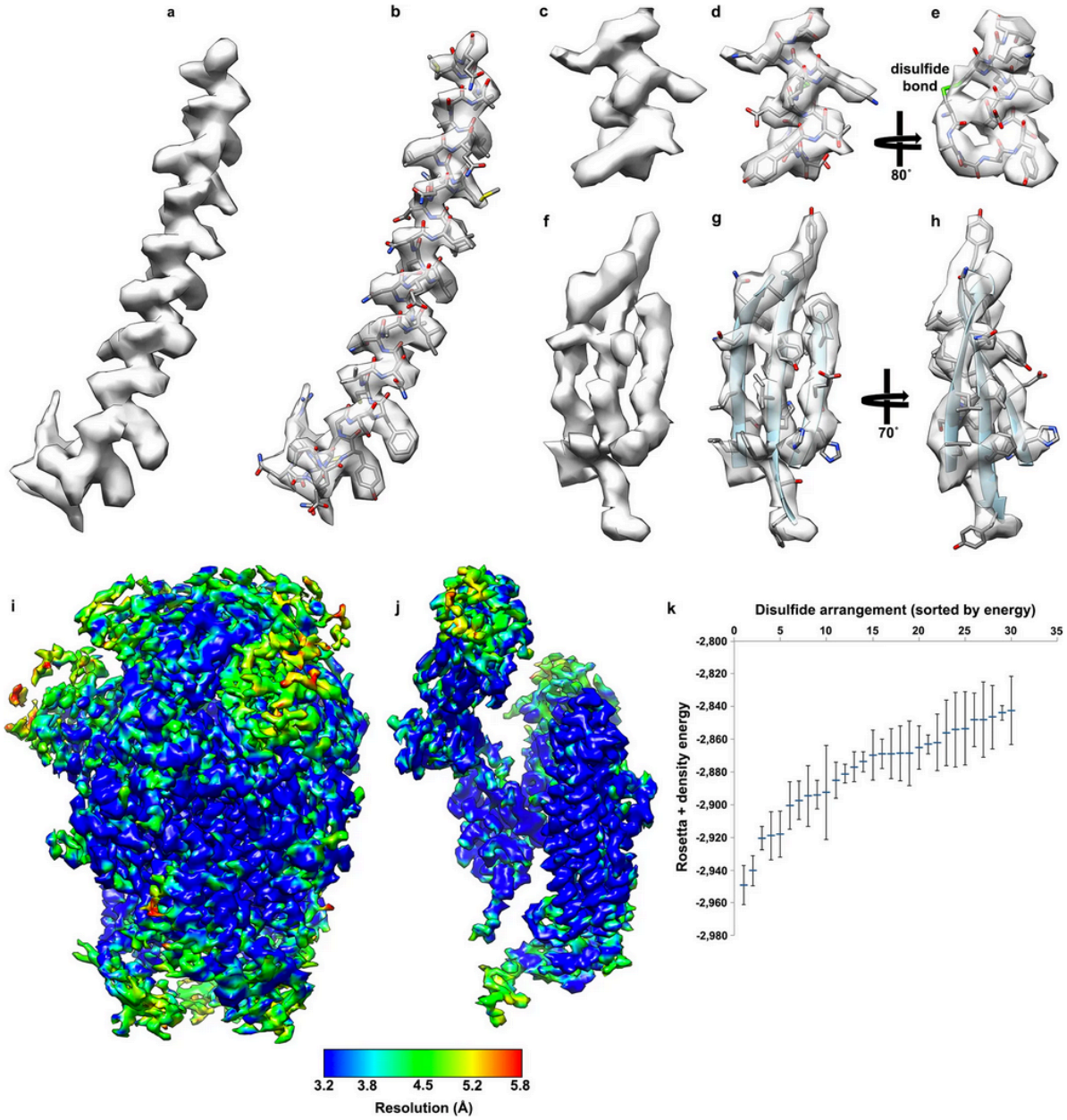
a–c, 3D map filtered at 4.0 Å resolution colored by protomer. Two different views of the S trimer (from the side (a) and from the top, looking towards the viral membrane (b)), and a side view of one S protomer (c) are shown. d–f, Ribbon diagrams showing the MHV S atomic model oriented as in a–c.

Figure II.III: CryoEM analysis of MHV S trimer.



(a)(b),Representative electron micrograph (defocus: 4.6 µm) (a) and class averages (b) of the MHV S trimer embedded in vitreous ice. Scale bars: 573 Å (micrograph) and 44 Å (class averages). c, Gold-standard (blue) and model/map (red) Fourier shell correlation (FSC) curves. The resolution was determined to 4.0 Å. The 0.143 and 0.5 cut-off values are indicated by horizontal grey bars.

Figure II.IV: CryoEM density for selected regions of the MHV S reconstruction, local resolution analysis and density-guided homology modelling of residues 453-535.

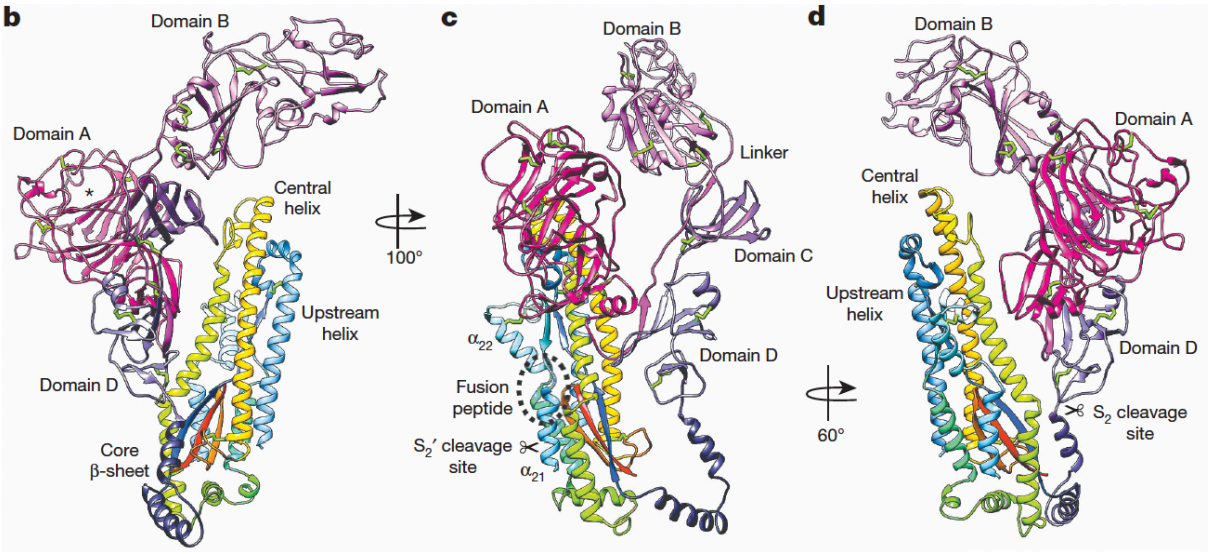


The atomic model is shown with the corresponding region of the map. a, b, Upstream helix. c–e, Helix belonging to domain A (residues 284–296). f–h, Core β-sheet. i, j, CryoEM density corresponding to the MHV S trimer (i) and a single protomer (j), colored according to local resolution determined with 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.

Figure II.V. Refinement and model statistics.

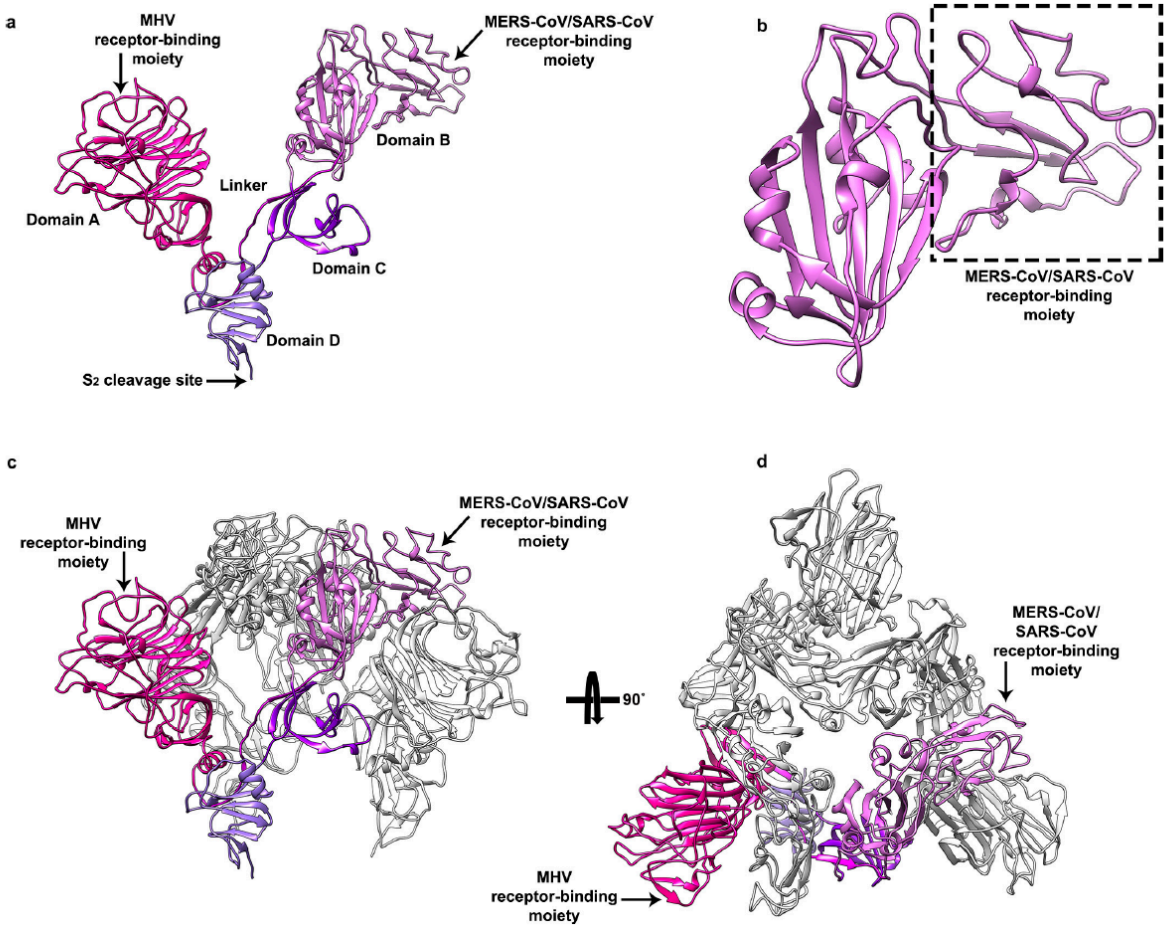
Parameter	Value
Data Collection	
Number of Particles	82,000
Pixel size, Å	1.315 (1.46)
Defocus range, μm	2-5
Voltage, kV	300
Electron Dose, $e^-/Å^2$	53
Refinement	
Resolution, Å	4.0
Map-sharpening B factor, $Å^2$	-220
Model validation	
Favored rotamers, %	99.32
Poor rotamers, %	0
Ramachandran favored, %	94.26
Ramachandran outliers, %	0.56
Clash score	3.68
MolProbity score	1.55

Figure II.VI. Architecture of MHV S protomer.



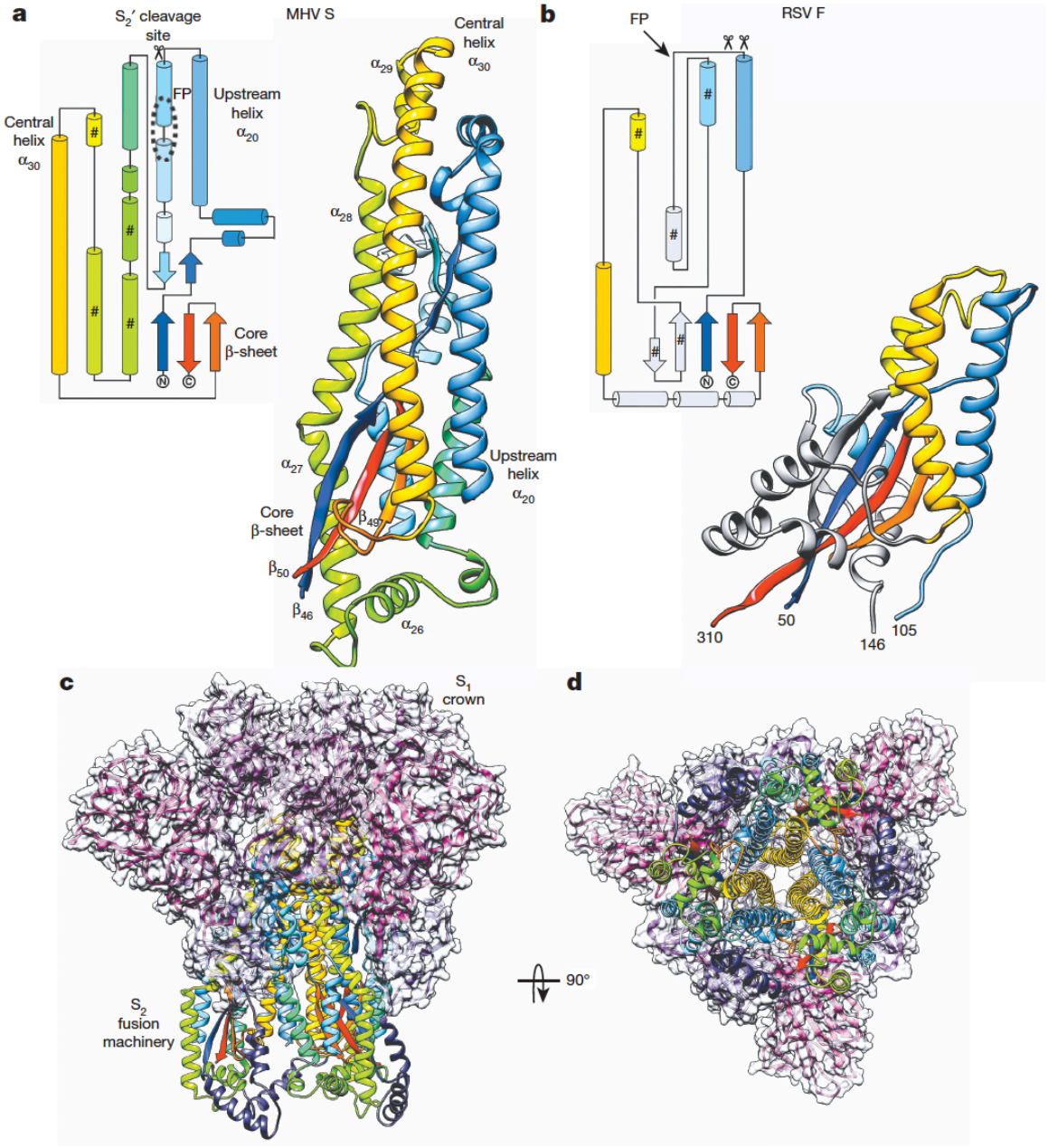
b–d, Ribbon diagrams depicting three views of the S protomer colored as in a. Asterisk denotes the MHV S receptor-binding region. Disulfide bonds are shown as green sticks except for residues 453–535, for which they are not shown.

Figure II.VII. Structural organization of the S_1 subunit.



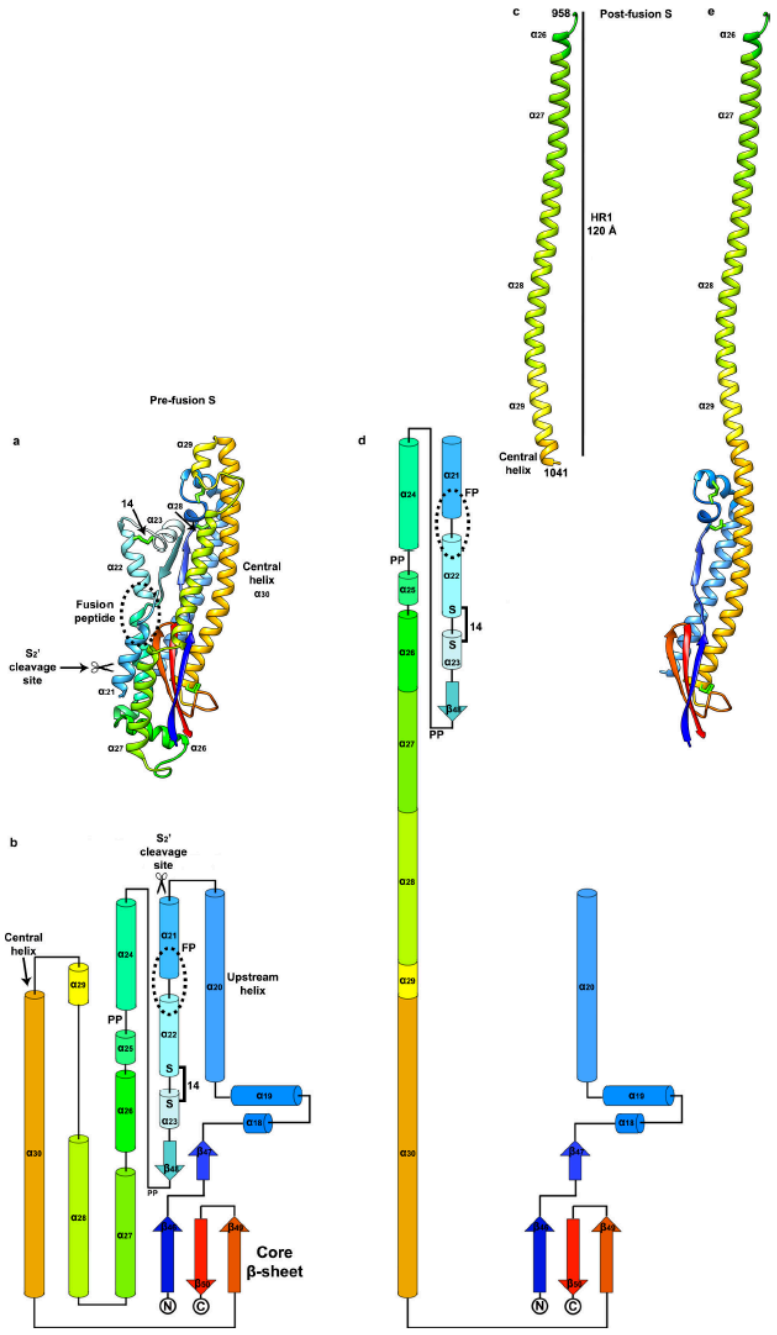
a, Ribbon diagram showing a single S_1 protomer. b, Close-up view of the MHV S domain B. The structural motif used as a receptor-interacting moiety by MERS-CoV and SARS-CoV is indicated. The density was too weak to allow tracing of this segment (residues 453–535), which has been traced by density-guided homology modelling using Rosetta. c, d, Ribbon diagrams of the S_1 trimer viewed from the side (c) and from the top (looking towards the viral membrane) (d).

Figure II.VIII. Pre-fusion structure of the coronavirus fusion machinery.



a, b, Topology and ribbon diagrams showing the structural similarity between coronavirus MHV S_2 (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 colored 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 S_1 (ribbon diagram and semi-transparent surface) wraps around the S_2 fusion machinery (ribbon diagram) to stabilize it.

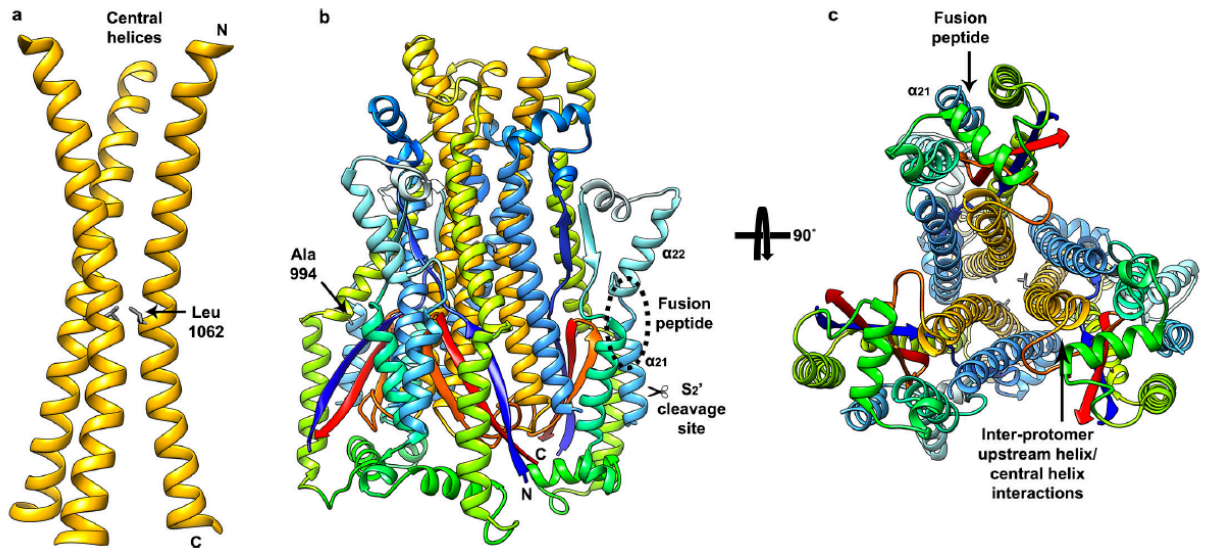
Figure II.IX. Mechanisms of membrane fusion promoted by CoV S glycoproteins.



a, Ribbon diagram of the MHV S_2 prefusion 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

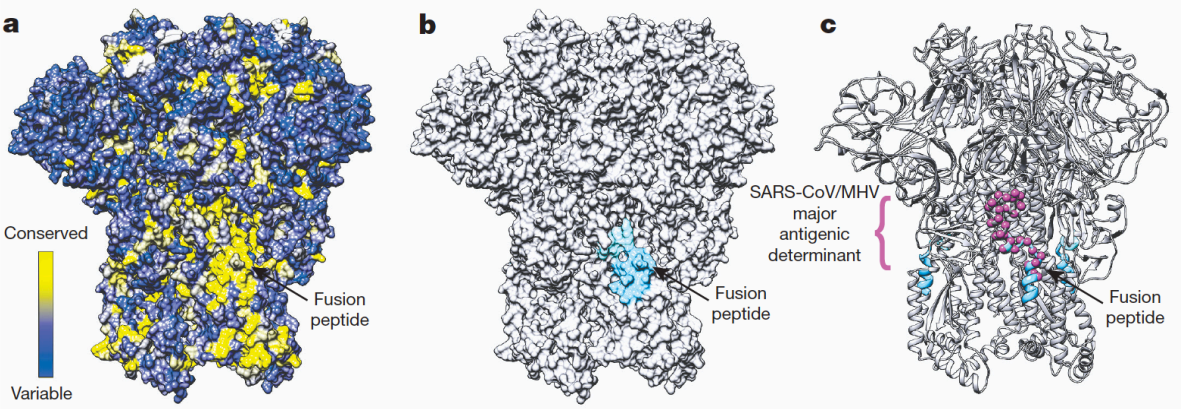
refolding of helices α22 and α23 as a single extended helix. c, Ribbon diagram of the SARS-CoV postfusion HR1 helix obtained by X-ray crystallography (PDB 1WYY). The residue numbers corresponding to the MHV A59 sequence are indicated. d, Topology diagram showing the expected coronavirus S post-fusion 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 α21, α22, α23, β48, α24 and α25 are not represented owing to a lack of structural information.

Figure II.X. Structural organization of the S_2 fusion machinery.



a, Ribbon diagram of the trimer of central helices. b, c, Ribbon diagrams of the S_2 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.

Figure XI. Potential strategy for neutralizing coronavirus infections.



a, Surface representation of the MHV S trimer colored according to sequence conservation. 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).

II.II Characterization of HCoV-NL63 Spike Protein

Our previously reported cryo-electron microscopy (cryoEM) reconstruction of the MHV S glycoprotein at 4.0 Å resolution revealed the pre-fusion architecture of the machinery mediating entry of β-coronaviruses into cells (Walls et al. 2016). It also demonstrated that coronavirus S and paramyxovirus F proteins share a common evolutionary origin. Here, we report the atomic resolution structure of the pathogenic HCoV-NL63 S glycoprotein trimer which belongs to the α-coronavirus genus. The significant resolution improvement compared to earlier studies allows visualization of the S glycoprotein with an unprecedented level of details, which is a prerequisite for guiding drug and vaccine design, and reveals shared and unique features of the α-genus of human pathogens. Our results suggest that NL63 and other coronaviruses use molecular trickery, based on masking of epitopes with glycans and activating conformational changes, to evade the immune system of infected hosts similarly to what has been described for HIV-1.

We produced the HCoV-NL63 S ectodomain in *Drosophila S2* cells, N-terminally fused to a GCN4 trimerization motif in register with the heptad-repeat 2 (HR2) helix. Frozen-hydrated HCoV-NL63 spike ectodomain particles were imaged using an FEI Titan Krios electron microscope equipped with a Gatan Quantum GIF energy-filter operated in zero-loss mode with a slit width of 20 eV and a Gatan K2 Summit electron counting camera (Li et al. 2013). We determined a three-dimensional reconstruction of the HCoV-NL63 spike at 3.4 Å resolution using the gold-standard Fourier shell correlation (FSC) criterion of 0.143 (Rosenthal & Henderson, 2003; Scheres & Chen, 2012) (**Figure II.XII, II.XIII**). The final model was built and refined using Coot (Brown et al. 2015) and Rosetta

(DiMaio et al. 2011; DiMaio et al. 2015; Song et al. 2013; Wang et al. 2015) and includes residues 28 to 1231 with an internal break between residues 994–998 (**Figure II.XII, II.XIV)**. The HCoV-NL63 S ectodomain is a 160 Å long trimer with a triangular cross-section.

A striking feature of this structure is the extraordinary number of N-linked oligosaccharides that cover the spike trimer. In the cryoEM reconstruction, we could observe density for 31 N-linked glycans extending tangentially relative to the protein surface (**Figure II.XIII, II. XV, II.XVI**). At least the two core N-acetyl glucosamine moieties are visible for the majority of glycosylation sites. Using on-line reversed phased liquid chromatography with electron transfer/high-energy collision-dissociation tandem mass-spectrometry (Frese et al. 2013), we detected 25 N-linked glycosylation sites overlapping with those observed in the cryoEM map and identified three additional sites (**Figure II.XV, II.XVI, II.XVII**) These sites were identified from both intact glycopeptides as well as peptides with the glycan trimmed down to the N-linked core N-acetyl glucosamine moiety. The cryoEM and mass-spectrometry data combined provides evidence for glycosylation at 34 out of 39 possible NXS/T glycosylation sequons. The intact glycopeptides detected by MS/MS for HCoV-NL63 S expressed in Drosophila S2 cells corresponded to either paucimannosidic glycans containing 3 mannose residues (with or without core fucosylation) or high-mannose glycans containing 4 to 9 mannose residues. Previous reports suggested that several coronavirus S glycans are of the high-mannose type, as a result of direct budding from the endoplasmic reticulum-golgi intermediate compartment (Jeffers et al. 2006; Ritchie et al. 2010), which underscores the biological relevance of the potential glycan structures identified.

In the refined model, N-linked glycans significantly cover the accessible surface of the trimer (**Figure II.XVI)**. The higher glycan density per accessible surface area detected for the S_2 subunits (865 $Å^2$/glycan) compared to the S_1 subunits (1350 $Å^2$/glycan) could explain why most coronavirus neutralizing antibodies isolated to date target the latter region. As many of the observed glycosylation sites are topologically conserved among coronavirus S proteins, we put forward that the glycan footprint observed here could be representative of other S proteins. Besides potentially contributing to immune evasion as discussed below, S glycans have been proposed to play a role in host cell entry (Zhou et al. 2010) using L-SIGN lectin which is an alternative receptor for SARS-CoV (Jeffers et al. 2004) and HCoV-229E (Jeffers et al. 2006) .

The HCoV-NL63 and MHV S_2 fusion machineries are structurally similar and can be superimposed with excellent agreement (**Figure II.XVIII, II.XIX**) (DALI (Holm & Rosenström, 2010) Z-score 30.4, rmsd 2Å over 312 residues). In contrast to our previous MHV S structure (Walls et al. 2016), most of the HCoV-NL63 S_2' trigger loop, which connects the upstream helix to the fusion peptide and participates in fusion activation, is resolved in the reconstruction (**Figure II.XVIII)**. It runs almost perpendicularly to the long axis of the S_2 subunit and forms three helical segments before looping back to connect to the fusion peptide. Multiple arginine residues, forming two putative furin cleavage sites, are present in the C-terminal region of the S_2' loop (863-RNIRSSR-870), which is characterized by weaker density as would be expected from a protease-sensitive polypeptide segment. These observations agree with previous studies suggesting fusion activation of the HCoV-NL63 S glycoprotein occurs upon S_2' proteolytic processing at the plasma membrane (by trypsin-like proteases such as TMPRS2) or in the endosomal

pathway (by furin or cysteine proteases) (Burkard et al. 2014; Kawase et al. 2012). The lack of strict amino acid sequence conservation at the S_2' cleavage site among coronavirus S proteins reflects the usage of different proteases found in distinct cellular compartments for fusion activation (Burkard et al. 2014; Millet & Whittaker, 2015). Similarly to the additional cleavage site present between the S_1 and S_2 subunits of MERS-CoV (Yang et al. 2015), protease sensitivity is likely further influenced by the presence of multiple glycans in the vicinity of the S_2' loop (**Figure II.XVIII***)*. However, we emphasize that S_2' processing occurs at topologically equivalent positions for HCoV-NL63 S, MERS-CoV S, MHV S and probably most coronavirus S glycoproteins.

The HCoV-NL63 S reconstruction (**Figure II.XVII**) resolves a large part of the S_2 C-terminal region that had not been observed in previous studies (Kirchdoerfer et al. 2016; Walls et al. 2016). Due to the significant resolution improvement, we were able to build an atomic model for the connector domain, so named as it connects to the HR2 region, and the stem helix. The connector folds as a β-rich domain decorated with one short α-helix. At its C-terminal end, the polypeptide chain folds as an α-helix (stem helix, **Figure II.XVIII**), aligned along the 3-fold molecular axis, which turns into the HR2 domain corresponding to 71 additional residues not resolved in our map. In the trimer, the connector domains assemble as a cup flanking the viral membrane proximal side of the ectodomain, and the stem helices form a bundle stabilized by hydrophobic interactions. The coronavirus S connector domain and the equivalent paramyxovirus F domain share a related topology although their tertiary structures are different and several structural motifs have been added to the latter domain throughout evolution (McLellan et al. 2013; Yin et al. 2006). Moreover, the trimer of stem helices assemble as a helical bundle

initiating the HR2 domain in a way reminiscent of the HRB region of paramyxovirus pre-fusion F structures (McLellan et al. 2013; Yin et al. 2006). These observations strengthen the evolutionary connection we previously proposed for the fusion machineries of these two viral families (Walls et al. 2016). Comparison of the pre-fusion HCoV-NL63 S_2 subunit with the structure of the post-fusion core implies that the C-terminal region of the connector domain and the stem helix must refold and/or change conformation to yield the canonical "trimer-of-hairpin" conformation that mediates fusion of the host and viral membrane in all class I fusion proteins (Harrison, 2008; Walls et al. 2016; Zheng et al. 2006) .

The HCoV-NL63 S structure shows the presence of an additional N-terminal domain, relative to β-coronaviruses. Phylogenetic analyses suggest this is a canonical feature of most α-coronavirus S glycoproteins (**Figure II.XX**). This domain, that we named domain 0, adopts a galectin-like β-sandwich fold supplemented with a three-stranded β-sheet similar to domain A (**Figure II.XX,** DALI Z-score 5.8, rmsd 4.1Å over 144 residues), suggesting a gene duplication event. Domain 0 interacts with the viral membrane proximal side of domain A and with domain D. We detected that domain 0 is also structurally similar to the VP8* sialic acid-binding domain of the rotavirus VP4 spike protein (**Figure II.XIX,** PDB 1KQR, DALI Z-score 7.8, rmsd 3.1Å over 114 residues). In line with this finding, domain 0 of transmissible gastro-enteritis coronavirus (TGEV) and of PEDV bind to sialic acid and deletion of this domain in α-coronavirus S appears to correlate with loss of enteric tropism (Krempl et al. 1997). We detected no sialic acid binding activity for HCoV-NL63 S_1 subunit (**Figure II.XXI)**, thus possibly explaining the strict respiratory tropism of this virus. Instead, host-cell heparan sulfate proteoglycans have been shown to

participate in HCoV-NL63 anchoring and infection (Milewska et al. 2014), and we detected binding of heparan sulfate to the HCoV-NL63 S protein by using surface plasmon resonance (SPR) (**Figure II.XXII),** or domain B, which has been reported to bind carbohydrates in the case of a bovine coronavirus (Peng et al. 2012).

Domain B, which is the HCoV-NL63 receptor-binding domain, exhibits a structure distinct from β-coronavirus B domains although a topological relatedness could be detected among these β-rich domains (Li, F., 2012). Superimposition of the HCoV-NL63 and MHV S_1 subunits highlights that their B domains feature opposite orientations related by a ~180° rotation (**Figure II.XXIII**). As a result, many of the HCoV-NL63 receptor-binding residues are buried through interaction with domain A of the same protomer, are masked by the glycan at residue Asn358 and are not available to engage the host-cell receptor (human angiotensin-converting enzyme 2, ACE2). Comparison of the HCoV-NL63 domain-B structure in our cryo-EM-derived model with the crystal structure of the same domain in complex with ACE2 (Wu et al. 2009) revealed that the receptor-binding loop containing residues 531–539 undergoes substantial conformational changes after binding (and is defined by weak density; **Figure II.XXIII)**. These findings explain the markedly higher ACE2 binding affinity of HCoV-NL63 domain B, compared with that of the full-length S_1 domain (**Figure II.XXIII).**

Because the receptor-binding loops are known to elicit potent neutralizing antibodies in the case of TGEV (Reguera et al. 2012), MERS-CoV (Ying et al. 2015) and SARS-CoV (Hwang et al. 2006; Prabakaran et al. 2006; Sui et al. 2004; Zhu et al. 2007), we speculate that HCoV-NL63 has evolved to limit exposure of this vulnerable site to B-cell receptors via protein-protein interactions and glycan masking. This mechanism is

reminiscent of the HIV-1 immune evasion strategy which relies on a glycan shield and conformational changes triggered by binding of CD4 to expose the chemokine receptor-interacting motif (V3 loop) (Chen et al. 2005; Huang et al. 2005).

Viruses have evolved several immune evasion strategies including rapid antigenic evolution, masking of epitopes, and exposure of non-neutralizing immune-dominant "decoy" epitopes. For instance, HIV-1 (Stewart-Jones et al. 2016), Lassa virus (Sommerstein et al. 2015), Hepatitis C virus (Falkowska et al. 2007) and Epstein Barr virus (Szakonyi et al. 2006) exhibit extensive N-linked glycosylation, covering exposed protein surfaces, and whose mass can exceed that of the protein component. The HCoV-NL63 S trimer is covered by an extensive glycan shield consisting of 102 N-linked oligosaccharides obstructing the protein surface. This observation is reminiscent of what has been described for the HIV-1 envelope trimer (Stewart-Jones et al. 2016) although the glycan density is 30% higher in the latter case. Furthermore, our data suggests that, similarly to HIV-1, coronavirus S glycans are masking the protein surface to limit access to neutralizing antibodies and thwart the humoral immune response. This strategy is illustrated by the presence of a glycan linked to asparagine 358 in the HCoV-NL63 structure reported here. This glycan contributes, along with the proteinaceous moiety of domain A, to masking the receptor-binding loops, that have been shown to elicit potent neutralizing antibodies *in vivo* (Reguera et al. 2012) and *in vitro* (Ying et al. 2015), and appear to represent a potential Achilles' heel of coronaviruses. This hypothesis is further supported by the observation of three additional glycans directly protruding from the viral membrane distal side of domain B. As a result, conformational changes are required for the HCoV-NL63 S glycoprotein to be able to interact with ACE2 (Wu et al. 2009). These

rearrangements and/or receptor binding likely participate to initiate the fusion reaction by disrupting the interactions formed between domain B and the HR1 C-terminal region. It is tempting to speculate that interactions with heparan sulfate proteoglycans present at the host cell surface could contribute to activating HCoV-NL63 S and promote subsequent interactions with ACE2. A common theme arising from the analysis of α- and β-coronavirus S glycoprotein structures is that domain B-mediated host anchoring involves major structural rearrangements to expose the binding motifs (Kirchdoerfer et al. 2016; Walls et al. 2016) .

Visualization of the glycan shield obstructing access to the S surface and deciphering the molecular trickery employed by some coronaviruses provides a rational basis to understand accessibility to neutralizing antibodies and paves the way for guiding future immunogen and therapeutics design. We previously suggested that targeting the fusion machinery bears the promise of finding broadly neutralizing inhibitors of coronavirus infection (Walls et al. 2016) and the high-density of glycans decorating this region will have to be taken into consideration to increase the likelihood of success.

II.IV Methods for Characterization of HCoV-NL63 Spike Protein and Glycan Shield

A gene fragment encoding the HCoV-NL63 S ectodomain (residues 16–1291, UniProt Q6Q1S2) was PCR-amplified from a plasmid containing the full-length S gene. The PCR product was ligated to a gene fragment encoding a GCN4 trimerization motif (LIKRMKQIEDKIEEIESKQKKIENEIARIKKIK) (Eckert et al. 1998; Walls et al. 2016; Yin et al. 2006), a thrombin-cleavage site (LVPRGSLE), an eight-residue-long Strep-Tag

(WSHPQFEK) and a stop codon. Subsequent cloning was performed in the pMT-BiP-V5-His expression vector (Invitrogen) in frame with the *Drosophila* BiP secretion signal downstream the metallothionein promoter.

To generate a stable *Drosophila* S2 cell line expressing the recombinant HCoV-NL63 S ectodomain, we used Effectene (Qiagen) and 2 µg of plasmid. Puromycin *N*-acetyltransferase was cotransfected and used as a dominant selectable marker. Stable HCoV-NL63 S–expressing cell lines were selected by addition of 7 µg/ml puromycin (Invivogen) to the culture medium 48 h after transfection. For large-scale production, the cells were cultured in spinner flasks and induced by 5 µM of $CdCl_2$ at a density of approximately 10^7 cells/mL. After one week at 28 °C, clarified cell supernatants were concentrated 40-fold with Vivaflow tangential filtration cassettes (Sartorius, 10-kDa cutoff) and adjusted to pH 8.0, before affinity purification with a StrepTactin Superflow column (IBA) followed by gel-filtration chromatography with a Superose 6 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, and 100 mM NaCl. The purified protein was quantified according to absorption at 280 nm and concentrated to approximately 3 mg/mL.

2 µl of purified HCoV-NL63 spike at 1.0 mg/mL was applied to a 1.2/1.3 C-flat grid (Protochips), which had been glow-discharged for 30 s at 20 mA. Grids were then plunge-frozen in liquid ethane with an FEI Mark I Vitrobot with 7.5-s blot time and an offset of −3 mm at 100% humidity and 25 °C. Data were collected with Leginon automatic data-collection software (Suloway et al. 2005) on an FEI Titan Krios operated at 300 kV and equipped with a Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit

width of 20 eV, and a Gatan K2 Summit direct electron detector camera. The dose rate was adjusted to 8 counts/pixel/s, and each movie was acquired in counting mode fractionated in 50 frames of 200 ms. 1,400 micrographs were collected in a single session with a defocus range between 2.0 and 4.0 μm.

Whole-frame alignment was carried out with DOSEFGPU DRIFTCORR (Li et al. 2013). The parameters of the microscope contrast-transfer function were initially estimated with CTFFIND4 (Rohou & Grigorieff, 2015) and then with GCTF (Zhang, 2016). Micrographs were manually masked with Appion (Lander et al. 2009) to exclude the visible carbon edge from images. Particles were automatically picked with DoGPicker (Voss et al. 2009). Particle images were extracted and processed with Relion 1.4(Scheres, 2012) with a box size of 320 pixels^2 and a pixel size of 1.36 Å. After reference-free 2D classification, we retained 180,000 out of 474,000 particles to run 3D classification with C1 symmetry (Scheres, 2012). We used the initial model previously generated for MHV (Walls et al. 2016) with Optimod (Lyumkis et al. 2013b) and low-pass-filtered the data to 60 Å as a starting reference for 3D classification. 118,000 particles were selected and used to run gold-standard 3D refinement with RELION (Scheres & Chen, 2012), thus yielding a map at 3.95-Å resolution. After particle-motion and radiation-damage correction with Relion particle polishing (Scheres, 2014), another round of 3D classification with C3 symmetry was performed to select 79,667 particles. After gold-standard 3D refinement with this subset of particles, we obtained a reconstruction at 3.76-Å resolution. Per-particle defocus parameters were estimated with GCTF and used to run an identical round of 3D refinement that yielded the final 3.4-Å-resolution map. Post processing was performed with Relion to apply an automatically generated *B* factor of

−129 Å^2. Reported resolutions were based on the gold-standard FSC = 0.143 criterion (Rosenthal & Henderson, 2003; Scheres & Chen, 2012), and FSC curves were corrected for the effects of soft masking by high-resolution noise substitution (Chen et al. 2013). The soft mask used for FSC calculation had a 10-pixel cosine-edge fall-off.

UCSF Chimera (Goddard et al. 2007) and Coot (Brown et al. 2015; Emsley et al. 2010) were used to fit atomic models into the cryo-EM map. The MHV S_2 subunit was fit into the density and rebuilt manually in Coot. The crystal structure of HCoV-NL63 domain B was then fit into the density, and the rest of the S_1 subunit was built with a combination of manual building in Coot and *de novo* building with Rosetta (DiMaio et al. 2015; Song et al. 2013; Wang et al. 2015). Glycan density coming after an NXS/T motif was initially manually built into the density, and glycan geometry was then refined with Rosetta, optimizing the fit-to-density as well as the energetics of protein-glycan contacts. The glycans were not as well defined as the protein region in the reconstruction, owing to flexibility and compositional heterogeneity. The final model was refined by application of strict noncrystallographic symmetry constraints with Rosetta, with 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 with the atomic model and preventing overfitting (DiMaio et al. 2013). The quality of the final model was analyzed with MolProbity (Chen et al. 2010) and Privateer (Agirre et al. 2015). Structure analysis was performed with the DALI server (Holm & Rosenström, 2010) and areaimol (Lee & Richards, 1971). Electrostatic-potential calculations were performed with PDB2PQR (Dolinsky et al. 2004) and APBS (Baker et

al. 2001). All figures were generated with UCSF Chimera (Goddard et al. 2007). Local resolution estimation was performed with Resmap (Kucukelbir et al. 2014) .

HCoV-NL63 S was prepared for MS analysis unaltered or subjected to Endo H (NEB), subjected to Endo F3 (Millipore) or subjected to combined Endo H and Endo F3 deglycosylation treatment. 2 µl of the relevant endoglycosidases was incubated with 20 µg of HCoV-NL63 S for 14 h overnight in 50 mM sodium acetate, pH 4.4, at 37 °C in a 20-µL reaction. 6 µg of HCoV-NL63 S was then incubated in a freshly prepared solution containing 100 mM Tris, pH 8.5, 2% sodium deoxycholate, 10 mM Tris(2-carboxyethyl)phosphine and 40 mM iodoacetamide at 95 °C for 5 min; this was followed by an incubation at 25 °C for 30 min in the dark. 1.6 µg of denatured, reduced and alkylated HCoV-NL63 S was then diluted into freshly prepared 50 mM ammonium bicarbonate and incubated for 14 h at 37 °C with 0.032 µg of either trypsin (Sigma Aldrich) or chymotrypsin (Sigma Aldrich). Formic acid was then added to a final concentration of 2% to precipitate the sodium deoxycholate in the samples. Samples were then centrifuged at 14,000 r.p.m. for 20 min. The supernatant containing the (glyco)peptides was collected and spun again at 14,000 r.p.m. for 5 min immediately before sample analysis. Between 4 and 7 µL was run on a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer. A 35-cm analytical column and a 3-cm trap column filled with ReproSil-Pur C18AQ 5 µm (Dr. Maisch) beads were used. Nanospray LC-MS/MS was used to separate peptides over a 110-min gradient from 5% to 30% acetonitrile with 0.1% formic acid. A positive spray voltage of 2,100 was used with an ion-transfer-tube temperature of 350 °C. An electron-transfer/higher-energy collision dissociation ion-fragmentation scheme (Frese et al. 2013) was used with calibrated charge-dependent

ETD parameters and a supplemental higher-energy collision dissociation energy of 0.15 for the samples with intact glycopeptides and 0.2 for the samples treated with endoglycosidases. A resolution setting of 120,000 with an AGC target of 2×10^5 was used for MS1, and a resolution setting of 30,000 with an AGC target of 1×10^5 was used for MS2. The data were searched against a custom database including recombinant coronavirus S-glycoprotein sequences, a list of common contaminant proteins including trypsin, chymotrypsin and the endoglycosidases, as well as 998 decoy reverse yeast sequences, with trypsin or chymotrypsin as the protease, allowing up to two missed cleavages. All searches included carbamidomethylation of cysteine as a fixed modification and oxidation of methionine as a variable modification. An initial comprehensive search for glycosylation revealed that (core-fucosylated) paucimannose and high-mannose structures were the only identified glycan species in the samples. On the basis of these findings, a final search was performed with COMET(Eng, Jahan, & Hoopmann, 2013) on the same data with the following list of variable modifications of asparagine residues: +HexNAc(2)Hex(3), +HexNAc(2)Hex(3)dHex(1), +HexNAc(2)Hex(3)dHex(2), +HexNAc(2)Hex(4), +HexNAc(2)Hex(5), +HexNAc(2)Hex(6), +HexNAc(2)Hex(7), +HexNAc(2)Hex(8) and +HexNAc(2)Hex(9). The samples treated with endoglycosidases were searched with +HexNAc, +HexNAc(1)dHex(1) and +HexNAc(1)dHex(2) as variable modifications of asparagine. We used a precursor mass tolerance of 20 p.p.m., 0.02 fragment bin size, including b/c/y/z fragments, with monoisotopic masses for both precursor and fragment ions. The search results were filtered for modification of asparagine residues and the presence of an NX(S/T) sequon at the protein level. All appropriate peptide spectrum matches (PSMs)

were manually inspected, and only those with reasonable peptide sequence coverage were kept. In addition, the spectra were inspected for the presence of glycan fragment ions. All glycosylation sites identified by MS listed are based on multiple PSMs, often with multiple different glycans and additional confirmation from overlap between the trypsin- and chymotrypsin-treated samples. The greatest number of glycopeptide identifications was made in the chymotrypsin-digested samples.

The S_1 subunit of HCoV-NL63 C-terminally tagged with the Fc portion of human IgG (S_1-Fc) was tested alone or premixed with 1 µl of Protein A–coupled, 200-nm-sized nanoparticles (nano-screenMAG-Protein A beads; Chemicell, cat.no. 4503-1) to increase the avidity of S_1-Fc proteins for sialic acids on the erythrocyte surface. The sialic acid–binding S_1 subunit of PEDV (strain GDU, GenBank AFP81695.1) C-terminally fused to the human Fc portion was used as a positive control. 'Mock' indicates the conditions in which no S_1 subunit was used (negative control). The initial concentration of S_1-Fc was 5 µg, and two-fold serial dilutions of S_1-Fc-nanoparticle mixtures were made in 50 µl phosphate-buffered saline supplemented with 0.1% bovine serum albumin. 50 µl erythrocyte suspension (0.5%) was mixed with 50 µl of S_1-Fc-nanoparticle dilution in V-shaped 96-well plates and incubated for 2 h on ice, after which the wells were photographed.

Different S_1 variants of HCoV-NL63 S protein, including S_1 (residues 1–718), S_1 domain 0 (S_1-0, residues 1–209) and S_1 domain B (S_1-B, residues 481–616), were C-terminally fused to the Fc region of mouse IgG (mFc), expressed in HEK-293T cells and affinity purified as previously described (Raj et al. 2018). Likewise, an S_1-mFc expression

plasmid was made for the SARS-CoV S_1 domain (isolate CUHK-W1, residues 1–676) and the PEDV S_1 domain (strain GDU; residues 1–728). Expression of the human angiotensin-converting enzyme ectodomain (ACE2; residues 1–614) fused to the Fc portion of human IgG (hFc) was performed as previously described (Raj et al. 2013) .

The ability of the HCoV-NL63 S_1-mFc and S_1-B-mFc chimeric proteins to bind the ACE2-hFc receptor was evaluated with an ELISA-based assay. 100 µl of hACE2-hFc (20 µg/ml, diluted in PBS) was coated on a 96-well MaxiSorb plate overnight at 4 °C. Nonspecific binding sites were subsequently blocked with a 3% (w/v) solution of bovine serum albumin in PBS. Plates were washed with washing buffer (PBS with 0.05% Tween 20) and subsequently incubated with serially diluted S_1-mFc proteins (starting with equimolar concentrations) for 1 h at room temperature, after which plates were washed three times with washing buffer. mFc-tagged S_1 proteins were detected with HRP-conjugated polyclonal rabbit-anti-mouse immunoglobulins (1:2,000 dilution in PBS with 0.1% BSA; DAKO, P0260), and a colorimetric reaction was produced after incubation with tetramethylbenzidine substrate (BioFX). The optical density (OD) was subsequently measured at 450 nm with an ELISA reader (EL-808, BioTEK). Background (signal from HRP-conjugated anti-mFc antibody alone) was subtracted from the OD_{450nm} values. The mFc-tagged SARS-CoV S_1 subunit was used as a positive control, whereas the mFc-tagged HCoV-NL63 S_1 domain 0 (HCoV-NL63 S_1-0-mFc) and PEDV S_1 subunit (PEDV S_1-mFc), both of which do not bind ACE2, were used as negative controls.

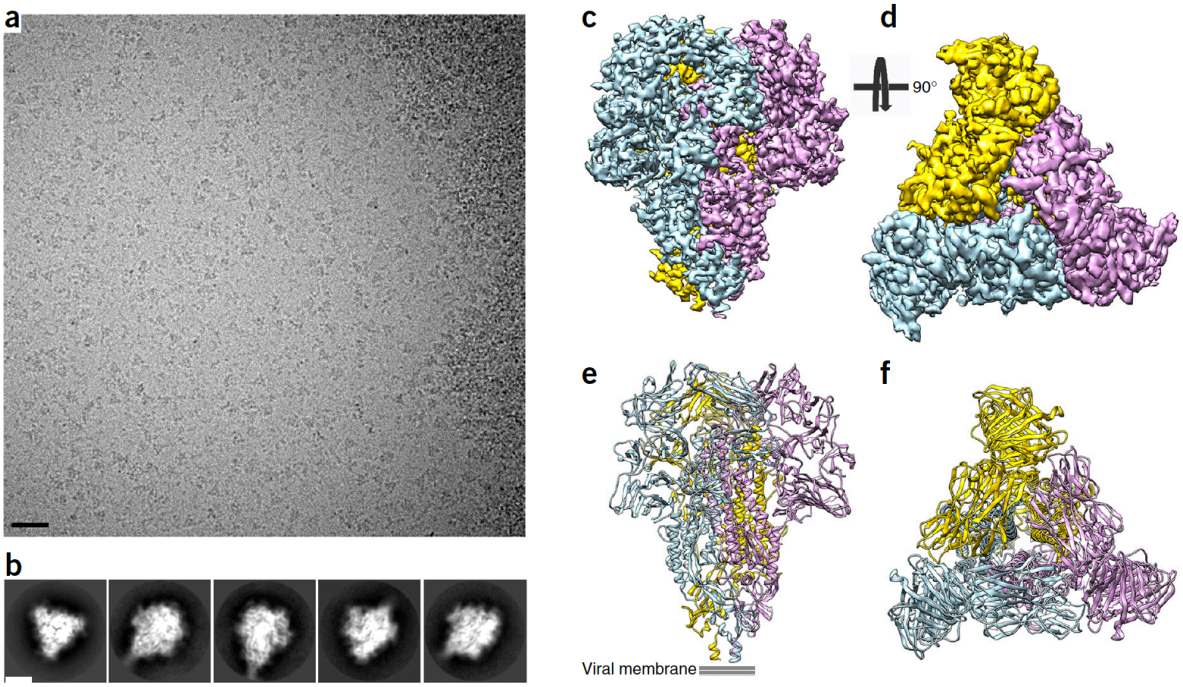
SPR was performed on a GE Healthcare Biacore T200 with a running buffer containing 20 mM HEPES, pH 7.5, 100 mM NaCl and 0.5% Tween-20, with a flow rate of

30 μL/min at 25 °C. A carboxymethylated dextran (CM5) chip (GE Healthcare) was activated with *N*-hydroxysulfosuccinimide (NHS) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). We then either quenched the CM5 surface with ethanolamine (yielding a blank flow cell) or immobilized HCoV-NL63 S before quenching. 10 μg of HCoV-NL63 S was diluted into 10 mM sodium acetate, pH 5.5 and was directly immobilized for 700 s, thus yielding 28,000 RUs. After immobilization quenching, running buffer was flowed for 10 min to ensure a steady baseline before experimental binding. Heparan sulfate (Sigma Aldrich) was reconstituted in running buffer at 5.0 mg/mL. Two concentrations of heparan sulfate, 5.0 mg/mL and 2.5 mg/mL, were injected for 80 s with a dissociation time of 400 s. All data were subtracted from the blank flow cell, to account for any nonspecific interactions of heparan sulfate with the CM5 chip, and the baseline was normalized to 0.

The cryo-EM map has been deposited in the Electron Microscopy Data Bank under accession code EMD-8331; the corresponding atomic model has been deposited into the Protein Data Bank under accession code PDB 5SZS. The MS data (including the raw data, COMET search results and annotated tandem MS spectra of all accepted glycopeptide identifications) have been deposited in the proteomics identifications (PRIDE) database under dataset PXD004557.

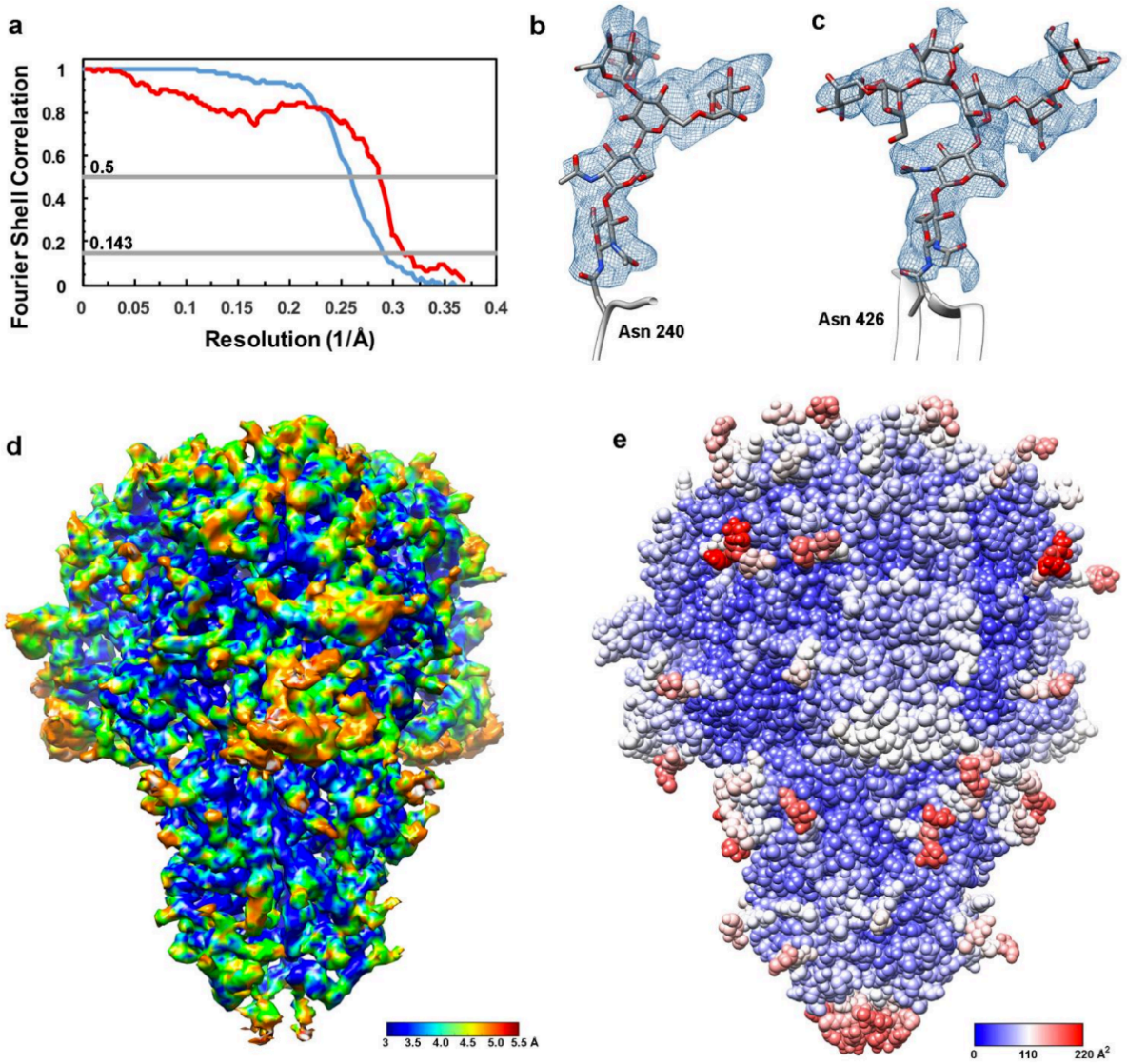
Figures Chapter II: II.XI-II.XXII: Adapted from Walls et al. 2016a

Figure II.XII: Cryo-EM structure of the HCoV-NL63 S trimer at atomic resolution.



(**a**) Representative micrograph of frozen-hydrated HCoV-NL63 S particles (defocus 3.4 µm). Scale bar, 355 Å. (**b**) Five selected class averages showing the particles along different orientations. Scale bar, 60 Å. (**c**,**d**) 3D map filtered at 3.4-Å resolution and colored by protomer. Two orthogonal views of the S trimer from the side (**c**) and from the top, facing toward the viral membrane, (**d**) are shown. (**e**,**f**) Ribbon diagrams showing the HCoV-NL63 S atomic model.

Figure II.XIII: Cryo-EM analysis of the HCoV-NL63 S trimer.



a, Gold-standard (blue) and model/map (red) Fourier shell correlation (FSC) curves. The resolution was determined to 3.4 Å. The 0.143and 0.5 cut-off values are indicated by horizontal grey bars. b,The glycan linked to Asn 240 is rendered as ball and sticks and the corresponding region of the cryoEM map is shown as a blue mesh. c,The glycanlinked to Asn 426isrendered as ball and sticks and the corresponding region of the cryoEM map is shown as a blue mesh. In panels (b-c), carbon, nitrogen and oxygen atoms are colored grey, blue and red, respectively. d,HCoV-NL63 S cryoEM map colored according to local resolution.e,HCoV-NL63 S atomic model colored according to refined B factors.

Figure II.XIV: Data collection and refinement statistics

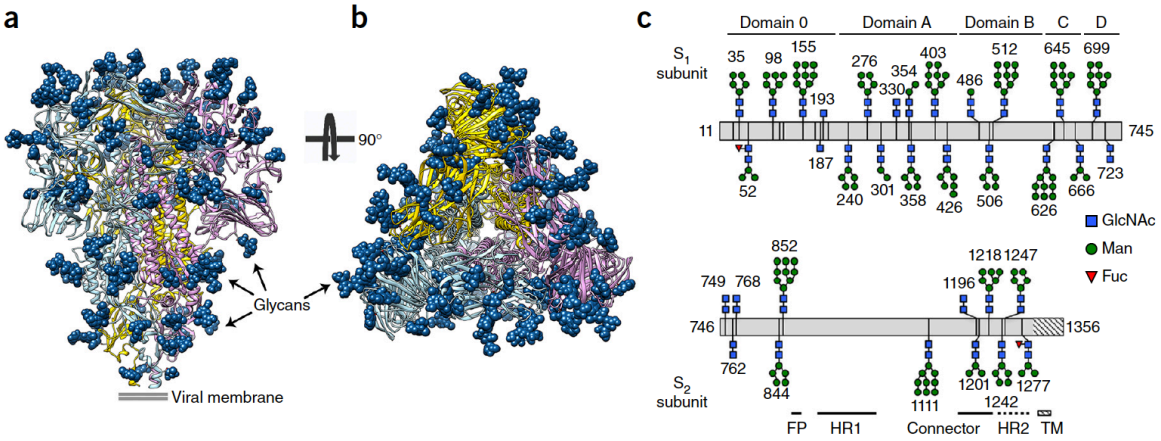
Parameter	Value
Data Collection	
Number of Particles	79,667
Pixel size, Å	1.36
Defocus range, μm	2-4
Voltage, kV	300
Electron Dose, $e^-/Å^2$	48
Refinement	
Resolution, Å	3.4
Map-sharpening B factor, $Å^2$	-129
Model validation	
Favored rotamers, %	97.87
Poor rotamers, %	0.68
Ramachandran favored, %	99.32
Ramachandran outliers, %	0.6
Clash score	3.3
MolProbity score	1.54

Figure II.XV: Characterization of the HCoV-NL63 S glycans by mass spectrometry and cryoEM

site	Sequon	Com-bined	MS deglyc	MS intact	EM	Glycan Composition from peptide ID	Glycan residues traced in map
24	NLSM	ND	ND	ND	ND	ND (Not detected)	ND
35	NSST	35	35	35	35	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5)	Not built
52	NQST	52	52	52	52	HexNAc(2)Hex(3)dHex(1)	Not built
98	NASV	98	98	98	98	HexNAc(2)Hex(5); HexNAc(2)Hex(6)	Not built
111	NTTF	ND	ND	ND	ND	ND	ND
119	NASS	ND	ND	ND	ND	ND	ND
155	NVTR	155	155	155	155	HexNAc(2)Hex(7); HexNAc(2)Hex(8); HexNAc(2)Hex(9)	HexNAc(1)
178	NYSC	ND	ND	ND	ND	ND	ND
187	NATV	187	187	ND	187	ND	HexNAc(1)
193	NVTT	193	193	ND	193	ND	HexNAc(1)
203	NYTV	ND	ND	ND	ND	ND	ND
240	NGST	240	240	240	240	HexNAc(2)Hex(5)	HexNAc(2)Hex(4)
276	NATG	276	276	276	276	HexNAc(2)Hex(5)	HexNAc(2)
301	NFSA	301	301	ND	301	ND	HexNAc(2)Hex(2)
330	NSSS	330	ND	ND	330	ND	HexNAc(2)
354	NSTI	354	ND	ND	354	ND	HexNAc(2)Hex(2)
358	NTTH	358	ND	358	358	HexNAc(2)Hex(5)	HexNAc(2)
403	NVTT	403	403	403	403	HexNAc(2)Hex(6); HexNAc(2)Hex(8)	HexNAc(2)
426	NVSA	426	ND	ND	426	ND	HexNAc(2)Hex(6)
486	NFTA	486	486	ND	486	ND	HexNAc(2)Hex(1)
506	NISL	506	ND	506	506	HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5)	HexNAc(2)
512	NTSV	512	ND	512	512	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(7); HexNAc(2)Hex(8); HexNAc(2)Hex(9)	HexNAc(1)
626	NCTK	626	626	626	626	HexNAc(2)Hex(5); HexNAc(2)Hex(8); HexNAc(2)Hex(9)	HexNAc(1)

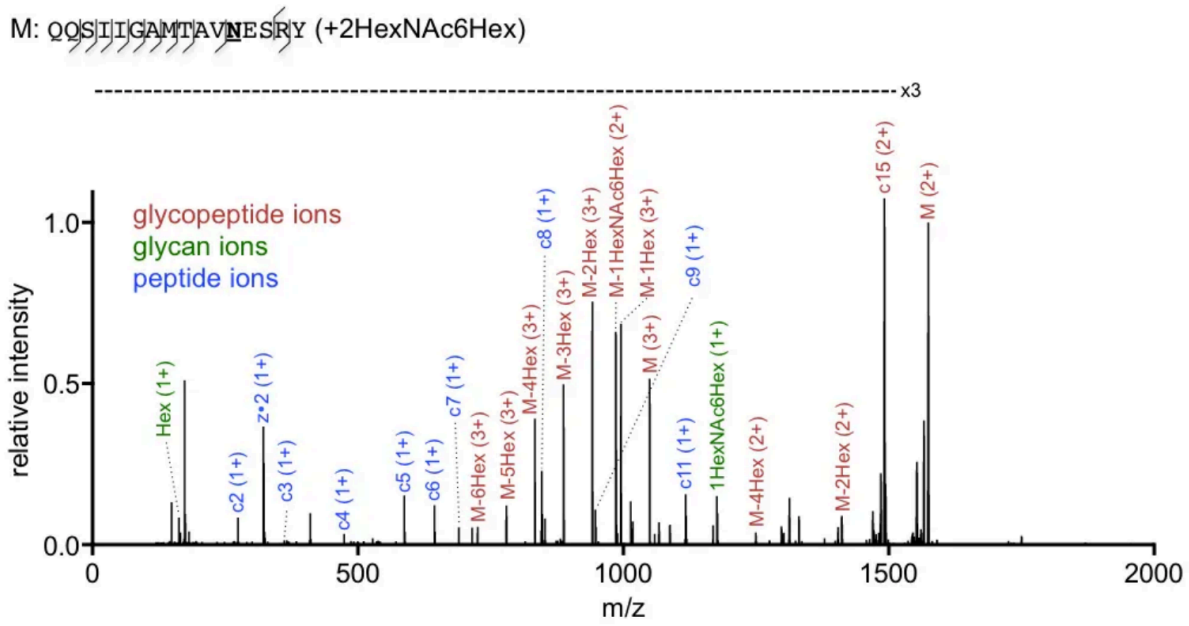
645	NQSL	645	645	645	645	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5); HexNAc(2)Hex(6); HexNAc(2)Hex(8)	HexNAc(1)
666	NVST	666	666	ND	666	ND	HexNAc(2)Hex(3)
699	NESR	699	699	699	699	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5); HexNAc(2)Hex(6); HexNAc(2)Hex(7); HexNAc(2)Hex(8)	HexNAc(2)Hex(1)
723	NCTT	723	ND	ND	723	ND	HexNAc(2)
762	NLSI	762	ND	ND	762	ND	HexNAc(2)
768	NWTT	768	768	ND	768	ND	HexNAc(2)
844	NVTS	844	844	844	844	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5)	HexNAc(2)
852	NLSS	852	852	852	852	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5); HexNAc(2)Hex(8); HexNAc(2)Hex(9)	HexNAc(2)Hex(2)
1111	NGTH	1111	1111	1111	1111	HexNAc(2)Hex(9)	HexNAc(2)Hex(3)
1196	NVTF	1196	1196	ND	1196	ND	HexNAc(2)
1201	NISR	1201	1201	1201	1201	HexNAc(2)Hex(3)	HexNAc(2)Hex(1)
1218	NKTL	1218	1218	1218	1218	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5); HexNAc(2)Hex(6)	HexNAc(1)
1242	NLTY	1242	1242	1242	ND	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5)	ND
1247	NLSS	1247	1247	1247	ND	HexNAc(2)Hex(3); HexNAc(2)Hex(3)dHex(1); HexNAc(2)Hex(5)	ND
1277	NSTY	1277	1277	1277	ND	HexNAc(2)Hex(3)dHex(1)	ND
TOTAL # of sites							
39		34	25	21	31		

Figure II.XVI: Organization of the HCoV-NL63 S-protein glycan shield, revealed by cryo-EM and MS.



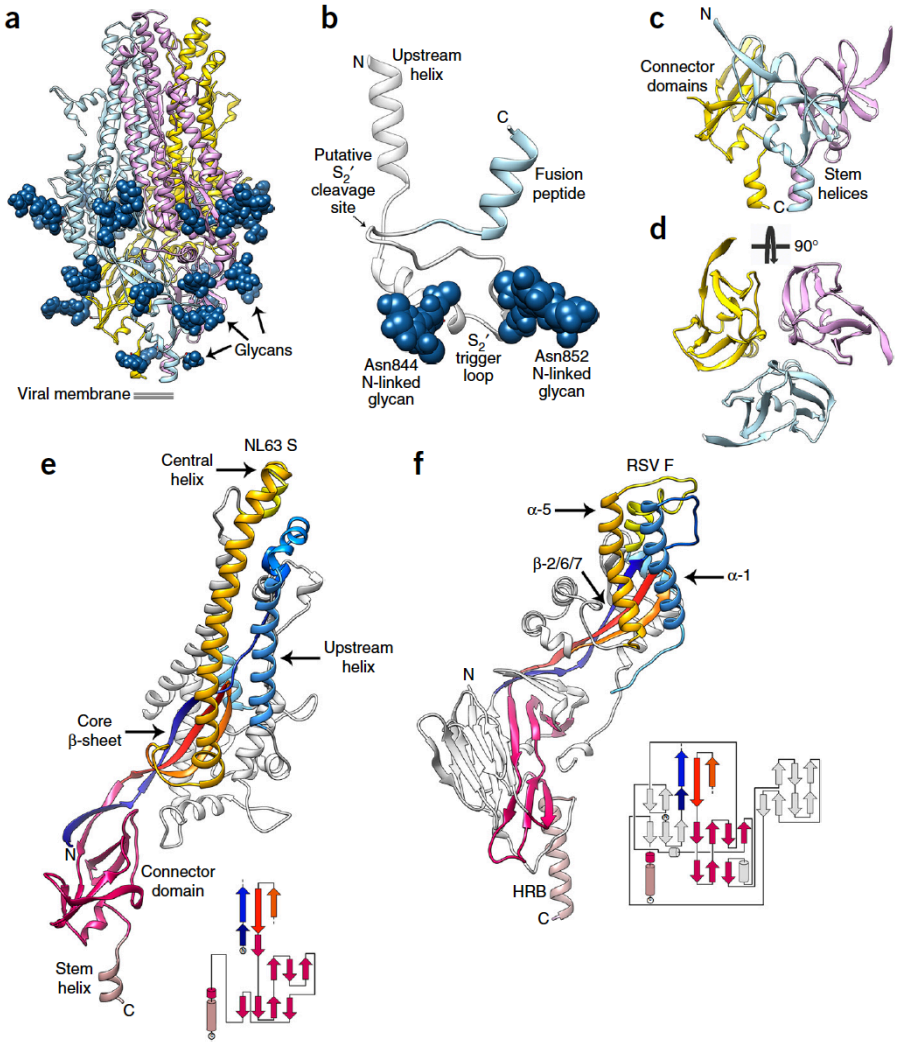
(**a**, **b**) Ribbon diagrams showing two orthogonal views of the S trimer, from the side (**a**) and from the top (**b**), facing toward the viral membrane. Glycans are shown as dark-blue spheres. (**c**) Residue-level schematic of N-linked glycans. The most extensive glycan structure detected by MS at each site is represented except for glycans observed only by cryo-EM, for which the resolved sugar moieties are shown. FP, fusion peptide; HR1, heptad-repeat 1 region; HR2, heptad-repeat 2 region (shown with a dashed line because it is not resolved in the map); TM, transmembrane domain (the striated texture indicates regions that are not part of the construct); GlcNac, *N*-acetylglucosamine; Man, mannose; Fuc, fucose.

Figure II.XVII Characterization of the HCoV-NL63 S glycans by using mass spectrometry.



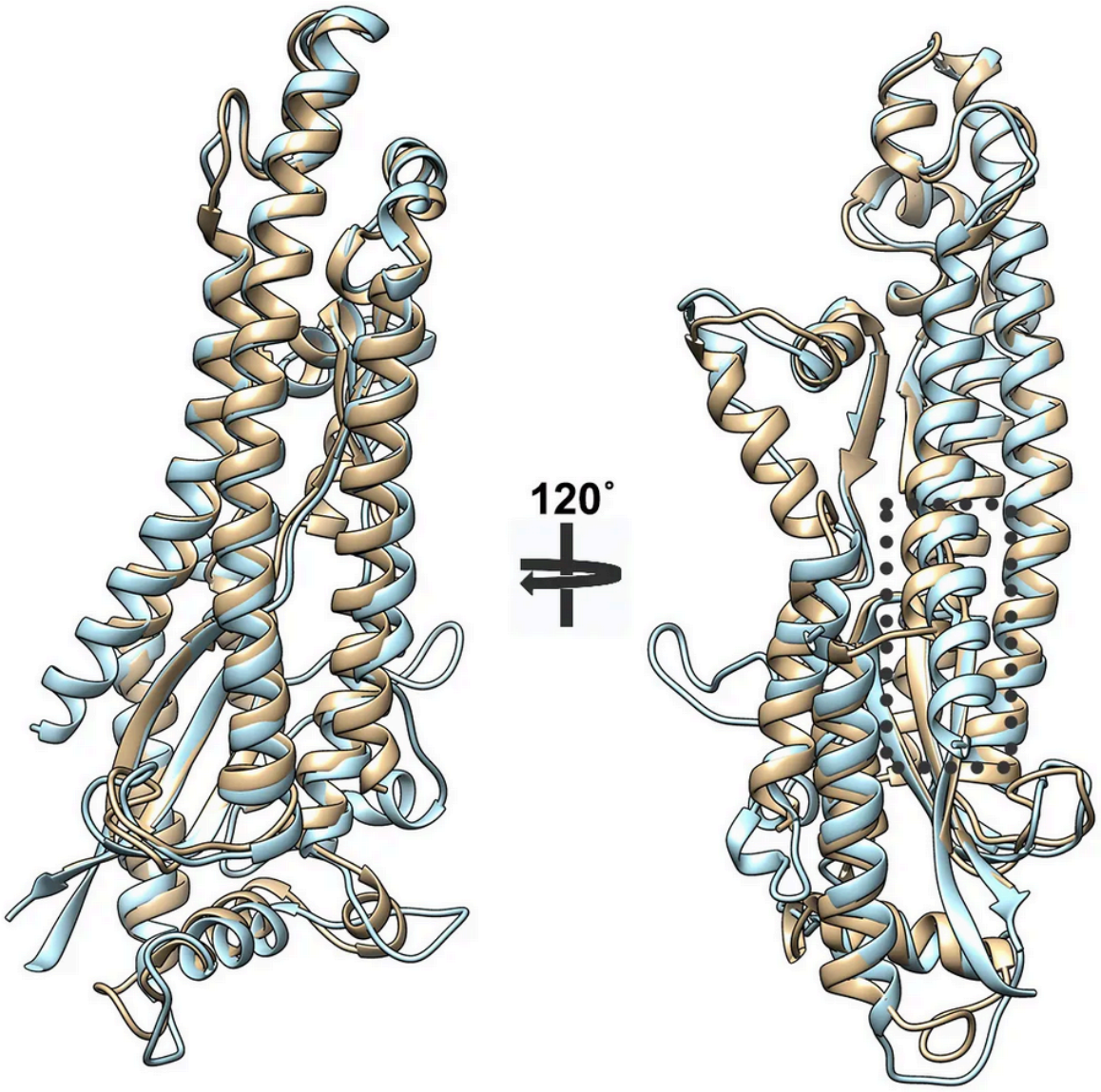
Tandem MS EThcD spectrum of a 3+ glycopeptide with HexNAc(2)Hex(6) attached to Asn 699 of the HCoV-NL63 S glycoprotein digested with chymotrypsin. The relative intensity normalized to the most intense ion is plotted against mass-to-charge ratio. The peaks under the horizontal dashed line are multiplied by 3 for visualization. "M" denotes the molecular ion. The charge state of the fragment ions is indicated in brackets. Ions relating to the glycopeptide, glycan and peptide fragments are colored red, green and blue, respectively. In this example of a glycopeptide identification the matched fragment ions define a large part of the peptide sequence and also provide detailed information about the glycan composition.

Figure II.XVIII: Architecture of the complete coronavirus fusion machinery.



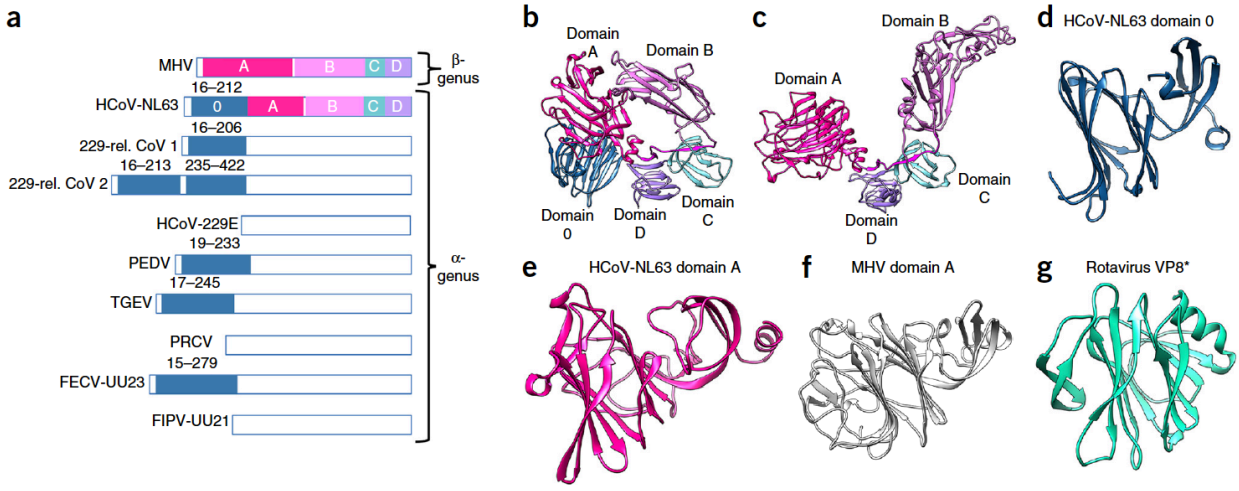
(a) Ribbon diagram of the S_2 trimer, colored by protomer with glycans rendered as dark-blue spheres. (**b**) Zoomed-in view of the S2′ _trigger-loop region comprising the central helix and the fusion peptide (light blue). N-linked glycans are shown as dark-blue spheres. The polypeptide segment corresponding to the putative cleavage site is poorly resolved in the density, and this part of the model should be considered to be hypothetical. (**c**, **d**) Ribbon diagrams showing two orthogonal views of the S2′ _C-terminal region, which is assembled from the connector domains and stem helices. (**e**, **f**) Ribbon diagrams of the HCoV-NL63 S_2 subunit (**e**) and of the RSV F protein (**f**). Conserved structural elements are colored identically to highlight the similar 3D organization of these two fusion machineries, whereas nonconserved regions are colored gray. The topology diagrams underscore the similar topology of the HCoV-NL63 S connector domain and the equivalent RSV F domain, although the tertiary structures of these domains are different, and several structural motifs have been added to the latter domain throughout evolution. The RSV F secondary-structural elements are annotated according to ref. 34. The N- and C-terminal extremities of the polypeptide segments are indicated.

Figure II.XIX: Structural similarity of coronavirus fusion machineries



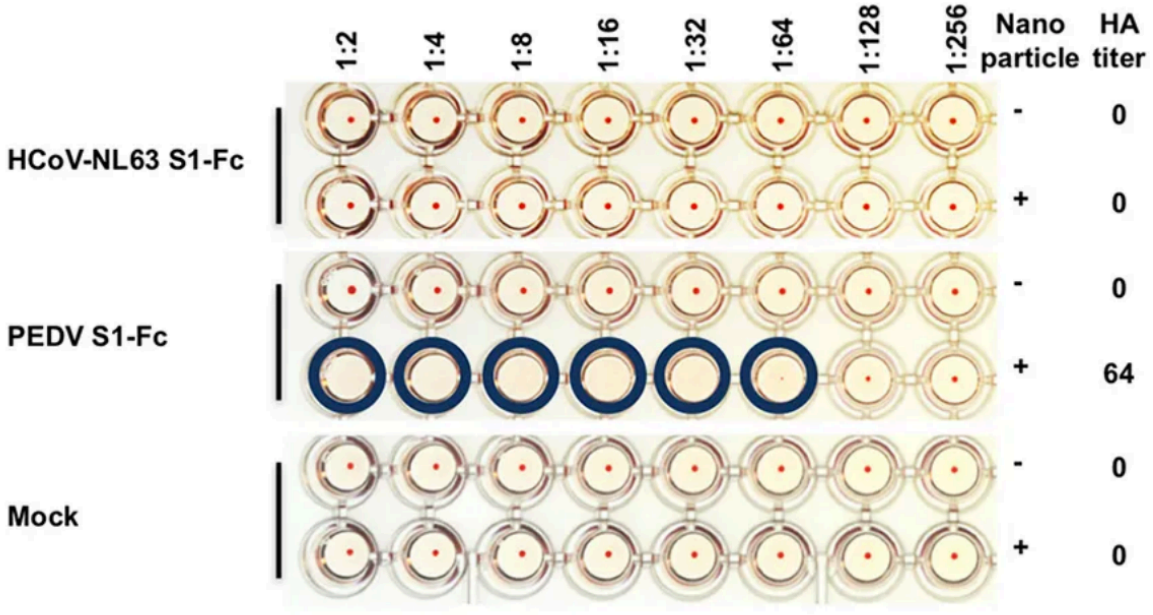
Ribbon diagram of the HCoV-NL63 (blue) and MHV (tan) S_2 fusion subunits. The dashed box highlights the two extra helical turns present in the S protein HR1 region of α-coronaviruses but not β-coronaviruses.

Figure II.XX: Evolution of the α-coronavirus S-glycoprotein fold appears to correlate with tissue tropism.



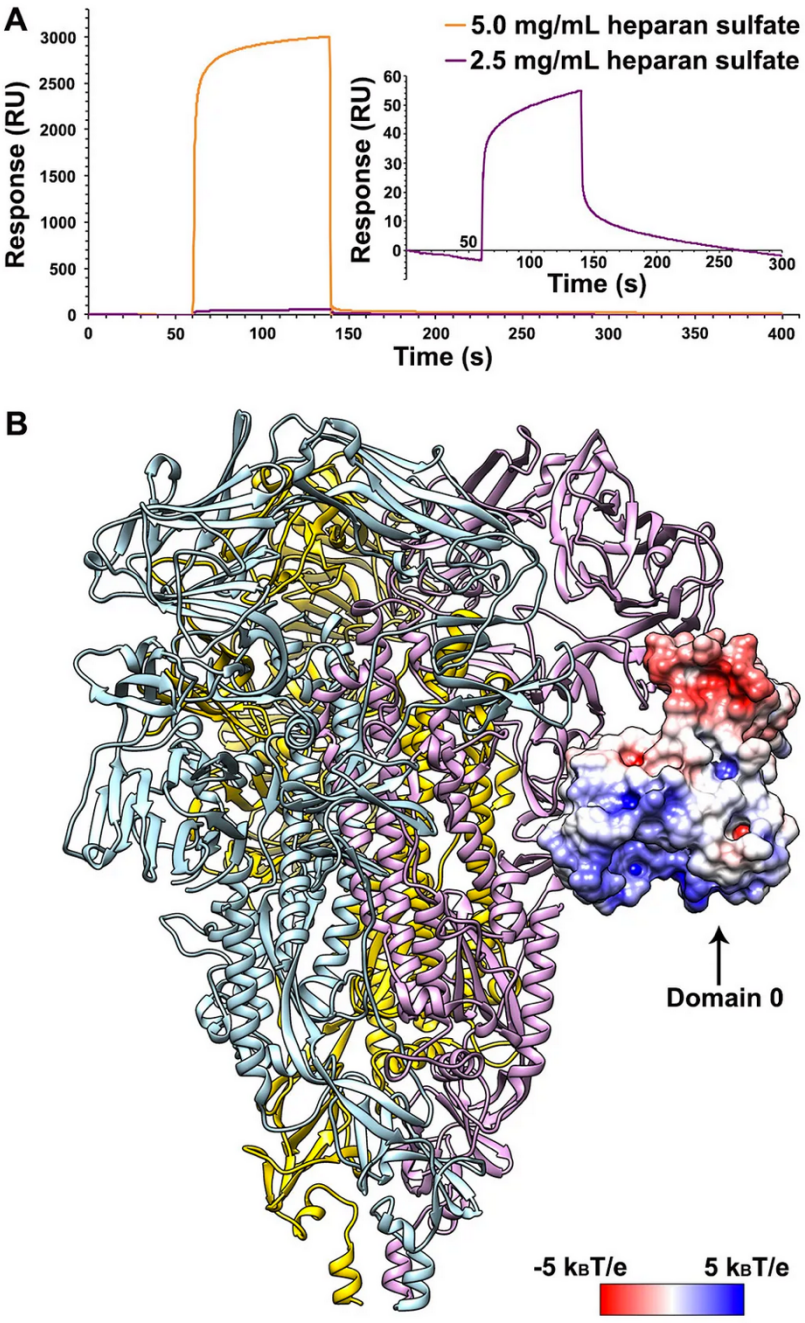
Schematic representation of several α-coronavirus S-glycoprotein S_1 subunits, highlighting the presence of one or several domains 0 (blue), as compared with β-coronaviruses. HCoV-NL63 (GenBank YP_003767.1), 229-rel. CoV 1 (GenBank ALK28775.1), 229-rel. CoV 2 (GenBank ALK28765.1), HCoV-229E (GenBank NP_073551.1), porcine epidemic diarrhea virus (PEDV; GenBank AAK38656.1), transmissible gastroenteritis virus strain Purdue P115 (TGEV; GenBank ABG89325.1), porcine respiratory coronavirus strain ISU-1 (PRCV; GenBank ABG89317.1), feline enteric coronavirus strain UU23 (FECV-UU23; GenBank ADC35472.1) and feline infectious peritonitis coronavirus strain UU21 (FIPV-UU21; GenBank ADL71466.1). The β-coronavirus MHV S_1 subunit is shown for comparison. Domains A–D are indicated for MHV and HCoV-NL63. (**b**) Ribbon diagram of the HCoV-NL63 S_1 subunit. (**c**) Ribbon diagram of the MHV S_1 subunit. (**d**–**g**) Ribbon diagrams of HCoV-NL63 domain 0 (**d**), domain A (**e**), MHV domain A (**f**) and rotavirus VP8* (**g**), showing their structural similarity, which suggests common ancestry. HCoV-NL63 domain 0 and A probably arose from a duplication event.

Figure II.XXI: The HCoV-NL63 S_1 subunit does not bind sialic acid



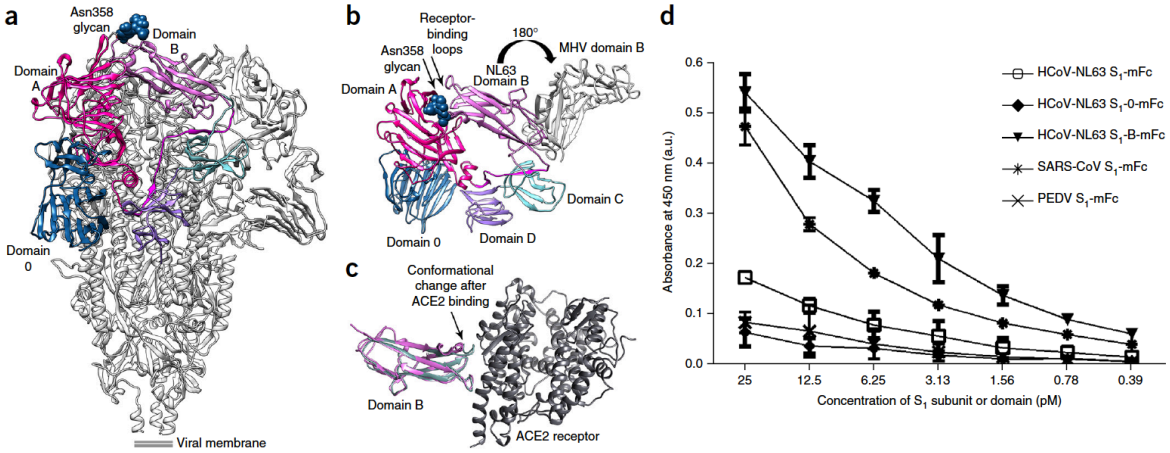
Binding of sialic acid by the HCoV-NL63 S_1 subunit (N-terminally fused to human IgG Fc) was assessed by probing the hemagglutination of human erythrocytes. The porcine epidemic diarrhea coronavirus S_1 subunit was used as a positive control. Mock indicates the absence of coronavirus S_1 subunit (negative control). The assays were performed using either free S_1-Fc or nanoparticle-displaying S_1-Fc to increase the avidity for sialic acid on the erythrocyte surface. Wells showing hemagglutination are circled.

Figure II.XXII: HCoV-NL63 S binds heparan sulfate



a,Surface plasmon resonance sensorgram showing binding of heparan sulfate to HCoV NL63 S. The right panel shows a blow-up view of the sensorgram corresponding to 2.5 mg/mL heparan sulfate. **b**, Ribbon diagram of the HCoV-NL63 S atomic model colored by protomer. Domain 0 is shown in surface representation colored according to its electrostatic surface potential for one protomer. The positively charged patch on its surface could putatively mediate binding to heparan sulfate.

Figure II.XXIII: Potential immune-evasion strategy used by HCoV-NL63.



(**a**) Ribbon diagram of the HCoV-NL63 S trimer, highlighting the conformation of the S_1 subunit. Domains 0, A, B, C and D are colored for one protomer. (**b**) The HCoV-NL63 receptor-binding loops are buried via interactions with domain A of the same protomer (including the glycan moiety at Asn358) and are not available to engage host-cell receptors. Superimposition of the HCoV-NL63 (purple) and MHV (light gray) S_1 subunits via their C domains highlights that their B domains feature opposite orientations related by an ~180° rotation, thus suggesting a putative trajectory for the conformational changes that must occur to engage the host-cell receptor. Only domain B is shown for MHV S. (**c**) Comparison of the HCoV-NL63 domain-B structure in our cryo-EM-derived model (purple) with the crystal structure of the same domain in complex with ACE2 (green and dark gray), showing that the receptor-binding loop containing residues 531–539 substantially changes its conformation after binding. (**d**) ACE2 binding ELISA showing that isolated HCoV-NL63 domain B (HCoV-NL63 S_1-B-mFc) binds ACE2 with higher affinity than does the full-length S_1 domain (HCoV-NL63 S_1-mFc). SARS-CoV S_1 (HCoV-NL63 S_1-mFc) is a positive control. HCoV-NL63 S_1 domain 0 (HCoV-NL63 S1-0-mFc) and PEDV S_1 (PEDV S_1-mFc), which do not bind ACE2, are negative controls. Mean values and s.d. of three independent experiments are shown.

Chapter III – Structure of Post-Fusion Coronavirus Spike Proteins

Prior to this dissertations work, there were no structures of coronavirus spike proteins in their post-fusion conformation, only crystal structures of the heptad repeats of various different CoVs. The S glycoprotein initiates infection through conformational changes that remain largely uncharacterized. Here we show the cryoEM structure of the MHV S glycoprotein in the postfusion state, showing large-scale secondary, tertiary, and quaternary rearrangements compared with the prefusion trimer and rationalizing the free-energy landscape of this conformational machine. Our data reveal that the postfusion S trimer adopts a 180-Å-long cone-shaped architecture arranged around a prominent central triple-helical bundle and is the longest structure observed for any class I fusion protein. It has later been noted that production of this protein is non-trivial, with most other groups purifying only rosettes or not well folded protein (communication, varied). We further characterized the refolding of the metastable prefusion S conformation using limited proteolysis, mass spectrometry, and single-particle EM. The observed similarity to paramyxovirus F structures demonstrates a conserved refolding trajectory and supports the evolutionary relatedness of their fusion subunits even with low sequence similarity. Finally, our data provide a structural framework for understanding antibody neutralization and for engineering vaccines against this medically important virus family. This structure laid the groundwork for identifying potential weak points of the virus that change from the pre to the post-fusion conformation and began to unravel the mechanism of the CoV fusion process.

Adapted from: Walls, A.C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B. J., Rey, F. A., Veesler, D. Tectonic conformational changes of a coronavirus spike glycoprotein

promote membrane fusion. Proceedings of the National Academy of Science. 2017. DOI: 10.1073/pnas.1708727114.

III.I Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion

The S glycoprotein exists as a metastable prefusion trimer at the viral surface, and its structure has recently been characterized (as of 2017) (Gui et al. 2017; Kirchdoerfer et al. 2016; Pallesen et al. 2017; Walls et al. 2016; Walls et al. 2016a; Yuan et al. 2017). Receptor binding and proteolytic processing promote large-scale conformational changes allowing initiation of the fusion reaction by insertion of the hydrophobic fusion peptide into the host membrane (Belouzard et al. 2009; Millet & Whittaker, 2014). The subsequent irreversible refolding of the fusion machinery provides the energy required to juxtapose the viral and host membranes, promoting fusion and delivery of the viral genome into the cytoplasm.

The only available structural information about the conformational changes undergone by coronavirus fusion machinery comes from X-ray crystallography studies of short polypeptide fragments spanning the heptad-repeat motifs (Duquerroy et al. 2005; Supekar et al. 2004; Xu et al. 2004). The data are limited to a small portion of the fusion machinery and do not reveal how most of the S_2 subunit refolds. A detailed knowledge of the conformational changes driving fusion is important to define the accessibility of epitopes targeted by neutralizing antibodies and to engineer improved subunit vaccine candidates, as was reported for the *Respiratory syncytial virus* (RSV) fusion (F) protein (McLellan et al. 2011; McLellan et al. 2013a; McLellan et al. 2013). Alternatively, heptad-repeat–mimicking peptides have been successfully used to inhibit type I fusion

machineries, including coronavirus S glycoproteins (Bosch et al. 2003). Furthering our understanding of the structural rearrangements underlying fusion bears the promise of developing next-generation inhibitors targeting this viral family.

Coronavirus S proteins harbor up to two protease cleavage sites located at the boundary between the S_1 and S_2 subunits (S_1/S_2 site) and upstream from the fusion peptide (S_2' site) (**Figure III.I**) (Belouzard et al. 2009; Millet & Whittaker, 2014). Cleavage at S_1/S_2 occurs upon biogenesis and viral egress for some coronaviruses, such as *Mouse hepatitis virus* (MHV) or MERS-CoV (Millet & Whittaker, 2015). This cleavage event, along with subsequent binding to the host receptor, is essential to promote cleavage at the S_2' site and fusion activation in the case of MERS-CoV (Millet & Whittaker, 2014). The critical importance of cleavage at the S_1/S_2 site is also exemplified by the *Bat coronavirus* (Bat-CoV) HKU4. Bat-CoV HKU4 shares a high degree of sequence similarity with MERS-CoV and can bind to the same human receptor (DPP4), although it is unable to infect human cells (Wang et al. 2014). Engineering two-point mutations in the Bat-CoV HKU4 S_1/S_2 region, which introduces two protease cleavage sites similar to the ones found in the MERS-CoV S sequence, is sufficient to allow efficient entry into human cells (Yang et al. 2015). These results demonstrate that both receptor and protease specificity are important determinants of host range. Proteolytic fusion activation at the S_2' site, which occurs for all coronaviruses, can take place in several cellular compartments (Millet & Whittaker, 2015). For instance, transmembrane protease/serine protease (TMPRSS) processing of SARS-CoV and MERS-CoV S at the cell membrane, furin-mediated processing of human coronavirus (HCoV)-NL63 and MERS-CoV S in the early endosomes, or lysosomal protease-mediated triggering of SARS-CoV S (cathepsin L) or

MHV S are key events that enable fusion activation and coronavirus entry into host cells (Bosch et al. 2008; Burkard et al. 2014; Millet & Whittaker, 2014; Millet & Whittaker, 2015). Several coronaviruses have redundancy built into their S protein sequence, enabling activation via multiple proteases and increasing the spectrum of cell types that can be infected (Millet & Whittaker, 2014) .

We incubated the prefusion uncleaved MHV, SARS-CoV, and MERS-CoV S glycoprotein ectodomains with trypsin, under limited-proteolysis conditions, before analysis by mass spectrometry and negative-staining EM. We identified that cleavage occurred at or near the S_1/S_2 and S_2' sites, which suggested that we recapitulated the proteolytic activation mechanism taking place in vivo (**Figure III.II**). We observed particles exhibiting the globular prefusion architecture and particles featuring an elongated cone-like structure reminiscent of the postfusion conformation described for paramyxovirus F proteins **(Figure III.I** *)* (Chen et al. 2001; McLellan et al. 2011; Swanson et al. 2011; Yin et al. 2005). These observations suggest that the three S ectodomains could refold to the postfusion conformation upon proteolytic activation and release of the S_1 subunit. We also detected the formation of rosettes, previously described for the SARS-CoV S protein (Li, F. et al. 2006), which are presumed to result from interactions of multiple postfusion trimers via their hydrophobic fusion peptides (**Figure III.II)**. In agreement with previous reports, our data show that β-coronavirus S proteins can undergo pH-independent conformational changes upon proteolytic cleavage to promote fusion of the viral and host membranes.

To understand the molecular basis of this transition, we used *Drosophila* S2 cells to produce only the MHV S_2 subunit ectodomain (residues 718–1,252) and to recapitulate

the expected large-scale conformational rearrangements of the fusion machinery (lacking the receptor-binding subunit) driving membrane fusion. This construct was fused to a GCN4 trimerization motif at the C-terminal end (Walls et al. 2017). Frozen-hydrated MHV S_2 trimers are 190-kDa cone-shaped particles resembling those shown in **Figure III.I**, which suggested that we captured a biologically relevant conformation and that the presence of the GCN4 motif was not sufficient to preserve the prefusion state (**Figure III.III)**. We determined a 3D reconstruction at 4.1-Å resolution and built and refined a model including residues 741–1,248, with an internal break between residues 808-971, using Coot(Brown et al. 2015), Rosetta de novo (Wang et al. 2015), RosettaES (Frenz et al. 2017), and Rosetta density-guided iterative refinement (DiMaio et al. 2015) (**Figure III.III, IV, V**).

The atomic model was validated using a multipronged approach including comparison with known fragment structures, analysis of disulfide bond positions, and mass spectrometry. First, available crystal structures of coronavirus S six-helix bundle fragments (Duquerroy et al. 2005; Supekar et al. 2004; Xu et al. 2004), spanning part of the heptad-repeat 1 (HR1) and 2 (HR2) motifs, could be superimposed with good agreement onto the refined structure (Figure III.IV). Also, the tertiary structure of the polypeptide chain corresponding to the upstream helix and the core β-sheet remain virtually identical to the prefusion structure, although the rest of the S_2 subunit undergoes refolding (Walls et al. 2016) (**Figure III.VI)**. Second, the reconstruction resolves several intrachain disulfide bonds (Cys775–Cys797, Cys780–Cys786, Cys1082–Cys1093, and Cys1132–Cys1177) between pairs of cysteine residues that are conserved in the MHV, HCoV-NL63, SARS-CoV, and HCoV-HKU1 S glycoproteins (Gui et al. 2017; Kirchdoerfer

et al. 2016; Walls et al. 2016; Walls et al. 2016a; Yuan et al. 2017) (**Figure III.VI**). As expected, these covalent links are maintained during the transition from the prefusion to the postfusion conformation. Finally, we identified several N-linked glycosylation sites using on-line reversed-phase liquid chromatography with electron transfer/high-energy collision-dissociation tandem mass-spectrometry (Frese et al. 2013). Some of these sites are resolved in the cryoEM map (Asn754, Asn1180, and Asn1190) with the correct sequence context, further validating the assigned register (**Figure III.VI, III.VII**).

Previous studies have shown that the HR1 and HR2 motifs of coronavirus S proteins assemble as a stable, six-helix bundle when expressed in isolation (Duquerroy et al. 2005; Supekar et al. 2004; Xu et al. 2004), a canonical feature of class I viral fusion proteins. Our structure of the MHV S_2 fusion machinery ectodomain resolves a long central triple helical bundle, assembled from the central helices and HR1 motifs, surrounded by three antiparallel HR2 helices forming a six-helix bundle at one end (**Figure III.IV).** At the opposite end exists a triangular pyramidal base comprised of the N terminus of the S_2 subunit and the connector domain (**Figure III.IV**). The HR1 motif of each protomer interacts exclusively with the other two protomers, forming an intertwined network (**Figure III.IV**). In line with studies on structural rearrangements of other class I fusion proteins, the antiparallel orientation of the HR2 helices relative to the central coiled coil is indicative of the postfusion conformation, as the fusion peptide and transmembrane regions are located at the same end of the cone-like structure, unlike in the prefusion structure (**Figure III.IV**). These conformational changes are believed to be crucial to mediate membrane fusion. Neither the fusion peptide region nor the adjacent N terminus of HR1 is resolved, suggesting a high degree of flexibility of this polypeptide segment in

the absence of a membrane. The conformation of the MHV S_2 trimer observed in this structure, along with the coronavirus S limited-proteolysis experiments, provides structural and biochemical evidence that the fusion machinery is maintained in its metastable prefusion conformation by the S_1 subunits, as removal of the latter results in spontaneous refolding to the more stable postfusion state (even in the presence of a GCN4 motif). Additionally, our data show that refolding can occur in the absence of proteolytic processing in the context of an isolated S_2 trimer ectodomain.

Transition of the MHV S_2 fusion machinery from the pre- to the postfusion state is accompanied by conformational changes of spectacular magnitude (**Figure III.VIII**). The prefusion S_2 trimer is short and wide (88 Å long and ~40 Å wide) (**Figure III.VIII).,** whereas the postfusion structure adopts an elongated cone-shaped conformation (185 Å long and 15–50 Å wide) (**Figure III.VIII).** The tertiary structure of the core β-sheet and of the upstream helix remains mostly unchanged in the two conformations, although their relative orientation is modified (**Figure III.VI**). Multiple disulfide bonds constrain S_2 around the core β-sheet and the upstream helix, which appear to act as scaffolds around which the conformational changes in the rest of the polypeptide chain take place (**Figure III.VI**). Based on the relatively high sequence conservation of the S_2 subunit among coronaviruses and the conservation of these disulfide bonds, we postulate that similar conformational changes take place in all coronaviruses during fusion.

One of the most noticeable rearrangements between the pre- and postfusion structures occurs at the level of the HR1 motif (residues 948–1,035). In the prefusion S trimer, HR1 consists of four α-helices connected by extended loops and runs in an antiparallel orientation relative to the central helix (Walls et al. 2016) (**Figure III.VIII**). In

the postfusion structure, the four HR1 α-helices and connecting loops refold and reorient to append to the N terminus of the central helix, leading to the formation of a single, continuous 130-residue-long α-helix (for which 113 residues are resolved in our reconstruction) (**Figure III.VIII**). This extremely long helix assembles as a homotrimeric helical bundle, stabilized partly via HR1-mediated coiled-coil interactions, and forms the core of the S_2 postfusion trimer (**Figure III.IV**). These conformational changes are accompanied by an alteration of the quaternary organization of the central helix. Although the fusion peptide region (Burkard et al. 2014; Madu et al. 2009) and the HR1 N-terminal moiety were resolved in the MHV S prefusion conformation (except for the S_2' loop), these residues are not visible in the postfusion map, suggesting that they are highly dynamic in the absence of the target membrane and transmembrane domain. Since this region is directly adjacent to HR1, which is completely refolded, it is expected that repositioning and potentially reorganization of the polypeptide chain occur for the fusion peptide region as well. As a result, we propose that targeting this polypeptide segment with antibodies or other types of inhibitors could prevent the insertion of the fusion peptide into the host membrane and putatively block refolding. The suitability of this approach is supported by in vivo neutralization data for MHV and SARS-CoV that identified this region as a major antigenic determinant, containing neutralization epitopes, upon infection by these viruses (Daniel et al. 1993; Zhang, H. et al. 2004) .

The C-terminal connector domain and stem helices are not resolved in the MHV S prefusion structure (Walls et al. 2016). These domains, however, are visible in the recently determined HCoV-NL63, SARS-CoV, and MERS-CoV S structures and can be used for comparison with the MHV S_2 postfusion structure (Gui et al. 2017; Pallesen et

al. 2017; Yuan et al. 2017). MHV S residues 1,120–1,210, which span the connector domain and stem helix, share 22%, 33%, and 36% sequence identity with the equivalent residues of the HCoV-NL63, SARS-CoV, and MERS-CoV S glycoproteins, respectively. Although the tertiary structures of the prefusion S and postfusion MHV S connector domains are similar, the latter is repositioned underneath the core β-sheet of a neighboring subunit due to the refolding of HR1 and the fusion peptide region (**Figure III.IX**). The most C-terminal part of the postfusion connector domain undergoes extensive refolding to adopt a domain-swapped organization with interactions formed exclusively with neighboring protomers. It contributes a fourth β-strand (residues 1,178–1,184) to the three-stranded core β-sheet of a different protomer and folds as an extended segment interacting with the central coiled coil in an antiparallel orientation (**Figure III.VIII, III.X).** The N terminus of the stem helix is refolded into an extended loop, whereas the previously unresolved C-terminally adjacent region folds as a short helix docked near the upstream helix, perpendicular to the central coiled coil (**Figure III.VIII, III.X**). In the context of the trimer, these two motifs form a pseudo-triangular scaffold wrapping around the central triple helical bundle, potentially stabilizing it in the postfusion conformation. Since the region corresponding to HR2 has not been observed in any prefusion S structure, the structural rearrangements taking place in this segment are unknown. However, it is likely that the C terminus of the polypeptide chain is oriented in the opposite direction to connect to the viral membrane before fusion occurs.

As for all other known class I viral fusion proteins, the postfusion state of coronavirus S glycoproteins represents the lowest-energy point of the conformational landscape (Ingallinella et al. 2004). This is supported by the spontaneous assembly of the isolated

heptad repeats of all characterized coronavirus S proteins as six-helix bundles recapitulating the postfusion state (Duquerroy et al. 2005; Supekar et al. 2004; Xu et al. 2004). The free energy released upon refolding is believed to promote merging of the viral and host membranes and subsequent fusion. Strikingly, the postfusion MHV fusion machinery buries 6,150 $Å^2$ of surface area per protomer at the interface with other subunits of the trimer through a large contribution from domain swapping of the connector domain C-terminal part and HR2. This observation provides a molecular basis explaining the unusually high stability of the postfusion S_2 conformation and represents the ground-state of the fusion reaction (Ingallinella et al. 2004) .

Previous genetic and biochemical studies have identified amino acid substitutions in the MHV S_2 subunit affecting fusogenicity, which can be analyzed based on the structural data reported here (Krueger et al. 2001; Taguchi & Matsuyama, 2002). The L1062F mutation introduces a bulkier amino acid side chain that would negatively affect packing of the central helix trimer and/or the interactions formed by the central helix with the adjacent upstream helix. Although modeling this mutation in the MHV S pre- and postfusion structures suggests comparable outcomes, the fact the upstream helix region appears to act as a pseudo-rigid scaffold around which refolding of the fusion machinery takes place may make this mutation more destabilizing for the postfusion conformation (**Figure III.XI**). The A994V substitution in the HR1 region introduces a bulkier amino acid side chain that could alter the packing of the postfusion S_2 central coiled coil as well as potentially disturb the formation of the six-helix bundle with the HR2 C-terminal helices, the latter being a major contributor of the extensive surface area buried between protomers, and might explain its destabilizing effect on the postfusion conformation

(**Figure III.XI**). In summary, it appears that substitutions at these two strictly conserved positions would be more destabilizing for the postfusion conformation, which is the ground-state for all known class I fusion proteins, than for the prefusion state, thereby explaining the reported reduction of S-mediated membrane fusion.

We previously reported that the prefusion architectures of coronavirus S and paramyxovirus F proteins share a similar topology, which suggests they use conserved mechanisms for mediating fusion of the viral and host membranes and that they may have arisen from a common ancestral gene (McLellan et al. 2013a; Walls et al. 2016; Walls et al. 2016a; Yin et al. 2006). Comparison of the F protein structures of RSV (McLellan et al. 2013a), *Parainfluenza virus 3* (PIV3) (Yin et al. 2006), or Newcastle disease virus (Chen et al. 2001) with MHV S_2 also shows that their tertiary structures are related in the postfusion conformation (**Figure III.IX, III.X**). Although large-scale conformational changes occur during fusion, the core β-sheet and upstream helix remain mostly unchanged in the prefusion and postfusion S and F states (**Figure III.X**). CoV S_2 features a 30-turn helix, resulting from refolding of the four α-helices and connecting loops spanning the HR1 motif in the prefusion S state and appending to the central helix. Refolding of the multiple secondary structural motifs spanning the HR1 sequence into a single helix is the signature of class I fusion proteins and is reminiscent of what has been reported for *Influenza virus* hemagglutinin (Bullough et al. 1994), HIV envelope (Buzon et al. 2010), and *Ebola virus* GP (Weissenhorn et al.1998) in addition to paramyxovirus F proteins (McLellan et al. 2011; McLellan et al. 2013; Swanson et al. 2011; Yin et al. 2005; Yin et al. 2006) .

The quaternary reorganization of the postfusion coronavirus S_2 and paramyxovirus F glycoproteins results in the formation of similar cone-shaped rods tapering toward the end proximal to the merged membranes (Figure III.VII). The coronavirus fusion machinery, however, is at least 10 Å longer than paramyxovirus F glycoproteins (without considering the unresolved S HR1 N-terminal region) and is the longest structure observed for any class I fusion protein. Both fusion subunits are characterized by the formation of a triple helical-bundle core, resulting from the association of the extended central/HR1 helices, which participates in the formation of a six-helix bundle with the HR2 C-terminal helices. The polypeptide segment corresponding to the C-terminal region of the connector domain and HR2 is also largely refolded in coronavirus S_2 and paramyxovirus F glycoproteins in which they are observed extending along the periphery of each neighboring protomer (Chen et al. 2001; Gui et al. 2017; McLellan et al. 2011; Yin et al. 2005). Of note, the core β-sheet of each fusion subunit is supplemented by a β-strand contributed in an identical manner by the connector domain of a neighboring subunit (**Figure III.X**).

These observations reinforce the similarity between coronavirus S and other characterized class I viral fusion proteins, especially paramyxovirus F proteins, as they share a conserved refolding trajectory to promote membrane fusion. Refolding of coronavirus S and paramyxovirus F likely proceeds through a similar zippering reaction during which the C-terminal moiety enhances its interactions with the central triple-helical bundle to bring the transmembrane region of the fusion machinery in proximity to the fusion peptide (**Figure III.XII**).

We previously postulated that the S_1 subunits make interactions with the S_2 subunits that are key to stabilizing the prefusion conformation of coronavirus S glycoproteins (Walls et al. 2016). Similar observations were also made based on the HCoV–HKU1 prefusion S structure (Kirchdoerfer et al. 2016). The results of our limited-proteolysis experiments and the near-atomic-resolution reconstruction of the MHV S_2 trimer further corroborate this hypothesis. Proteolytic processing at the S_1/S_2 and S_2' cleavage sites removes the covalent linkage between the two functional subunits and frees the fusion peptide, which allows shedding of the S_1 crown and subsequent refolding of the fusion machinery (Figure III.IX). Inspection of coronavirus prefusion structures show that the HR1 region of each protomer contacts neighboring protomers via the B and C domains (Gui et al. 2017; Kirchdoerfer et al. 2016; Walls et al. 2016; Walls et al. 2016a). These interactions act as a molecular clamp, stabilizing the metastable prefusion trimer until receptor binding and cleavage triggering at the S_2' site. Similar mechanisms of stabilization have been described for influenza hemagglutinin in which the HA_1 globular head contacts the HR1 polypeptide segment to prevent early refolding. Subsequent exposure to the acidic environment of the endosomes promotes refolding of HA_2 and membrane fusion (Bullough et al. 1994). Structurally related paramyxovirus F proteins exist in a metastable prefusion state in the viral membrane. Upon receptor engagement, conformational changes in the receptor-binding protein H, HN, or G on the surface of measles virus, PIV5, or *Nipah virus*, respectively, promote F-triggering and subsequent membrane fusion (Liu et al. 2013). These similarities between influenza HA, paramyxovirus F, coronavirus S, and other class I viral-fusion proteins suggest that

comparable mechanisms have evolved to ensure proper spatial and temporal coordination of their fusion proteins and productive infection.

The tertiary structure of the fusion machinery is highly conserved among coronaviruses across multiple genera, in agreement with the relatively high sequence conservation of the S_2 subunit. Based on these observations, along with the presence of several conserved disulfide bonds constraining the fusion machinery, we propose that the tectonic conformational changes reported here for MHV S_2 are representative of the rearrangements taking place in all coronavirus S proteins during fusion. Most vaccine-design initiatives aim at targeting the prefusion state of viral fusion proteins, which correspond to the conformation that could be detected by the immune system before infection. Comparison of the prefusion S trimer structures with the MHV postfusion structure reported here provides a blueprint to analyze the accessibility of neutralization epitopes in each conformation. Although most known neutralization epitopes characterized to date are present in the S_1 subunit, due to its higher immunogenicity than the S_2 subunit, the marked sequence and structural diversity of S_1 has so far led to the elicitation of species-specific antibodies (Pascal et al. 2015; Rockx et al. 2008). In contrast, the higher conservation of the S_2 fusion machinery bears the promise that epitopes could potentially be targeted by broadly neutralizing antibodies cross-reacting with multiple coronaviruses. For instance, the fusion peptide region is highly conserved among coronavirus S glycoproteins and has been identified as a major antigenic determinant upon infection by MHV and SARS-CoV (Daniel et al. 1993; Zhang et al. 2004). The postfusion MHV S_2 structure supports our previous hypothesis that antibodies binding to this site could hinder insertion of the fusion peptide into the target membrane

and putatively prevent fusogenic conformational changes. The stem helix and part of the HR2 motif are also conserved among S glycoproteins and therefore represent attractive targets to elicit broadly neutralizing antibodies. In the prefusion state, the yet unresolved HR2 C terminus of the fusion machinery would likely point in the opposite direction, compared with the postfusion conformation, corresponding to a 180° reorientation of the polypeptide chain with potential changes in secondary, tertiary, and quaternary structures. The large-scale nature of the predicted conformational changes in HR2 reinforces the idea that antibodies targeting this region will be neutralizing, in agreement with previous reports (Elshabrawy et al. 2012; Routledge et al. 1991). For example, the 10G monoclonal antibody is known to inhibit MHV S-mediated cell-to-cell fusion and to block virus infectivity (Routledge et al. 1991) upon binding to an epitope comprised within residues 1,212–1,226 that fold as an extended loop upstream of the HR2 C-terminal helix in the postfusion structure. Finally, the data reported here provide a structural framework for designing coronavirus S-based subunit vaccine immunogens. Knowledge of the precise structural rearrangements taking place during the fusion reaction paves the way for rationally engineering modifications enhancing the stability of the prefusion trimer. Designed disulfide bonds cross-linking residues that are close to each other in the prefusion state but far apart in the postfusion conformation and/or cavity-filling mutations are strategies that have been successfully used to enhance the stability of the prefusion RSV F trimer and could also be implemented for S glycoproteins (McLellan et al. 2013). Finally, the recent stabilization of the MERS-CoV prefusion S via introduction of proline residues at the junction between the central helix and the HR1 motif can also be rationalized with our data (Pallesen et al. 2017) .

III.II Methods for characterization of a post-fusion coronavirus spike protein

CryoEM data were collected on an FEI Titan Krios microscope operated at 300 kV and equipped with a Gatan Quantum GIF energy filter and a Gatan K2 Summit direct electron detector. Frame alignment was carried out with dose weighting using MotionCor2 (Li et al. 2013), and Relion 2.0 was used for downstream processing and 3D reconstruction (Scheres, 2012) .

A human codon-optimized gene encoding for the MHV S glycoprotein (UniProt accession no. P11224) was synthesized, and the fragments encoding the MHV S ectodomain (residues 15–1,252) or the MHV S_2 ectodomain (residues 718–1,252) were PCR amplified and ligated to a gene fragment encoding a GCN4 trimerization motif (Eckert et al. 1998; Walls et al. 2017a), a thrombin cleavage site, an eight-residue-long Strep-tag (IBA), and a stop codon. A human-codon–optimized gene encoding for the MERS-CoV S glycoprotein (UniProt accession no. W6A028) with the R748K substitution was synthesized, and the fragment encoding the S ectodomain (residues 19–1,262) was ligated to a gene fragment encoding a foldon trimerization motif, a thrombin cleavage site, an eight-residue-long Strep-tag, and a stop codon. These genes were cloned into the pMT\BiP\V5\His expression vector (Invitrogen) in frame with the *Drosophila* BiP secretion signal downstream of the metallothionein promoter.

A human codon-optimized gene encoding for the ectodomain (residues 14–1,180) of the SARS-CoV S protein (UniProt accession no. P59594) was cloned into a modified pOPING vector (Addgene) (Berrow et al. 2007) introducing an N-terminal Mu-phosphatase signal peptide and a C-terminal *Tobacco etch virus* (TEV) protease cleavage site, a foldon, and a hexa-histidine tag at the C terminus of the construct.

To generate a stable *Drosophila* S2 cell line expressing the recombinant MHV-S_2 or MERS S ectodomain, we used Effectene (Qiagen) and 2 μg of plasmid. Puromycin *N*-acetyltransferase was cotransfected and used as a dominant selectable marker. Stable MHV-S_2– or MERS S-expressing cell lines were selected by the addition of 7 μg/mL puromycin (Invivogen) to the culture medium 48 h after transfection. For large-scale production, the cells were cultured in spinner flasks and induced by 5 μM of $CdCl_2$ at a density of ~10^7 cells/mL. After 1 week at 28 °C, clarified cell supernatants were concentrated 40-fold with Vivaflow tangential filtration cassettes (10-kDa cutoff) (Sartorius) and adjusted to pH 8.0 before affinity purification with a Strep-Tactin Superflow column (IBA) followed by gel-filtration chromatography with a Superose 6 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris·HCl (pH 7.5) and 100 mM NaCl or, in the case of MERS S, 20 mM sodium phosphate (pH 7.0) and 250 mM NaCl. The purified protein was concentrated to ~3 mg/mL.

Transient transfection of 250 mL of HEK293F cells at a density of 10^6 cells/mL was performed using 293fectin (Thermo Fisher) and Opti-MEM (Thermo Fisher). After 3 d the cells were harvested before affinity purification with a Talon 5-mL cobalt column equilibrated in 25 mM sodium phosphate (pH 8.0), 300 mM NaCl, 5 mM imidazole, and 0.5 mM PMSF. The purified protein was buffer exchanged into 20 mM Tris (pH 8.0), 150 mM NaCl and was concentrated to 1.0 mg/mL.

Five microliters of 1 mg/mL protein were incubated with either 1:100 or 1:10 trypsin (Sigma Aldrich) at room temperature for 1 h. This reaction was then used for negative-staining EM (3 μL of ~0.001 mg/mL).

Three microliters of trypsin-treated CoV S proteins were incubated for 30 s on carbon-coated copper mesh grids before staining with 2% uranyl formate. Micrographs were acquired using the Leginon software (Suloway et al. 2005) on a 120-kV FEI Tecnai G2 Spirit transmission electron microscope with a Gatan UltraScan 4000 4k × 4k CCD camera at 52,000 or 67,000 nominal magnification (pixel size of 2.07 Å or 1.62 Å at the specimen level). The defocus ranged from 1.0 to 2.0 µm. Particles were picked automatically in a reference-free manner using DoGpicker (Voss et al. 2009) and were extracted using a box size of 320 Å. Contrast transfer function (CTF) estimation was performed using GCTF (Zhang, 2016). 2D classifications were performed using RELION 2.0 and included correction for the effect of the microscope CTF. Each displayed class contains ~1,000–5,000 particles.

Two microliters of purified MHV S_2 spike at 0.6 mg/mL in 0.025% amphipole (Anatrace) were double blotted (Snijder et al. 2017) onto a 1.2/1.3 C-flat grid (Protochips), which had been glow-discharged for 30 s at 20 mA. Grids were then plunge-frozen in liquid ethane with an FEI Mark I Vitrobot with a 13-s blot time and an offset of −1 mm at 100% humidity and 25 °C. Data were collected with Leginon automatic data-collection software on an FEI Titan Krios electron microscope operated at 300 kV and equipped with a Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV, and a Gatan K2 Summit direct electron detector camera (Li et al. 2013). The dose rate was adjusted to eight counts per pixel per second, and each movie was acquired in counting mode fractionated in 50 frames of 200 ms. In a single session, 2,200 micrographs were collected with a defocus range between 1.5 and 3.0 µm.

Movie frame alignment was carried out with dose weighting using MotionCorr2 (Li et al. 2013). The parameters of the microscope CTF were estimated with GCTF (Zhang, 2016). Particles were manually selected to generate a template. The template was low pass filtered to 25 Å and used to pick particles on all micrographs with FindEM (Roseman, 2004) integrated in Appion (Lander et al. 2009). Particle images were extracted and processed with Relion 2.0 (Scheres, 2012) with a box size of 320 × 320 pixels and a pixel size of 1.36 Å. Following two rounds of 2D classification, 150,000 of 900,000 particles remained to run 3D classification with C3 symmetry. Then 106,000 particles were used to run gold-standard 3D refinement with Relion using the solvent_correct_fsc_flag and a solvent mask with a soft edge of five pixels and a cosine-edge falloff of five pixels. Postprocessing was performed with Relion to apply a B-factor of −280 $Å^2$. Reported resolutions were based on the gold-standard criterion [Fourier shell correlation (FSC) = 0.143], and FSC curves were corrected for the effects of soft masking by high-resolution noise substitution. The soft mask used for FSC calculation had a 10-pixel cosine edge falloff.

University of California, San Francisco (UCSF) Chimera (Goddard et al. 2007) and Coot (Brown et al. 2015) were used to fit and rebuild atomic models into the cryoEM map. The central helix, core β-sheet, and upstream helix of the prefusion MHV S structure were fit into the S_2 map and were both manually adjusted and extended with Rosetta ES (Frenz et al. 2017). The connector domain was built using the HCoV-NL63 (Walls et al. 2016a) and SARS-CoV (Gui et al. 2017) prefusion connector domains. The final model was refined by application of strict noncrystallographic symmetry constraints with Rosetta (DiMaio et al. 2013), with a training map corresponding to one of the two maps generated

by the gold-standard refinement procedure in Relion. The second map (the testing map) was used only for calculation of the FSC compared with the atomic model and preventing overfitting (DiMaio et al. 2013). The quality of the final model was analyzed with MolProbity (Chen et al. 2010). All figures were generated with UCSF Chimera (Goddard et al. 2007) or UCSF ChimeraX (Goddard et al. 2018) .

MHV-S after limited proteolysis with trypsin was separated by SDS/PAGE and cut from gel for digestion with AspN. The gel bands were cut to small pieces, washed in Milli-Q water, and treated with acetonitrile to shrink the gel pieces. The sample was then incubated in 1,4-DTT (1 g/L) for 60 min at 60 °C, treated with acetonitrile, alkylated with iodoacetamide (10 g/L) for 30 min at room temperature in the dark, and subsequently washed with ammonium bicarbonate and treated with acetonitrile, twice. The gel pieces were then incubated with 30 mg/L AspN for 90 min on ice. Excess AspN was removed, the gel pieces were covered in ammonium bicarbonate, and the samples were subsequently incubated overnight at 37 °C. The digested samples were collected, and remaining sample was extracted from the gel pieces by treatment with acetonitrile. The solution with the peptides was subsequently dried in a SpeedVac, and the peptides were resuspended in 10% formic acid, 5% dimethyl sulfoxide in water.

For analysis of N-linked glycosylation, an estimated 250 pmol of MHV S_2 was denatured, reduced, and alkylated by dilution to 5 µM in 50 µL of buffer containing 100 mM Tris (pH 8.5), 10 mM Tris(2-carboxyethyl)phosphine (TCEP), 40 mM iodoacetamide, and 2% (wt/vol) sodium deoxycholate. Samples were first heated to 95 °C for 10 min and then incubated for an additional 20–30 min at room temperature in the dark. The samples were split in two for digestion with trypsin and chymotrypsin (Sigma Aldrich) in parallel,

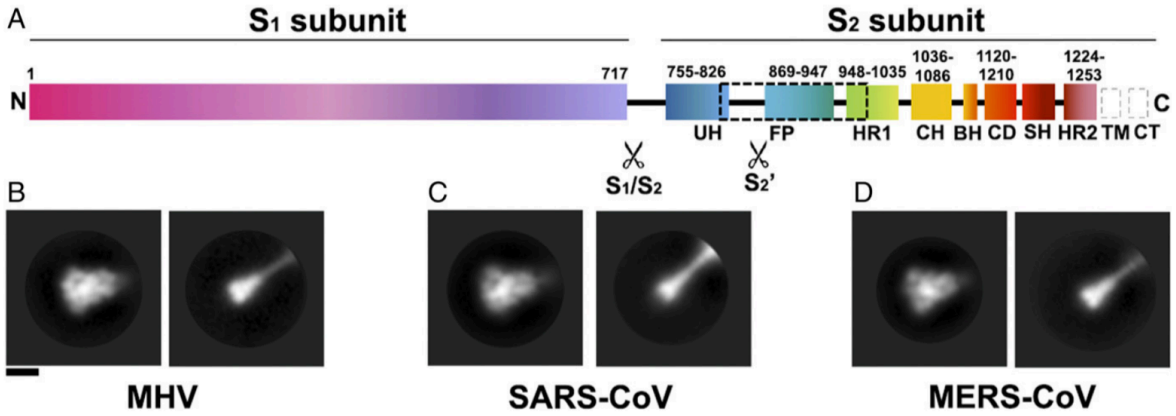
by diluting 20 µL of sample for each protease in a total volume of 100 µL 50 mM ammonium bicarbonate (pH 8.5). Protease was added to the samples in a ratio of 1:75 by weight and left to incubate at 37 °C overnight. After digestion, 2 µL of formic acid was added to the samples to precipitate the sodium deoxycholate from solution. After centrifugation at 17,000 × *g* for 20 min, 80 µL of the supernatant was collected.

For each sample 8 µL was injected on a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer. A 35-cm analytical column and a 3-cm trap column filled with ReproSil-Pur C18AQ 5 µm (Dr. Maisch) beads were used. Nanospray LC-MS/MS was used to separate peptides over a 110-min gradient from 5 to 30% acetonitrile with 0.1% formic acid. A positive spray voltage of 2,100 was used with an ion transfer tube temperature of 350 °C. An electron-transfer/higher-energy collision dissociation ion-fragmentation scheme (Frese et al. 2013) was used with calibrated charge-dependent entity-type definition (ETD) parameters and supplemental higher-energy collision dissociation energy of 0.15. A resolution setting of 120,000 with an AGC target of 2×10^5 was used for MS1, and a resolution setting of 30,000 with an AGC target of 1×10^5 was used for MS2. Data were searched with Protein Metrics Byonic software (Bern et al. 2012), using a small custom database of recombinant protein sequences including the proteases used to prepare the glycopeptides. Reverse decoy sequences were also included in the search. Specificity of the search was set to C-terminal cleavage at R/K (trypsin) or F/W/Y/M/L (chymotrypsin), allowing up to two missed cleavages, with EThcD fragmentation (b/y- and c/z-type ions). We used a precursor mass and product mass tolerance of 12 ppm and 24 ppm, respectively. Carbamidomethylation of cysteines was set as fixed modification, methionine oxidation as variable modification, and all four

software-provided N-linked glycan databases were used to identify glycopeptides. All glycopeptide hits were manually inspected, and only those with quality peptide sequence information are reported here.

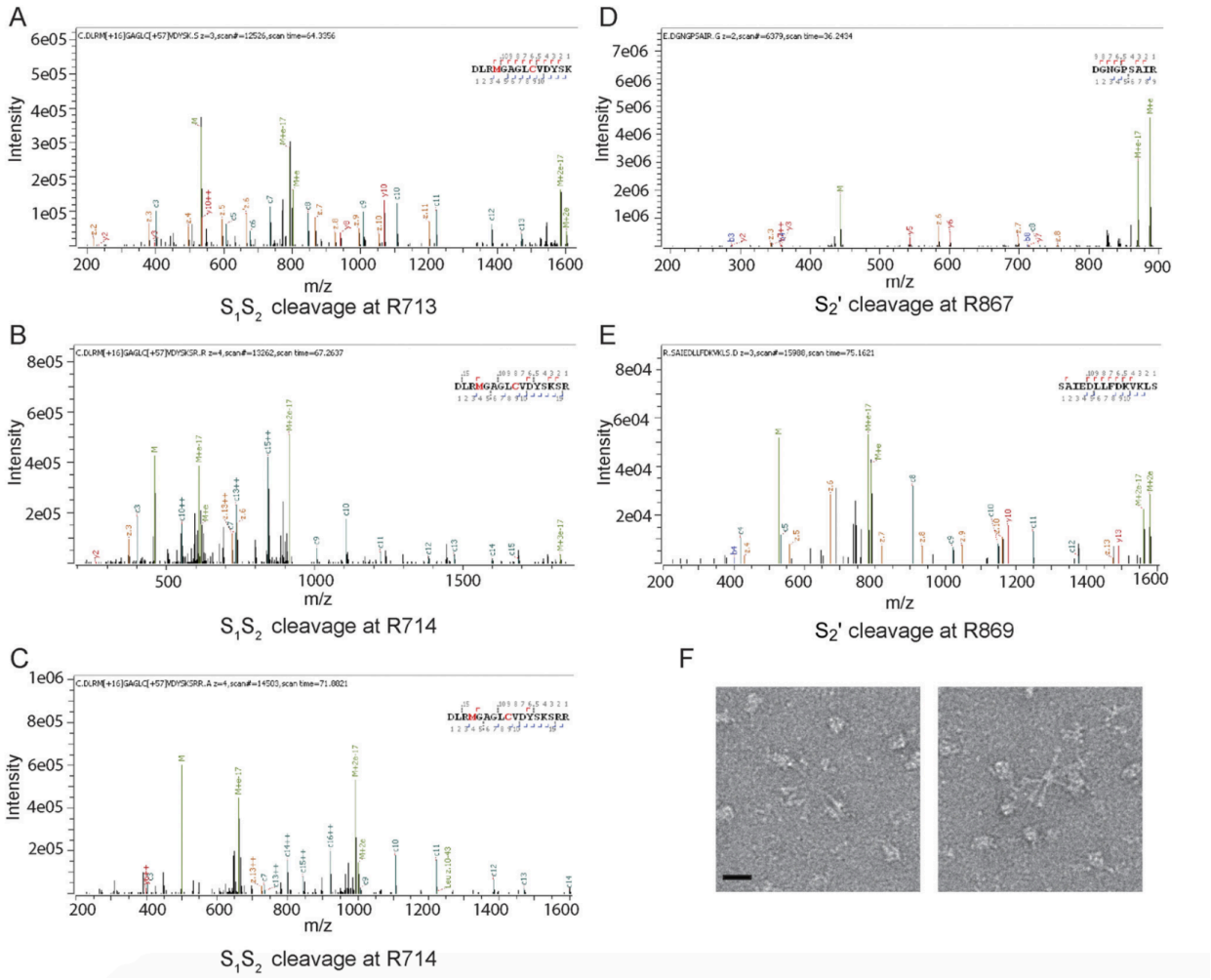
Figures Section III: III.I-III.XII: Adapted from Walls et al. 2017

Figure III.I: Proteolytic activation of coronavirus S proteins.



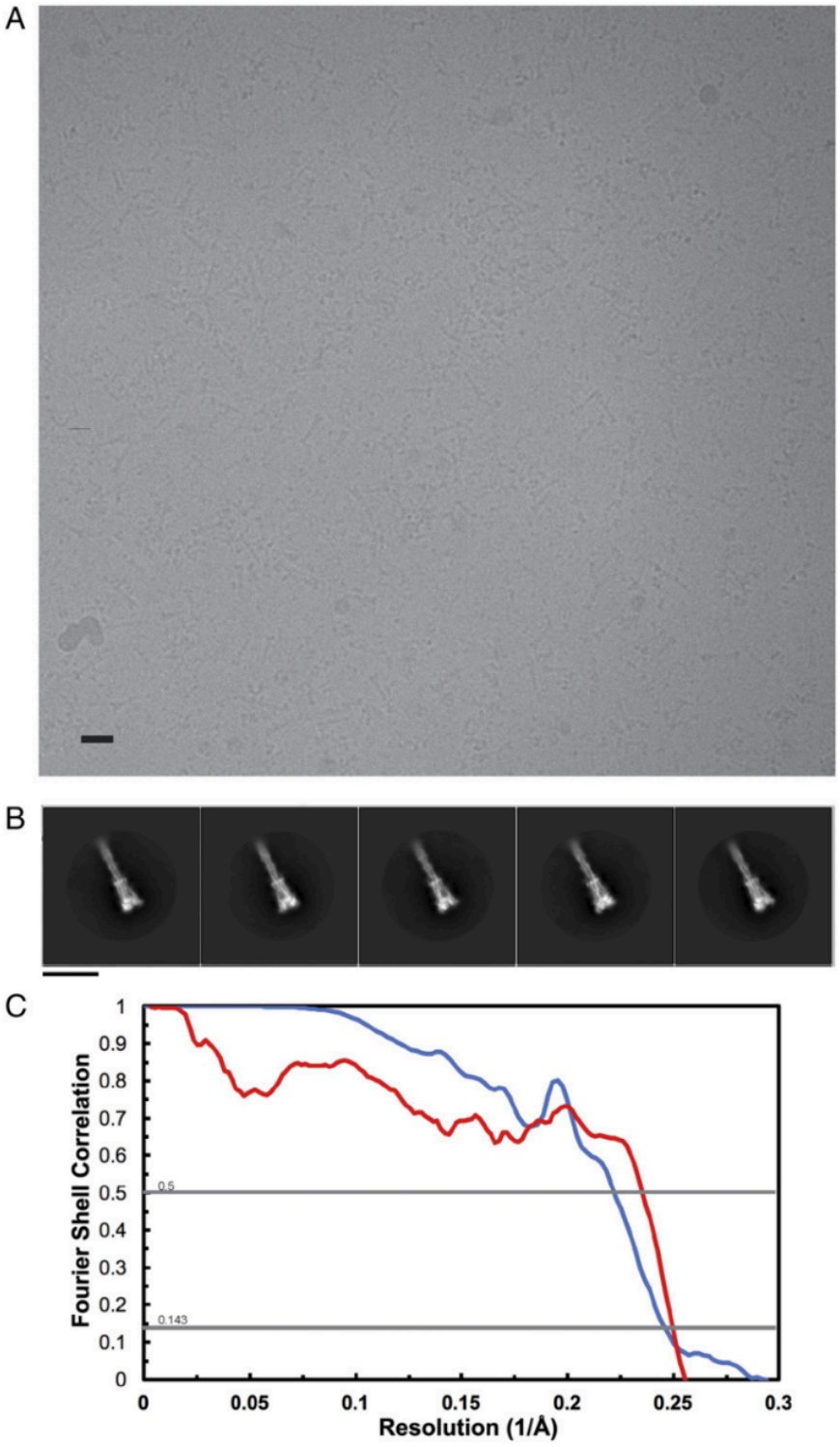
(A) Schematic of the MHV S glycoprotein organization with emphasis on the S2subunit. The dashed black box shows the region of the S2polypeptide chain that is unresolved in the map. Gray dashed boxes show regions that were not part of the construct. BH,beta hairpin; CD, connector domain; CH, central helix; CT, cytoplasmic tail; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; SH, stem helix; TM,transmembrane domain; UH, upstream helix. (B–D) 2D class averages of negatively stained MHV S (B), SARS-CoV S (C), and MERS-CoV S (D) trimers in the prefusion state (Left) and in the trypsin-cleaved postfusion state (Right). (Scale bar, 10 nm.)

Figure III.II: Characterization of the MHV S pre- to postfusion transition.



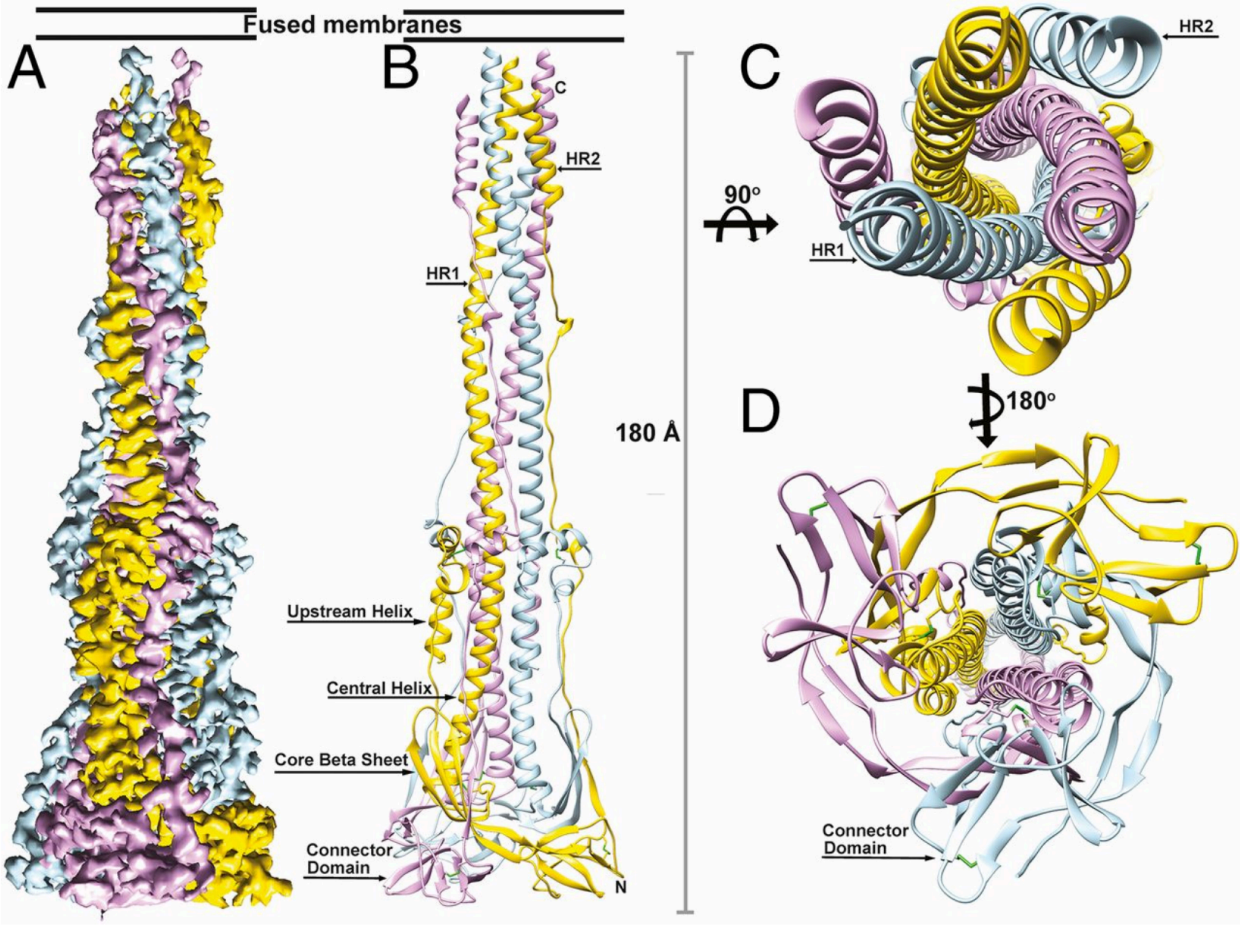
(A–E) Annotated mass spectra showing that trypsin cleavage can occur at or near the predicted S_1/S_2 (A–C) and S_2' (D and E) sites. (F) Section of a micrograph with negatively stained MHV S2rosettes formed upon incubation of prefusion S with 1:100 trypsin. (Scale bar, 10 nm.)

Figure III.III: CryoEM analysis of MHV S_2



(A) Representative micrograph. (Scale bar, 10 nm.) (B) 2D class averages of frozen-hydrated MHV S2particles. (Scale bar,10 nm.) (C) Gold-standard (blue) and model/map (red) Fourier shell correlation curves. The 0.143 and 0.5 cutoff values are indicated by horizontal gray lines. The resolution was determined to be 4.1 Å.

Figure III.IV: CryoEM structure of the MHV S_2 postfusion machinery

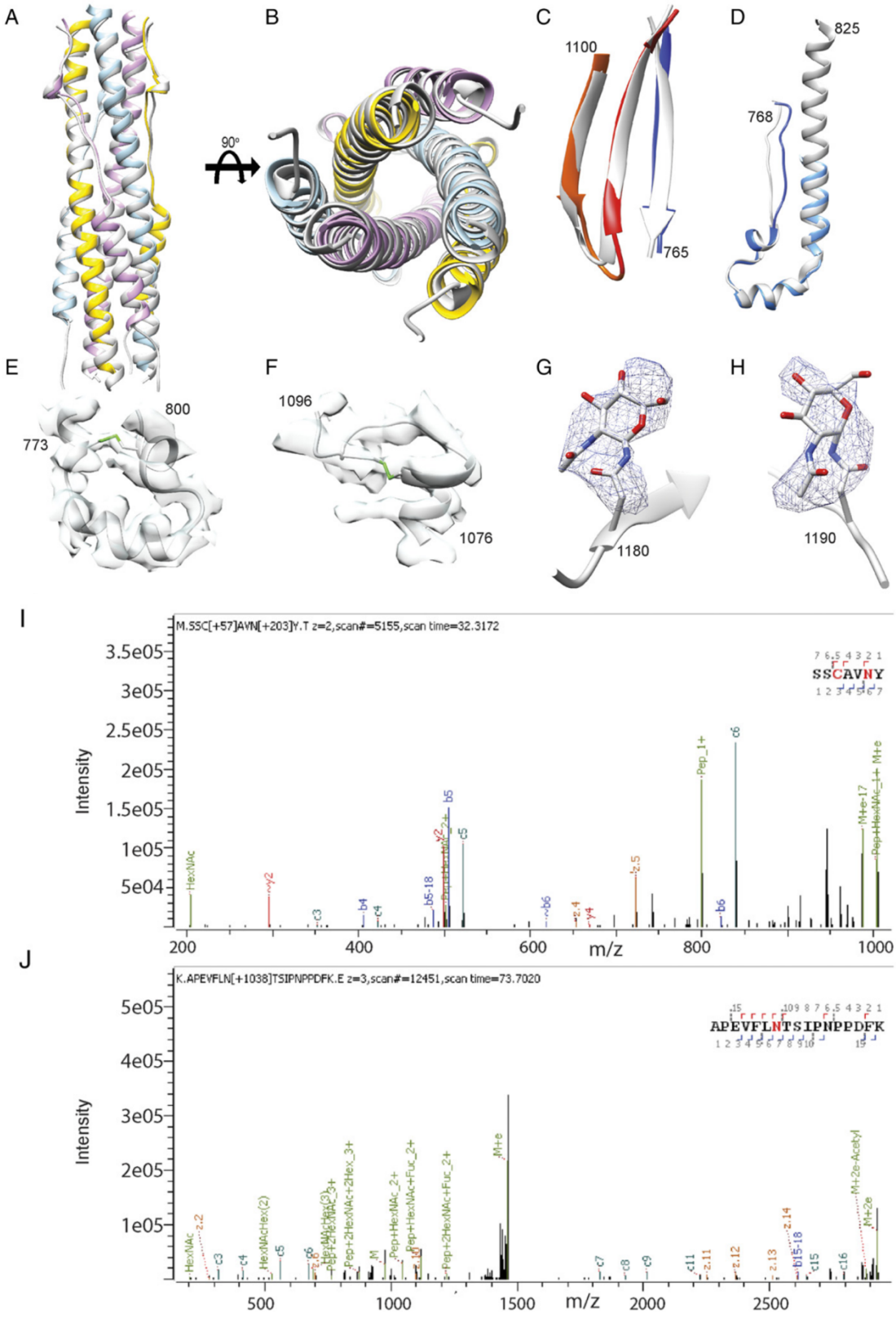


(A) 3D map colored by protomer and viewed from the side with the fused membranes located at the top. (B) Ribbon diagram of the MHV S2atomic model oriented as in (A). (C and D) Ribbon diagram showing the atomic model from the extremity proximal (C) or distal (D) to the fused membranes.

Figure III.V: Data collection, refinement, and model validation statistics

Parameter	Value
Data Collection	
Number of Particles	106,272
Pixel size, Å	1.36
Defocus range, μm	1.5-3.0
Voltage, kV	300
Electron Dose, $e^-/Å^2$	48
Refinement	
Resolution, Å	4.1
Map-sharpening B factor, $Å^2$	-280
Model validation	
Favored rotamers, %	99.32
Poor rotamers, %	0.34
Ramachandran favored, %	97.65
Ramachandran outliers, %	0.29
Clash score	1.77
MolProbity score	1.01

Figure III.VI: Validation of the atomic model

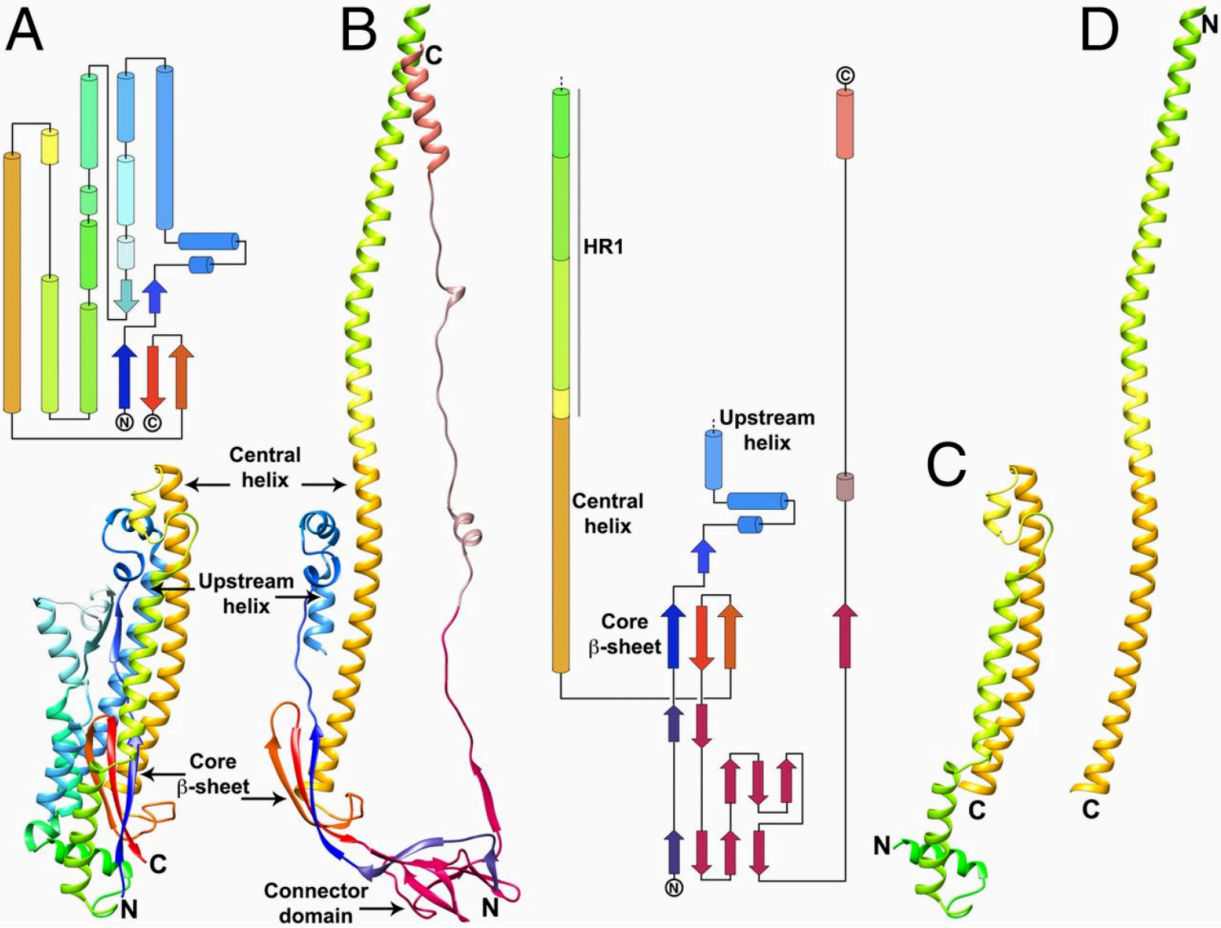


(A and B) Two orthogonal views of the superimposition of the HR1/HR2 six-helix bundle structure derived from the MHV S2cryoEM map (colored by protomer) and from X-ray crystallography (Protein Data Bank ID code 1WDF, gray). (C) Comparison of the coreβ-sheet structure in the prefusion (gray) and post-fusion (colored as in Fig. 2) MHV S spike glycoprotein. (D) Comparison of the upstream helix structure in the prefusion (gray) and post-fusion (colored as in Fig. 2) MHV S spike glycoprotein. (E and F) The MHV S2model is shown in gray ribbon with the cryoEM density shown in transparent light gray. Disulfide bonds are shown in green for residues 775–797 (E) and 1,082–1,093 (F). (G and H) Glycans linked to Asn-1180 (G) and Asn-1190(H) are rendered as ball and sticks with the corresponding region of the cryoEM map shown as a blue mesh. Carbon, nitrogen, and oxygen atoms are colored gray, blue, and red, respectively. (I and J) Mass spectra for Asn1180 (I) and Asn1190 (J) showing the presence of glycans at these glycosylation sequons.

Figure III.VII: Identification of MHV S_2 N-linked glycosylation sites using glycoproteomics and CryoEM

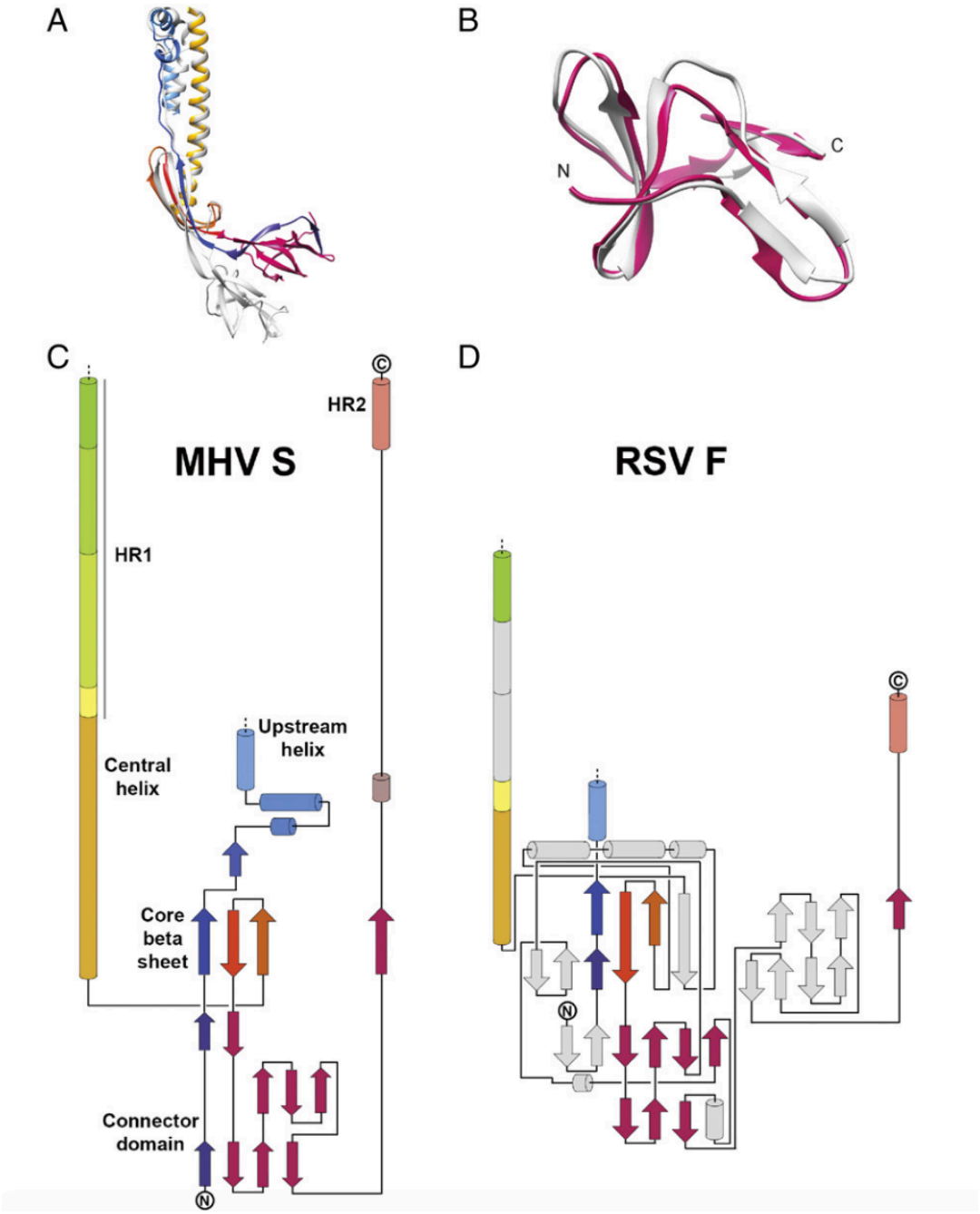
Site	Sequon	Combined	Intact	EndoH and EndoF treated	Un-glycosylated	CryoEM map
737	NDSV	737	737	737	ND	ND
754	NFTI	754	754	754	ND	754
844	NFSP	ND	ND	ND	844	ND
893	NCTG	893	893	893	ND	ND
1,126	NVSP	ND	ND	ND	1,126	ND
1,180	NYTK	1,180	1,180	1,180	1,180	1,180
1,190	NTSI	1,190	ND	1,190	1,190	1,190
1,209	NQTS	ND	ND	ND	ND	ND
1,225	NVTL	ND	ND	ND	ND	ND
Total number of sites	8	5	4	5	4	3

Figure III.VIII: Conformational changes association with the fusion reaction



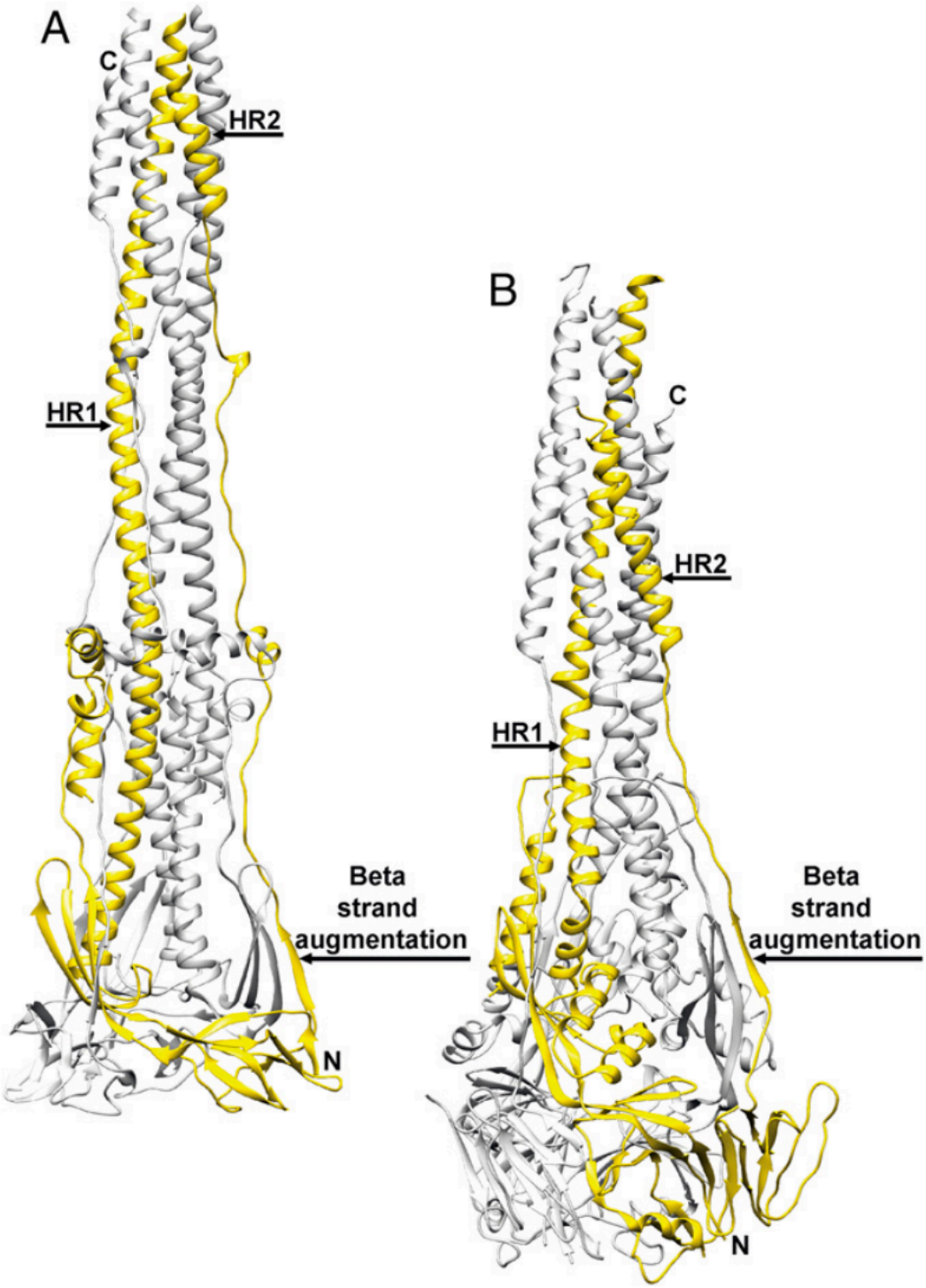
(A) Ribbon and topology diagrams of the MHV S2subunit in the prefusion conformation (6). (B) Ribbon and topology diagrams of the MHV S2subunitin the postfusion conformation. (C and D) Ribbon rendering of the MHV S central helix and HR1 in the prefusion (C) and postfusion (D) states highlighting the jack-knife refolding of the four HR1 helices and intervening regions into a single continuous helix.

Figure III.IX: Comparison of the MHV S postfusion S_2 structure with other fusion proteins.



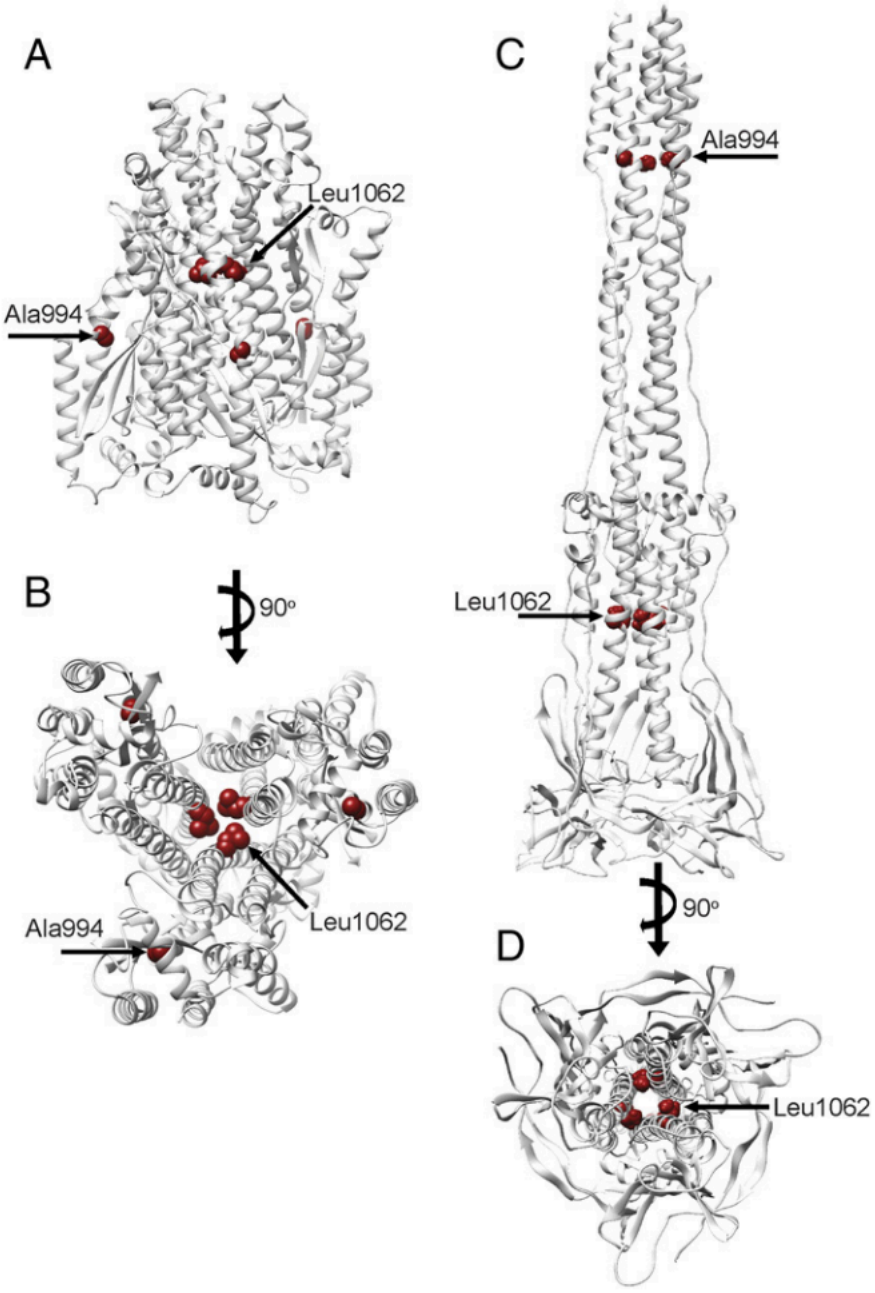
(A) Superimposition of the HCoV-NL63 S2(gray) and MHV postfusionS2(colored as in Fig. 2) subunits. Only the upstream and central helices, coreβ-sheet, and connector domain are shown. (B) Superimposition of the HCoV-NL63S2(gray) and MHV postfusion S2(pink) connector domains emphasizing their structural conservation. (C and D) Topology diagrams comparing postfusion MHV S2and RSV F glycoproteins. Conserved elements in the RSV F subunit are shown using the same colors; non conserved elements are shown in gray.

Figure III.X: Comparison of the coronavirus and paramyxovirus fusion machineries.



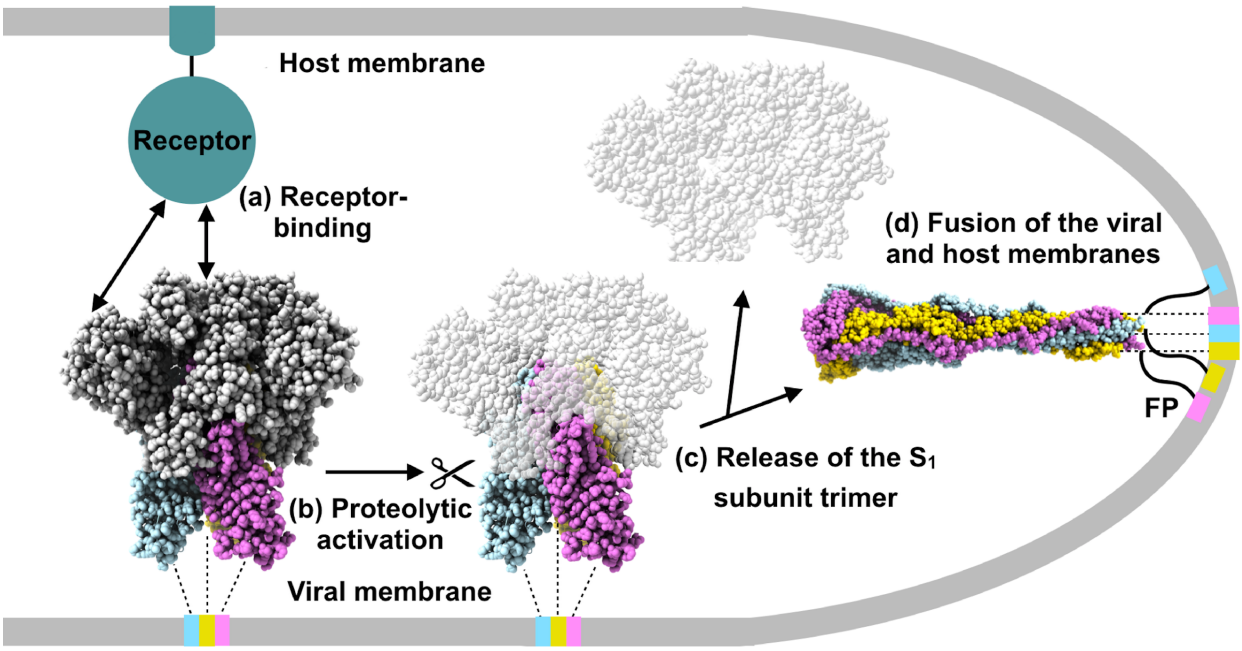
(A and B) Postfusion structures of MHV S (A) and RSV F (B) with one protomer of each trimer colored yellow and the other two-colored gray.

Figure III.XI: Location of fusion-inhibiting mutations in MHV S.



(A and B) Prefusion MHV S_2 subunit shown in gray ribbon from the side (A) and from the viral membrane (B). (C and D) Postfusion MHV S_2 shown in gray ribbon from the side (C) and looking toward the long axis (D). Two residues known to attenuate fusogenicity upon substitution are shown as red spheres

Figure III.XII: Proposed model of coronavirus entry



(A) The S glycoprotein pro-motes virus attachment to a host cell via binding to a transmembrane receptor using either domain A (e.g., MHV S) or domain B (e.g., SARS-CoV or MERS-CoV S). The prefusion MHV S trimer is shown with the S_1 subunitd epicted in gray and the S2subunits colored by protomer. (B) Upon receptor binding, activation of the S trimer occurs via protease cleavage at the S_2' site. (C) Shedding of the S_1 subunit trimer frees the fusion machinery, as reported for MERS-CoV (10). (D) Subsequent conformational changes of the S glycoprotein result in fusion of the viral and host membranes. The postfusion MHV S_2 trimer is depicted with each protomer in a different color. The transmembrane helices and the fusion peptides (FP) are connected to the MHV S trimer with dotted and solid lines, respectively

Chapter IV – Coronavirus Spike proteins in Complex with Neutralizing Antibodies

Proteolytic processing and receptor-binding act in synergy to induce large-scale S conformational changes promoting coronavirus entry. Priming involves S cleavage by host proteases at the boundary between the S_1 and S_2 subunits (S_1/S_2), in a subset of coronaviruses, and at a conserved site upstream from the fusion peptide (S_2') in all known coronaviruses (Belouzard et al. 2009; Burkard et al. 2014; Millet and Whittaker, 2014; Park et al. 2016). The latter site is believed to be the most important for membrane fusion activation. The SARS-CoV and MERS-CoV S receptor-binding domain, designated domain B, exhibits multiple conformational states that modulate the accessibility of the receptor-binding motifs (RBM) and in turn the ability to interact with host cells. Based on these findings, it was hypothesized that receptor-binding may initiate membrane fusion (Gui et al. 2017; Pallesen et al. 2017; Yuan et al. 2017). However, the S glycoprotein of other coronaviruses adopt a closed domain B conformation, incompatible with receptor engagement, indicating that structural rearrangements prior to receptor binding are required for entry (Kirchdoerfer et al. 2016; Shang et al. 2018; Shang et al. 2017; Walls et al. 2016;Walls et al. 2016a; Xiong et al. 2017). Due to this unusual mechanism of proteolytic activation and RBM conformational masking, our understanding of coronavirus membrane fusion activation remains limited.

Since the S glycoprotein densely decorates the viral surface and is a key player in pathogenesis, it is the main target of neutralizing antibodies and the focus of intense interest for vaccine design. We previously showed that coronavirus S glycoproteins are decorated with an extensive glycan shield comprising up to one hundred N-linked oligosaccharides (Walls et al. 2016a; Xiong et al. 2017). Molecular-level information of

the carbohydrates attached to the surface of highly pathogenic coronaviruses, however, is lacking. These glycans contribute to S stability during biogenesis, since inhibition of glycosylation by tunicamycin was reported to yield virions lacking S glycoproteins (Rossen et al. 1998), and likely participate in immune evasion via epitope masking (Du et al. 2016; Walls et al. 2016a; Xiong et al. 2018). Studies of the S glycoproteins of MERS-CoV and the bat-specific HKU4 coronavirus suggested that glycans also impact zoonosis by modulating cleavage site accessibility to proteases for membrane fusion activation (Yang et al. 2015). The outcome of the arms race between viral evolution mechanisms and the immune system of infected individuals can also lead to the elicitation of antibodies binding glycan-containing epitopes, such as in the case of HIV-1 (Scharf et al. 2015; Stewart-Jones et al. 2016) or Epstein-Barr virus (Snijder et al. 2018). These findings emphasize the necessity to obtain a detailed understanding of the carbohydrates covering coronavirus S glycoproteins to accelerate the development of subunit vaccines and therapeutics.

We previously reported the isolation of highly potent monoclonal neutralizing antibodies from rare memory B-cells obtained from SARS-CoV and MERS-CoV survivors (Corti et al. 2015; Traggiai et al. 2004). To understand neutralization, we report here an analysis of the fine molecular structure of the SARS-CoV and MERS-CoV S glycan shields and the cryo-electron microscopy (cryoEM) study of these glycoproteins in complex with two potent human neutralizing antibodies. Although both antibodies block receptor-interaction, the anti-SARS-CoV antibody acts by functionally mimicking receptor-attachment and promoting S fusogenic conformational rearrangements through a ratcheting mechanism. We found that the dynamics of the receptor-binding domain alter

the apparent binding affinity of coronavirus S trimers for their cognate receptors, a phenomenon that cannot be detected using isolated B domains. The findings presented here elucidate the unique nature of the coronavirus membrane fusion activation pathway. These results provide a structural framework for understanding coronavirus neutralization by human antibodies and shed light on activation of coronavirus membrane fusion, which takes place through a receptor-driven ratcheting mechanism.

Adapted from: Walls, A.C.*, Xiong, X.*, Park, Y.J., Tortorici, M.A., Snijder, J., Quispe, J., Cameroni, E., Gopal, R., Dai, M., Lanzavecchia, A., Zambon, M., Rey, F.A., Corti, D., Veesler, D. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. Cell. 2019.

IV.I Glycan shields, structures in complex with human antibodies, and membrane fusion activation mechanism

We set out to obtain a blueprint of individual N-linked carbohydrates attached to the surface of the MERS-CoV and SARS-CoV S glycoproteins produced in HEK293F cells. We used the stabilized MERS-CoV S 2P and SARS-CoV S 2P ectodomains, which harbor a double proline substitution between the central helix and the heptad-repeat 1 motif to enhance the stability of the prefusion conformation (Pallesen et al. 2017). We determined that oligosaccharides account for a molecular weight of 121 kDa and 98 kDa for MERS-CoV and SARS-CoV S, respectively **(Figure IV.I)**, using multi-angle light scattering, refractometry and UV spectrophotometry (Veesler et al. 2009). These results suggest that ~25% of the molecular weight of a S ectodomain corresponds to oligosaccharides whereas 50% of the HIV-1 Mosaic M Env molecular weight is accounted

for by glycans (Nkolola et al. 2014). We detected N-linked glycosylations at 22 and 16 unique sites for the MERS-CoV and SARS-CoV S protomers, respectively, using liquid chromatography coupled with electron transfer/high-energy collision dissociation tandem mass spectrometry (LC-MS/MS) (Frese et al. 2013). The data revealed an extraordinary heterogeneity of carbohydrates decorating the two S trimers **(Figure IV.II, IV.III, IV.IV)**. Individual sequons were linked to 21 unique glycans on average and up to 64 distinct oligosaccharides for the most heterogeneous sites. The intact glycopeptides were a variety of oligomannose, hybrid and complex glycans, with or without sialylation and (core) fucosylation. Most of the observed diversity corresponded to hybrid/complex carbohydrates **(Figure IV.I, IV.II, IV.III, IV.IV)**. The detection of glycans with different extent of processing on coronavirus S glycoproteins is reminiscent of the composite nature of the HIV Env (Struwe et al. 2018) or Lassa virus GP (Watanabe et al. 2018) glycan shields. The large number and combinatorial diversity of N-linked glycans covering the surface of MERS-CoV and SARS-CoV S could represent a challenge to overcome for antigen recognition.

Previous observations of SARS-CoV budding directly from the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) and/or from swollen Golgi sacs led to the hypothesis that coronaviruses might only be exposed to a subset of the enzymes of the glycosylation machinery (Ng et al. 2003; Stertz et al. 2007). To address this question, we used LC-MS/MS to characterize in parallel S oligosaccharides on the MERS-CoV England1 and Erasmus Medical Center (EMC)/2012 isolates produced using African green monkey VeroE6 cells, as well as on the purified MERS-CoV S 2P ectodomain trimer recombinantly produced using HEK293F cells. We observed extensive

overlap of the types of glycans detected across virion-derived and ectodomain-derived MERS-CoV samples, including the presence of hybrid and complex glycans **(Figures IV.III, IV.IV, IV.V, IV.VI)**. Our data agree with previous work detecting hybrid/complex glycans N-linked to S present in the membrane of authentic SARS-CoV virions (Krokhin et al. 2003; Ritchie et al. 2010), which is consistent with our analysis of the SARS-CoV S 2P ectodomain trimer. Our data indicate that at least a fraction of the MERS-CoV and SARS-CoV virions produced in a cell are exposed to the glycan-processing enzymes residing in the Golgi apparatus during assembly and budding.

The asymmetric cryoEM reconstructions of MERS-CoV S and SARS-CoV S in complex with human antibodies presented in the following sections respectively resolve glycan densities at 17 and 19 sites, virtually all of them overlapping with the MS/MS data **(Figure IV.II)**. At least the two core N-acetyl glucosamine moieties are resolved for the majority of glycosylation sites which extend from the surface of both the S_1 and S_2 subunits. Although both glycoproteins are densely decorated with N-linked glycans, by comparison the surface of the HIV Env trimer is obstructed by twice as many glycans per accessible surface area. MERS-CoV S and SARS-CoV S trimers share that the regions surrounding the S_1/S_2 cleavage site and the fusion peptide (near the S_2' cleavage site) are more sparsely glycosylated than the rest of each trimer **(Figure IV.II)**. These glycan holes could be important for providing access to activating host proteases and for allowing membrane fusion to take place unimpeded at the onset of infection (Walls et al. 2016a; Yang et al. 2015). We suggest the identified glycan holes could be exploited for epitope-focused immunogen design or therapeutic intervention against coronaviruses. This is supported by the presence of neutralizing antibodies targeting the fusion peptide region

in SARS-CoV survivor sera (Zhang et al. 2004), the high sequence conservation of this region among coronavirus S glycoproteins (Walls et al. 2016, Walls et al. 2016a) and the identification of a neutralization epitope within a comparable breach of the HIV-1 Env glycan shield (McCoy et al. 2016).

The LCA60 antibody was isolated from memory B cells of a MERS-CoV infected individual and potently neutralized multiple MERS-CoV isolates (i.e. England1, EMC/2012 and Jordan-N3/2012) (Corti et al. 2015). CryoEM analysis of the stabilized MERS-CoV S 2P glycoprotein ectodomain in complex with the LCA60 Fab fragment showed full saturation with one Fab bound to each B domain of the homotrimeric S. 3D classification revealed that the complex adopts two distinct conformational states **(Figure IV.VII, IV.VIII, IV.IX)** corresponding to two B domains closed and one B domain open (state 1, ~50%) as well as one B domain closed and two B domains open (state 2, ~50%). We determined asymmetric cryoEM reconstructions at 3.5 Å and 3.6 Å resolution for state 1 and 2, respectively **(Figure IV.VII, IV.VIII, IV.IX)**, as well as a crystal structure of the isolated LCA60 Fab at 3.0 Å resolution **(Figure IV.X).** LCA60 Fabs bound to open B domains are weakly resolved in the maps, suggesting marked conformational heterogeneity, whereas the quality of the density is enhanced for closed LCA60/B domains despite radial disorder **(Figure IV.VIII)**. Analysis of the S/LCA60 contacts was done using the state 1 structure.

LCA60 CDRH2, CDRH3, CDRL1 and CDRL3 interact with the MERS-CoV B domain **(Figure IV.XI)**, which also mediates binding to human dipeptidyl-peptidase 4 (DPP4), the MERS-CoV receptor (Raj et al. 2013). The binding interface of LCA60 and MERS-CoV S involves protein/protein and protein/glycan contacts. The 17-residue long

CDRH3 contacts the β8 strand and the preceding loop within the RBM. Specifically, residue $W108_{LCA60}$ inserts between the B domain residues $P531_{MERS\text{-}CoV}$ and $Y541_{MERS\text{-}CoV}$, on one side, and the glycan N-linked to the A domain residue $N166_{MERS\text{-}CoV}$ from a neighboring protomer, on the other side **(Figures IV.XI, IV.XII)**. CDRH2 also contacts the loop preceding strand β8 via electrostatic interactions. CDRL3 contacts the MERS-CoV strands β6, β9 and the loop preceding strand β8 whereas CDRL1 interacts with the loops preceding strands β6 and β8 **(Figure IV.XI)**. Domain A glycans at positions $N166_{MERS\text{-}CoV}$ and $N236_{MERS\text{-}CoV}$ appear to contact the Fab bound to the B domain of a neighboring protomer **(Figures IV.XI, IV.XII)**. The glycan N-linked to $N487_{MERS\text{-}CoV}$ of one protomer extends toward the Fab CDRH2 bound to the closed B domain of the neighboring protomer **(Figure IV.XI)** but is much less well resolved or unresolved in the other two protomers due to the absence of contact with neighboring LCA60 Fabs. The multiple contacts observed between LCA60 and oligosaccharides on the surface of the protein illustrates that the antibody accommodates MERS-CoV S glycan for binding and neutralization. Since the residues participating to the epitope and the glycosylation sequons at positions N166, N236 and N487 are conserved in >99.5% of MERS-CoV isolates, LCA60 is predicted to broadly neutralize most MERS-CoV viruses sequenced to date, although the chemical composition and occupancy of each glycan could modulate binding. LCA60 therefore shares similarities with several HIV-1 broadly neutralizing antibodies which interact to varying degree with carbohydrates N-linked to the Env glycoprotein (Scharf et al. 2015; Stewart-Jones et al. 2016).

Our structure rationalizes the effect of the previously described $E536A_{MERS\text{-}CoV}$ escape mutant that would disrupt putative electrostatic interactions with CDRH2 residues

$R50_{LCA60}$ and $K52_{LCA60}$ and/or CDRH3 residue $R103_{LCA60}$ (Corti et al. 2015) **(Figure IV.XI)**. It also suggests that the $K493E_{MERS\text{-}CoV}$ substitution could alter the positioning of the β5-β6 loop, which might in turn prevent CDRL1 from approaching the B domain, thereby providing a potential explanation for the observed loss of binding (Corti et al. 2015). The $T489A_{MERS\text{-}CoV}$ substitution would remove the glycan that extends toward the Fab CDRH2 bound to the closed B domain of the neighboring protomer **(Figure IV.XII)**. The observed abrogation of LCA60 binding for the $T489A_{MERS\text{-}CoV}$ mutant (Corti et al. 2015) might be explained by a loss of these protein/glycan contacts although only one out of the three bound Fabs interact with the glycan at position $N487_{MERS\text{-}CoV}$.

Comparison of the MERS-CoV S/LCA60 structure with the crystal structure of the MERS-CoV B domain in complex with DPP4 (Lu et al. 2013) shows that the two proteins would clash upon binding to MERS-CoV S and that the Fab and receptor epitopes partially overlap **(Figure IV.XII)**. This observation supports the finding that LCA60 and DPP4 compete for binding to MERS-CoV S (Corti et al. 2015). Whereas LCA60 can bind to all possible conformations of the B domain, DPP4 binding strictly requires opening of the B domain due to burial of its binding site in the closed conformation.

The S230 antibody was isolated from memory B cells of a SARS-CoV infected individual and potently neutralized a broad spectrum of SARS-CoV isolates of human and animal origin (Rockx et al. 2008; Traggiai et al. 2004). CryoEM characterization of the stabilized SARS-CoV S 2P glycoprotein in complex with the S230 Fab revealed that the particle images could be divided into SARS-CoV S/S230 complex with (state 1) or without (state 2) one closed B domain **(Figure IV.VIII, IV.IX, IV.XIII).** 3D classification of the cryoEM data revealed that the state 1 complex features multiple orientations of each of

the two S230 Fabs associated with intermediate and open conformations of the B domains. In the SARS-CoV S/S230 state 2 complex structure, the three B domains are open, albeit with conformations deviating from 3-fold symmetry, with the three bound Fabs protruding away from the S trimer apex **(Figure IV.XIII)**. We obtained asymmetric reconstructions at 4.2 Å and 4.5 Å resolution of the SARS-CoV S/S230 complex in states 1 and 2, respectively **(Figures IV.VIII, IV.IX, IV.XIII)**. The large conformational heterogeneity of the SARS-CoV S/S230 complex contrasts with the limited number of structural states detected for the MERS-CoV S/LCA60 complex. The marked mobility of the B domain/S230 regions, which adopt a continuum of conformations in both structures, limited the resolution of this part of the map. We therefore determined a crystal structure of the isolated S230 Fab at 1.5 Å resolution to assist interpretation of the data **(Figure IV.X)**.

Docking the S230 crystal structure in the cryoEM map indicated that CDRH2, CDRH3, CDRL1 and CDRL3 contact the B domain **(Figure IV.XIV)**, which also mediates binding to human angiotensin-converting enzyme 2 (ACE2), the SARS-CoV receptor (Li et al. 2003). The S230 epitope is centered around $L443_{SARS\text{-}CoV}$ and places the CDRH2 residue $F59_{S230}$ and the CDRH3 residues $Y106_{S230}$, $F107_{S230}$ and $Y110_{S230}$ near $Y408_{SARS\text{-}CoV}$, $Y442_{SARS\text{-}CoV}$, $F460_{SARS\text{-}CoV}$ and $Y475_{SARS\text{-}CoV}$. The fitting is therefore consistent with the identification of the $L443R_{SARS\text{-}CoV}$ substitution as the only escape mutant isolated thus far using the SARS-CoV Urbani strain (Rockx et al. 2008), since introduction of a charged residue would affect binding to this epitope.

Comparison of the SARS-CoV S/S230 structure with the crystal structure of the SARS-CoV B domain in complex with ACE2 (Li et al. 2005a) shows that the two proteins

would clash upon binding to SARS-CoV S and that the Fab and receptor epitopes partially overlap **(Figure IV.XIV)**. This observation supports our previous results suggesting that S230 and ACE2 compete to bind to SARS-CoV S (Rockx et al. 2008). In contrast to LCA60, which could recognize all possible arrangements of the B domain, S230 only interacts with intermediate and open states but not with the closed conformation. Since the interaction sites of both S230 and ACE2 are only accessible in the partially or fully open B domain conformations, binding of either of these two proteins to SARS-CoV S would sterically prevent sampling of the closed state and displace the conformational equilibrium towards open B domains **(Figure IV.XIII)**.

Based on the findings that (i) LCA60 and DPP4 or S230 and ACE2 compete to bind to the MERS-CoV or SARS-CoV domain B, respectively **(Figures IV.XI, IV.XIV)** and (ii) binding of DPP4, S230 or ACE2, but not LCA60, prevents closure of the B domain, we speculated that the conformational changes we captured by cryoEM are related to membrane fusion triggering. Incubation of the wild-type SARS-CoV S ectodomain trimer (i.e. without stabilizing mutations) with 1.6 µg/mL trypsin (w/v), to recapitulate proteolytic priming, led to cleavage at the S_1/S_2 boundary (Belouzard et al. 2009), as observed by SDS-PAGE **(Figure IV.XVI)**. Analysis of the negatively stained sample showed that the SARS-CoV S trimers largely remained in the prefusion conformation and were stable for several days on ice or at room temperature **(Figure IV.XVI)**. Complex formation between Fab S230 and wild-type SARS-CoV S led to formation of postfusion rosettes and the frequency of this transition was enhanced with longer incubation times and/or by trypsin cleavage **(Figures IV.XV, IV.XVI)**. The S_1 subunit acts as a “chaperone” contributing to stabilize S_2 in the prefusion state by reducing its propensity to transition to the postfusion

conformation. Proteolytic cleavage at the S_1/S_2, S_2' sites or shedding of the S_1 subunit, however, were not strictly required for refolding of the S ectodomain trimer to occur **(Figure IV.XVI).** These results are in agreement with our observation that the isolated MHV S_2 trimer ectodomain, uncleaved at the S_2' site, adopted the post-fusion state (Walls et al. 2017).

Incubation of the wild-type SARS-CoV S with the ACE2 ectodomain at up to 8 μM (4-fold molar excess) unexpectedly revealed that only a few percent of the particle images corresponded to complexes and a limited number of postfusion S trimers could be detected even in the presence of trypsin **(Figures IV.XV, IV.XVI)**. Prolonged incubation along with trypsin cleavage acted in synergy to increase the frequency of refolding, as attested by the presence of postfusion rosettes, although a significant fraction of unbound prefusion S trimers remained **(Figures IV.XV, IV.XVI)**. We also carried out the same set of experiments with SARS-CoV S pre-incubated with trypsin and similarly observed that S230 was more effective than ACE2 at inducing fusogenic conformational changes **(Figure IV.XVII)**.

To explore whether these findings could be detected in the context of virions, we analyzed the structural rearrangements of the full-length SARS-CoV S trimer embedded in the membrane of infectious murine leukemia (pseudo)virus (MLV)(Millet and Whittaker, 2016). Refolding of coronavirus S glycoproteins to the postfusion conformation can be detected by the appearance of a proteinase K-resistant band migrating at approximately 55 kDa when analyzed by SDS-PAGE. This molecular species comprises the postfusion 6-helix bundle, assembled from the heptad-repeats 1 and 2, since its formation could be inhibited in the presence of a heptad-repeat-derived peptide that prevents completion of

the conformational transition (Matsuyama and Taguchi, 2009). Addition of the S230 Fab **(Figure IV.XVIII)** or the ACE2 ectodomain **(Figure IV.XVIII)** to SARS-CoV S-pseudotyped MLV, in the presence of trypsin led to the formation of a proteinase K-resistant band migrating at approximately 55 kDa, thereby demonstrating both the antibody and the natural receptor promoted S fusogenic conformational changes in the context of a membrane-embedded full-length S.

Our negative staining EM findings suggest that ACE2 bound to the SARS-CoV S trimer with a lower affinity than S230 **(Figure IV.XVI)**. Since binding of ACE2 to the isolated B domain was reported to have an equilibrium dissociation constant (K_D) in the nanomolar range (, one would expect that full saturation should occur rapidly in the conditions of our experiments (100-fold concentration excess over the K_D). Previous reports, however, suggested a lower affinity of the full-length S trimer, relative to the isolated B domain, for ACE2 (Kirchdoerfer et al. 2018; Matsuyama and Taguchi, 2009; Song et al. 2018). We propose that the conformational landscape of the B domains in the context of the SARS-CoV S trimer, sampling open and closed conformations, decreases the apparent binding affinity for ACE2 and S230 due to masking of their binding sites in the closed state. A similar effect has previously been described for ubiquitin (. The higher binding affinity of S230 for SARS-CoV S, compared to ACE2 **(Figure IV.XIX)**, resulted in increased complex formation which explains the more efficient Fab-mediated membrane fusion activation compared to the receptor **(Figures IV.XVI, IV.XVII)**.

LCA60 could interact with all possible B domain states. In contrast, S230 binding was restricted to partially or fully open B domains and promoted S fusogenic conformational changes in a way reminiscent of the natural receptor, ACE2. These

outcomes indicate that S triggering could work through a ratcheting mechanism involving sequential binding of a receptor molecule to each of the RBM, which in turn traps the B domains in the open conformation and displaces the conformational equilibrium toward the open state. Opening of the B domains releases the constraints imposed on the fusion machinery, thereby allowing the S trimer to refold and fuse the viral and host membranes.

Both LCA60 and S230 bind to epitopes partially overlapping with the RBM on the B domain and in turn block attachment to DPP4 or ACE2, respectively, via a competitive mechanism. Moreover, S230 (or ACE2) binding triggered the SARS-CoV S transition to the postfusion conformation. This finding is an unprecedented example of functional mimicry, whereby an antibody activates membrane fusion by recapitulating the action of the receptor. It remains to be investigated whether binding of S230 triggers virus-cell fusion when particles are bound to the surface of cells.

Comparison of the binding modes of LCA60 and S230 with those of DPP4 and ACE2, provide a framework to understand the unique mechanism of coronavirus membrane fusion activation. Opening of each of the three B domains reduces the number of interactions between the S_1 and S_2 subunits and specifically releases the constraints imposed on the heptad-repeat 1-central helix hairpin, which is known to completely refold during the membrane fusion process (Walls et al. 2017). The 660 $Å^2$ decrease in buried surface area at the interface between the S_1 and S_2 subunits upon opening of the three B domains destabilizes the pre-fusion state and increases its probability of refolding, as observed upon complex formation with ACE2 or S230. Proteolytic activation is likely required to ensure that S glycoproteins will work in synergy, with proper spatial and temporal coordination, to drive fusion of the viral and host membranes. The large number

of S glycoproteins decorating the surface of coronaviruses suggests that multiple trimers work in concert for formation of a fusion pore. We expect this activation mechanism to hold true for all coronaviruses using domain B for interacting with their proteinaceous receptor. We view the S_1 subunit as a molecular chaperone, stabilizing the prefusion state of the fusion machinery, and coordinating receptor-binding with membrane fusion. In summary, coronaviruses appear to have evolved a fine-tuned balance between masking of the RBM, to limit neutralization by the humoral host immune response, and their necessary exposure, to enable receptor-recognition and infection of host cells.

IV.II Methods for coronavirus spike proteins in complex with neutralizing antibodies

Cell lines.

HEK293F is a female human embryonic kidney cell line transformed and adapted to grow in suspension (Life Technologies). HEK293F cells were grown in 293FreeStyle expression medium (Life Technologies), cultured at 37˚C with 5% CO_2 and 150 rpm. HEK293T/17 is a female human embryonic kidney cell line (ATCC). HEK293T/17 cells were cultured at 37˚C with 5% CO2 in flasks with DMEM + 10% FBS + penicillin-streptomycin + 10mM HEPES. Expi293F is a female human embryonic kidney cell line transformed and adapted to grow in suspension (Life Technologies). Expi293F cells were grown in Expi293F expression medium, cultured at 37˚C, 5% CO2 and 150 rpm. VeroE6 is a female kidney epithelial cell from African green monkey. VeroE6 cells were grown in serum-free medium (VP-SFM; ThermoFisher) at 37˚C and 5% CO_2. Cells lines were not tested for mycoplasma contamination nor authenticated except for the VeroE6 cells.

Construct design.

Genes encoding for the ectodomains of both MERS-CoV and SARS-CoV S were synthesized by GeneArt (ThermoFisher Scientific) and placed into a modified pOPING vector with its original N-terminal mu-phosphatase signal peptide, and an engineered C-terminal extension: SG-R*ENLYFQG (TEV protease site),* GGGSG-YIPEAPRDGQAYVRKDGEWVLLSTFL (foldon trimerization motif), G-HHHHHH (hexa-histidine tag), just upstream of the predicted transmembrane region (YNK for MERS-CoV S and YIK for SARS-CoV S). The MERS-CoV S_1/S_2 furin cleavage site was mutated and both SARS-CoV and MERS-CoV S were stabilized with the 2P mutations (. The SARS-CoV S_1 and ACE2 ectodomain constructs were fused to a sequence encoding a thrombin cleavage site and a human Fc fragment at the C-terminal end (kind gifts from Berend-Jan Bosch), as previously described (Raj et al. 2013).

Protein expression and purification.

The SARS-CoV S 2P and MERS-CoV S 2P ectodomains were produced in 500mL HEK293F cells grown in suspension using FreeStyle 293 expression medium (Life technologies) at 37°C in a humidified 8% CO2 incubator rotating at 130 r.p.m. The cultures were transfected using 293fectin (ThermoFisher Scientific) with cells grown to a density of 1 million cells per mL and cultivated for three days. The supernatants were harvested, and cells resuspended for another three days, yielding two harvests. Clarified supernatants were purified using a 5mL Cobalt affinity column (Takara). Purified protein was filtered or concentrated and flash frozen in Tris-saline (50 mM Tris pH 8.0, 150 mM NaCl) prior to negative staining and cryo-EM analysis. Wild-type SARS-CoV S was expressed in 293F cells and purified the same way as the SARS-CoV S 2P. Wildtype

SARS-CoV S was filtered through 0.2 μm filter and the quality of the protein was assessed by negative stain electron microscopy and SDS-PAGE. Wildtype SARS-CoV S was not frozen and was stored at 4°C and used within 7 days of purification.

Expression and purification of human angiotensin-converting enzyme ectodomain (ACE2, residues 1–614) fused to the Fc region of human IgG (hFc) was performed following the protocol previously described for CEACAM1a (Walls et al. 2016; Walls et al. 2016a) and cleavage of the Fc fragment was achieved using trypsin or thrombin. SARS-CoV S_1-Fc was produced in 250mL HEK293F cells grown in suspension using FreeStyle 293 expression medium (Life technologies) at 37°C in a humidified 8% CO2 incubator rotating at 130 r.p.m. The cultures were transfected using 293fectin (ThermoFisher Scientific) with cells grown to a density of 1 million cells per mL and cultivated for six days. The supernatant was harvested, clarified and incubated with 0.25mL protein A beads (0.5 mL slurry) (GenScript) overnight. Beads were washed thoroughly with phosphate buffered saline prior to elution with 0.1M glycine pH3.0 which was immediately neutralized with 1M Tris pH8.5. LCA60 and S230 Fabs were expressed using transient transfection of Expi293F cells, purified by affinity chromatography using CaptureSelect™ CH1-XL column (ThermoFisher), buffer exchanged to phosphate buffer saline using a HiTrap fast desalting column (GE Healthcare) and sterilized through a 0.2 μm filter.

Pseudovirus production.

MLV-based SARS-CoV S-pseudotyped were prepared as previously described (Millet and Whittaker, 2016). HEK-293T cells were cotransfected with a SARS-CoV S encoding-plasmid, an MLV Gag-Pol packaging construct and the MLV transfer vector encoding a luciferase reporter using the Lipofectamine 2000 transfection reagent (Life

Technologies) according to the manufacturer's instructions. Cells were incubated for 5 hours at 37 °C with 5% CO_2 with transfection medium. Cells were then washed with DMEM 2X and DMEM containing 10% FBS was added for 12 hours. The cells were washed again 2X with DMEM and were kept in DMEM without FBS for 48 hours (Wang et al. 2014). The supernatants were then harvested and filtered through 0.45-µm membranes.

MERS-CoV production.

Virions were purified from the supernatant of infected Vero E6 cells. Cells were infected at 90% confluence at multiplicity of infection of 0.01, in Serum Free Medium (VP-SFM; ThermoFisher Scientific). Cultures were maintained at 37˚C, 5% (v/v) CO_2 until 3 days post infection when the culture supernatant was removed. Approximately 50% of cells were rounding, but most remained attached to the growth surface. For virion purification, culture supernatant was clarified by centrifugation at 1,400 g for 10 minutes then overlaid on a cushion of 20% (w/w) Sucrose in Hanks Buffered Saline solution (HBSS: ThermoFisher Scientific) .Supernatant virions were centrifuged through the cushion at 100,000g for 90 minutes at 4˚C. Pelleted virions were resuspended in HBSS and disrupted by heating for 10 minutes at 95˚C in 1% (w/v) sodium dodecyl sulfate. All live virus work was undertaken under UK ACDP Containment level 3 conditions.

MERS-CoV S/LCA60 complex formation.

Flash frozen MERS-2P stabilized S at 5mg/mL was incubated in a 1:1.5 molar ratio with LCA60 at 2.3 mg/mL on ice for three hours. Trypsin at 3 µg/mL was added for twenty minutes at 37°C prior to injection over a gel filtration column (Superose 6 10/300 from GE Life Sciences) equilibrated in 20mM Tris pH8.0 and 150mM NaCl. The complex peak

fractions were concentrated to 0.5 mg/mL. 3 μL was applied to a glow discharged (20 seconds at 20 mA) lacey carbon copper grid with a thin layer of evaporated continuous carbon prior to plunge freezing into liquid ethane with a Mark IV vitrobot (ThermoFisher Scientific) using -1 blot force and 2 second blot time at 95% humidity and 25°C.

SARS-CoV S/S230 complex formation.

Lacey carbon copper grids were coated with a thin-layer of continuous carbon using a carbon evaporator and glow discharged for 20 seconds at 20 mA. Flash frozen SARS 2P stabilized S was diluted to 0.05 mg/ml and 2 µl was applied to the grid and incubated for 1 minute. Subsequently, the grid was washed three times in Tris-saline buffer before 2μL of S230 Fab at 0.02 mg/ml was applied to the grid and incubated for five minutes. The grid was washed again three times with Tris-saline after the fab incubation and manually blotted prior to plunge freezing into liquid ethane at room temperature.

Cryo-EM data collection.

Data were acquired using the Leginon software (to control 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 at a nominal magnification of 105,000x with a pixel size of 0.685 Å. The dose rate was adjusted to 8 counts/pixel/s, and each movie was acquired in super-resolution mode fractionated in 50 frames of 200 ms. For MERS-CoV S/LCA60, 2,400 micrographs were collected in a single session with a defocus range comprised between 0.6 and 3.0 µm.

For SARS-CoV S/S230, 5,900 images were collected in a single session with a defocus range comprised between 0.8 and 3.0 μm.

CryoEM data processing of MERS CoV S/LCA60.

Movie frame alignment was carried out with MotionCorr2 (Zheng et al. 2017) using a dose weighting of 40 $e^-/Å^2$. The parameters of the microscope contrast-transfer function were estimated with GCTF (Zhang, 2016). Particles were automatically picked with DoGPicker (Voss et al. 2009) through the Appion interface (. Particle images were extracted and processed with Relion3.0(Kimanius et al. 2016) with a box size of 768 $pixels^2$ binned to 384 $pixels^2$ yielding a pixel size of 1.37Å. Reference free 2D classification was used to parse 240,000 particles from the original 500,000 particle images using Relion3.0. An initial model was generated from PDB: 5W9J with all MERS-CoV S domain B closed and removing the G4 Fab (Pallesen et al. 2017). Relion3.0 3D classification with C1 symmetry was utilized to select the ~126,000 best particles for state 1 and the ~114,000 best particles for state 2. CTF refinement in Relion3.0 was used to refine per-particle defocus values. Particles images from each state were subjected to the Bayesian polishing procedure implemented in Relion3.0 (Zivanov et al. 2018; Zivanov et al. 2019) and 3D refinement before performing 3D classification, without refining angle and shifts, using soft masks focusing on the B domains and Fabs with a tau factor of 50. 82,000 and 78,000 particles were selected for states 1 and 2, respectively. Particle images from state 1 underwent another round of domain B/LCA60 focused classification to select 54,000 particles used for the final reconstruction. Finally, CryoSPARC non-uniform refinement (Punjani et al. 2017) was used to obtain the final reconstructions at 3.5 and 3.6 Å resolution, respectively, using soft masks excluding the open B domains

and bound Fabs. Reported resolutions are based on the gold-standard FSC=0.143 criterion (Rosenthal and Henderson, 2003; Scheres and Chen, 2012) and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution (Chen et al. 2013). Local resolution estimation, filtering and sharpening was carried out using CryoSPARC. Analysis was carried out using the state 1 structure since the reconstruction is best resolved.

CryoEM data processing of SARS-CoV S/S230.

Movie frame alignment was carried out with MotionCorr2 (Zheng et al. 2017) using a dose weighting of 40 $e^-/Å^2$. The parameters of the microscope contrast-transfer function were estimated with GCTF (Zhang, 2016). Particles were automatically picked with DoGPicker (Voss et al. 2009) through the Appion interface (. Particle images were extracted and processed with Relion3.0(Kimanius et al. 2016) with a box size of 768 $pixels^2$ binned to 72 $pixels^2$ (for 2D classification), 128 $pixels^2$ (for 3D classification) or 384 $pixels^2$ (for 3D refinement, yielding a pixel size of 1.37Å). Two rounds of Relion reference-free 2D classification were used to select ~350,000 particles clustering in well-defined class averages. An initial model was generated from a SARS-CoV S structure with all B domains closed (PDB: 5X58) (Yuan et al. 2017). Relion 3D classification without applying symmetry was utilized to select ~162,000 particle images for state 1 (including numerous relative conformations of the two bound Fabs as well as partial occupancy) and ~126,000 particle images for state 2 (including numerous relative conformations of the three bound Fabs as well as partial occupancy). Particle images in each subset were then subsequently subjected to a new round of 3D classification without imposing symmetry and using the same initial model as in the previous round. A subset of ~65,000

particle images from state 1 and ~71,000 particle images from state 2 were selected based on reduced conformational heterogeneity of the bound Fabs (which exhibit a continuum of conformational states) and independently used to run 3D auto-refinements using Relion. Particles images from each state were subjected to the Bayesian polishing procedure implemented in Relion3.0 (Zivanov et al. 2018; Zivanov et al. 2019) and 3D refinement before performing 3D classification, without refining angle and shifts, using soft masks focusing on the B domains and Fabs with a tau factor of 30. ~23,000 and 20,000 particles were selected for states 1 and 2, respectively, based on mobility of the bound Fabs which were less well resolved than SARS-CoV S (especially the S_2 subunit) in the maps. Finally, CryoSPARC non-uniform refinement (Punjani et al. 2017) was used to obtain the final reconstructions at 4.2 Å and 4.5 Å resolution for states 1 and 2, respectively. Reported resolutions are based on the gold-standard FSC=0.143 criterion (Rosenthal and Henderson, 2003; Scheres and Chen, 2012) and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution (Chen et al. 2013). Local resolution estimation, filtering and sharpening was carried out using CryoSPARC.

CryoEM Model building and analysis.

UCSF Chimera (Goddard et al. 2007) and Coot were used to fit atomic models (PDB: 5W9J for MERS and PDB: 5X58 for SARS) into the cryoEM maps. The crystal structure of LCA60 or S230 were also fit into density using UCSF Chimera. The models were subsequently manually rebuilt using Coot (Brown et al. 2015; Emsley et al. 2010). N-linked glycans were hand-built into the density where visible and the models were refined and relaxed using Rosetta (Wang et al. 2016). Glycan refinement relied on a

dedicated Rosetta protocol, which uses physically realistic geometries based on prior knowledge of saccharide chemical properties (Frenz et al. 2019), and was aided by using both sharpened and unsharpened maps. Models were analyzed using MolProbity (Chen et al. 2010) and privateer. Figures were generated using UCSF ChimeraX (Goddard et al. 2018).

Negative stain electron microscopy.

Protein was adsorbed to glow-discharged carbon-coated copper grids for 1 min prior to a 3X wash with water and 2% uranyl formate staining. Micrographs were recorded using the Leginon software (on a 100kV FEI Tecnai G2 Spirit with a Gatan Ultrascan 4000 4k x 4k 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. Particles were picked automatically in a reference free manner using DogPicker. Contrast transfer function (CTF) estimation was performed using GCTF(Zhang, 2016). Class averages were generated using Relion2.1(Kimanius et al. 2016).

Crystallization, data collection and data processing of LCA60 and S230.

Fab LCA60 crystals were grown in hanging drop set up with a mosquito at 20˚C using 150 nL protein solution and 150 nL mother liquor solution containing 30 % (v/v) Jeffamine ED-2001 pH 7.0 and 0.1 M HEPES pH 7.0. The LCA60 dataset was collected at the ALS beamline 5.0.1 and processed to 3.0 Å resolution (Karplus and Diederichs, 2012) in space group $P2_1$, with cell dimensions a = 70.72 Å b = 69.57 Å and c = 93.84 Å, using XDS (Kabsch, 2010) and Aimless (Evans and Murshudov, 2013). Fab S230 crystals were grown in hanging drop set up with a mosquito at 20˚C using 150 nL protein solution and 150 nL mother liquor solution containing 3.5 M sodium formate pH 7.0. Crystals were

cryo-protected using the mother liquor solution supplemented with 30% glycerol. The S230 dataset was collected at the ALS beamline 5.0.1 and processed to 1.5 Å resolution in space group $P2_12_12_1$ using XDS and Aimless. The structures of Fab LCA60 and Fab S230 were solved by molecular replacement using Phaser (McCoy et al. 2007) and homology models as search models. The coordinates were subsequently improved and completed using Buccaneer (Cowtan et al. 2006) and COOT (Brown et al. 2015; Emsley et al. 2010) and refined with BUSTER-TNT (Blanc et al. 2004) and REFMAC5 (Vagin et al. 2004). Simulated annealing and omit map calculations were performed in Phenix (Adams et al. 2010). The quality of the crystal structures were analyzed using MolProbity (Chen et al. 2010). Crystallographic data collection and refinement statistics are shown in **Figure IV.X**.

Ligand-induced conformational change analysis using negative staining EM.

The wildtype SARS-CoV S ectodomain trimer at 1 mg/ml (7.5 μM spike monomer) was incubated with equimolar amount of Fab S230 or 4-fold molar excess of ACE2 for 30 minutes or overnight (20-30 hours) on ice. Samples were taken at the end of the incubation and imaged by negative staining EM. The incubation mixtures were also subjected to digestion with trypsin at 1.6 μg/ml at room temperature and reactions were analyzed by SDS-PAGE at different time points. The digestion reaction was stopped by addition of 1.5 mM PMSF after 15 minutes for imaging by negative staining EM. To visualize the effect of ligand binding on pre-digested wildtype SARS-CoV S ectodomain trimer, the protein at 1 mg/ml was digested with 1.6 μg/ml trypsin for 15 minutes at room temperature before stopping the reaction using 1.5 mM PMSF. The pre-digested wildtype

SARS-CoV S ectodomain trimer was incubated with S230 or ACE2 on ice for 30 min or overnight before negative staining EM imaging.

Ligand-induced conformational change analysis using Western-blotting.

SARS-CoV S pseudovirions were thawed and incubated either with an equal volume of buffer, 10 μM of the ACE2 ectodomain or 500 nM S230 Fab for 30 minutes or 2 hours on ice. Trypsin (5 μg/mL) was then added to these samples for 30 minutes at 37°C. Subsequently, the samples were supplemented with 10 μg/mL proteinase K for 30 minutes at 4°C. 6X SDS-PAGE loading buffer was then added to all samples prior to boiling at 95°C for 5 min. Samples were run on a 4-15% gradient Tris-Glycine Gel (BioRad) and transferred to PVDF membranes. An anti-S_2 SARS-CoV S polyclonal primary antibody (1:1,500 dilution, ThermoFisher) and an Alexa Fluor 680-conjugated goat anti-rabbit secondary antibody (1:20,000 dilution, Jackson Laboratory) were used for Western-blotting. A LI-COR processor was used to develop images.

Mass spectrometry.

In solution: MERS-CoV S or SARS-CoV S were either unaltered or subjected to Endo-F and Endo-H deglycosylation treatment. Two microliters of the deglycosylases were allowed to incubate with 20 μg of S for 4-14 hrs in 50mM sodium acetate pH 4.4 at 37°C in a 20 μL reaction. 6 μg of S were then incubated in 100mM Tris pH 8.5, 2% sodium deoxycholate, 10mM tris(2-carboxyethyl)phosphine, and 40mM iodoacetamide at 95°C for five minutes and 25°C for thirty minutes in the dark. Denaturated, reduced, and alkylated S (1.6 μg) was then diluted into fresh 50mM Ammonium bicarbonate and incubated for 14 hrs at 37°C either with 0.032 μg of trypsin (Sigma Aldrich), chymotrypsin (Sigma Aldrich), trypsin then sequentially chymotrypsin, alpha lytic protease (Sigma), or

AspN protease (Pierce). Formic acid was then added to a final concentration of 2% and the samples spun at 14,000 rpm for 20 minutes and then again for 5 minutes to precipitate the sodium deoxycholate and collect the peptides from the supernatants. The SARS S digests with trypsin, chymotrypsin and alpha lytic protease were processed in parallel after cotton-enrichment of glycopeptides. Approximately 0.5 mg of cotton wool was packed into 100 µL pipette tips. Glycopeptides were diluted in 80% acetonitrile with 0.1% formic acid and bound to the cotton plug, followed by washing of the cotton plug with 80% acetonitrile with 0.1% formic acid. Glycopeptides were eluted in 0.1% formic acid in water.

In gel: 120 µL of MERS-CoV England1 and EMC2012 virions and 40 µg of purified MERS2P were run on a 4-15% gradient Tris-Glycine Gel (BioRad). One lane of each sample was loaded with 40 µL while a second lane was loaded with 40 µL, allowed to run for 2 minutes, and loaded with another 40 µL. The bands were excised and washed with milliQ water and acetonitrile. 1 mg/mL DTT in 50mM ammonium bicarbonate was incubated with the gel pieces at 60°C for 1 hour. Following an acetonitrile wash, 10 mg/mL iodoacetamide in 50mM ammonium bicarbonate was incubated for 30 minutes at room temperature in darkness. Following extensive washes, either 30 µg/mL trypsin or chymotrypsin in ammonium bicarbonate was incubated with the gel pieces for 1.5 hours on ice. The solution was replaced with ammonium bicarbonate and incubated overnight at 37°C. The supernatant was collected and dried down in a speedvac for 1 hr. The peptides were reconstituted in 40 µL 10% formic acid, 5% DMSO.

Peptide supernatant was collected and between 4 and 8 µL were run on a Orbitrap Fusion Tribrid (ThermoFisher Scientific) mass spectrometer. A 35 cm analytical column and a 3cm trap column filled with ReproSil-Pur C18AQ 5 µm (Dr. Maisch GMBH) beads

was used. Nanospray LC-MS/MS was used to separate peptides over a 110-minute gradient from 5% to 30% acetonitrile. A positive spray voltage of 2100 was used with an ion transfer tube temperature of 350°C. An electron-transfer/higher energy collision dissociation ion fragmentation scheme was used (Frese et al. 2013). The supplemental higher energy collision dissociation energy was 0.15 for the runs. A resolution setting of 120,000 with an AGC target of $2*10^5$ was used for MS1, and a resolution setting of 30,000 with an AGC target of $1*10^5$ was used for MS2. The data was then analyzed using Byonic (Bern et al. 2012) against a custom database of recombinant coronavirus spike proteins, searching for glycan modifications with 12/24 ppm search windows for MS1/MS2, respectively. The in-solution digestions of the recombinant ectodomain were searched against a custom database of recombinant coronavirus spike proteins, proteases, and common contaminants. The in-gel digestion experiments were first searched against a database comprised of the full MERS-CoV proteome and the full Chlorocebus sabaeus (virus produced in VeroE6 cells) or human proteome (ectodomain produced in HEK293F cells) retrieved from UniProt, with variable N-linked glycosylations omitted from the search to identify background contaminants. The top 16 contaminants were then assembled together with the S sequence into a focused database for the glycopeptide identifications. In all experiments, up to three missed cleavages were permitted using C-terminal cleavage at R/K for trypsin, F/Y/W/M/L for chymotrypsin, R/K/F/Y/W/M/L for trypsin-chymotrypsin, T/A/S/V for alpha lytic protease, and N-terminal cleavage at D for AspN. Carbamidomethylation of cysteine was set as fixed modification, methionine oxidation and carbamidomethylation of lysine as variable common 1, glycan modifications as variable common 2, with a variable common max. parameter of 3. The N-linked glycan

databases of Byonic were merged into a single non-redundant set of glycan compositions for the searches. All reported glycopeptides in the Byonic results files were manually inspected for quality of fragment assignments. All glycopeptide identifications were merged into a single non-redundant list per sequon. Glycans were classified based on HexNAc content as high-mannose (2 HexNAc), hybrid (3 HexNAc) or complex (>3 HexNAc). Byonic search results were exported to mzIdentML format to build a spectral library in Skyline (MacLean et al. 2010) and extract peak areas for individual glycoforms from MS1 scans. The full database of variable N-linked glycan modifications from Byonic was manually added to the Skyline project file in XML format. Byonic search results from in-gel digestions of both virion samples and ectodomain were merged into a single spectral library and manually inspected for quantitation. Reported peak areas were pooled based on the number of HexNAc, Fuc or NeuAc residues to distinguish high-mannose/hybrid/complex glycosylation, fucosylation and sialylation, respectively.

SEC-MALS.

30 µl of each sample at 2.5 mg/ml were loaded onto a Vanquish column (ThermoFisher Scientific) with Acquity UPLC protein BEH SEC, 450Å, 2.5 µm, 4.6 mm x 150mm (Waters) and passed through a Wyatt µDAWN coupled to a Wyatt Optilab UT-rEX differential refractive index detector. Data were analyzed using Astra 6 software (Wyatt Technology Corp) to quantify the protein and glycan weight.

Biolayer interferometry.

Assays were performed on an Octet Red instrument at 30°C with shaking at 1,000 RPM. ARG2 biosensors were hydrated in water then activated for 300 seconds with an NHS-EDC solution (ForteBio) prior to amine coupling. 30 µg/mL SARS-S_1-Fc was amine

coupled to ARG2 (ForteBio) sensors in pH5.0 10mM acetate (ForteBio) respectively for 600 seconds (4nm shift) and then quenched with 1M ethanolamine for 600 seconds. A baseline in 25mM Phosphate pH8.0, 150mM NaCl, 1% BSA was collected for 120 seconds prior to immersing the sensors in a serial dilution of either ACE2 or S230 for 600 seconds prior to dissociation in buffer for 600 seconds. ACE2 concentrations used were 1,000 nm, 333 nM, 111 nM, 37 nM, 12.3 nM, and 4.12 nM. S230 concentrations used were 100 nM, 33.3 nM, 11.1 nM, 3.7 nM, 1.23 nM, and 0.412 nM. Curve fitting was performed using a 1:1 binding model and the ForteBio data analysis software. Mean k_{on}, k_{off} values were determined with a global fit applied to all data.

QUANTIFICATION AND STATISTICAL ANALYZES

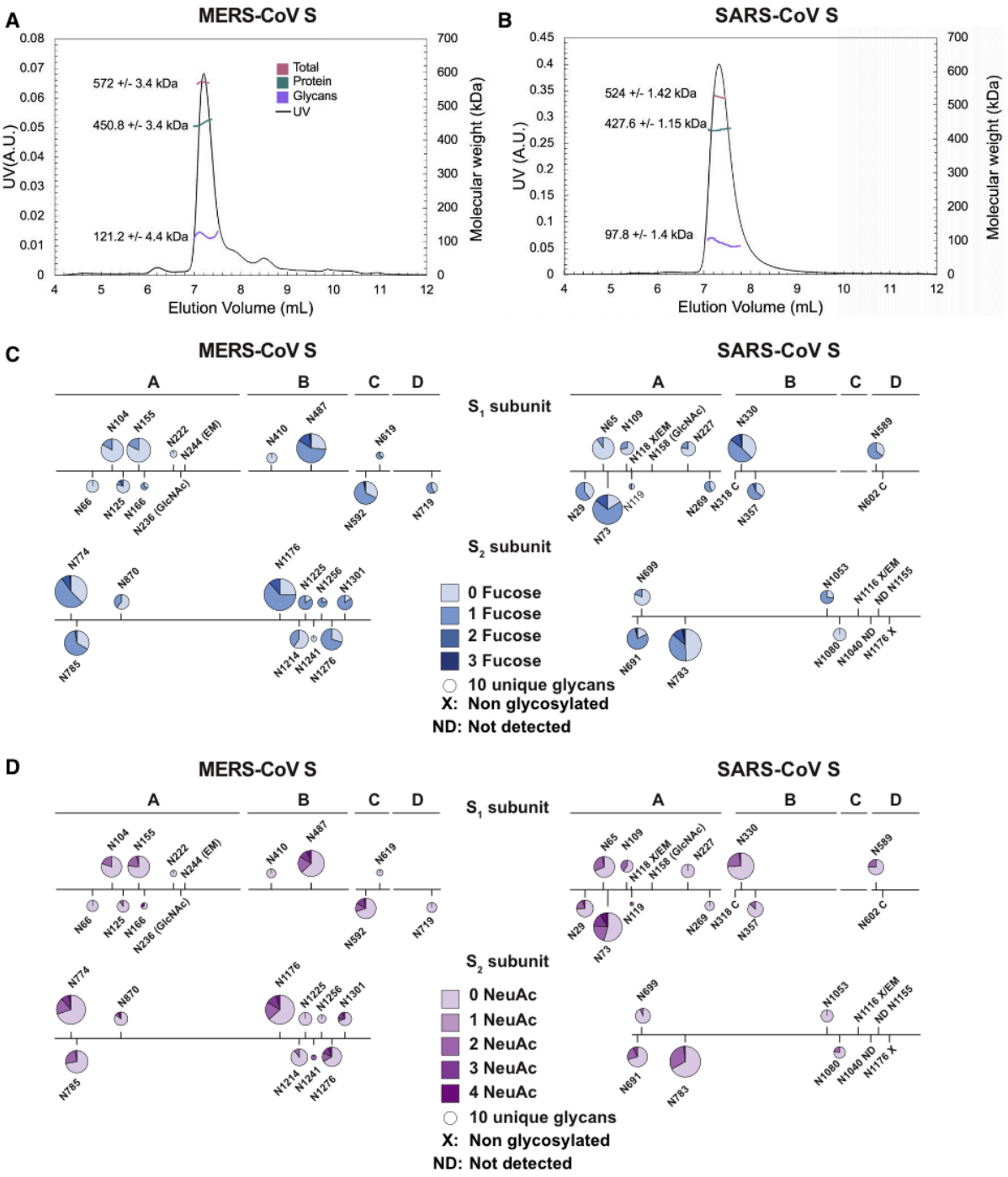
Digestion followed by Western-blotting experiments as well as biolayer interferometry assays were performed in triplicates.

DATA AND SOFTWARE AVAILABILITY

The cryoEM maps and atomic models have been deposited at the Electron Microscopy Data Bank and the Protein Data Bank with accession codes EMDB: 0401 and PDB: 6NB3 (MERS-CoV S/LCA60 state 1), EMDB: 0402 and PDB: 6NB4 (MERS-CoV S/LCA60 state 2), EMDB: 0403 and PDB: 6NB6 (SARS-CoV S/S230 state 1), EMDB: 0404 and PDB: 6NB7 (SARS-CoV S/S230 state 2). The crystal structures have been deposited in the Protein Data Bank with accession codes PDB: 6NB5 (LCA60) and PDB: 6NB8 (S230). The mass spectrometry data have been deposited in the ProteomeXchange consortium (PRIDE) databank with accession codes PRIDE: PDX010494.

Figures Chapter IV: IV.I-IV.XIX: Adapted from Walls et al. 2018

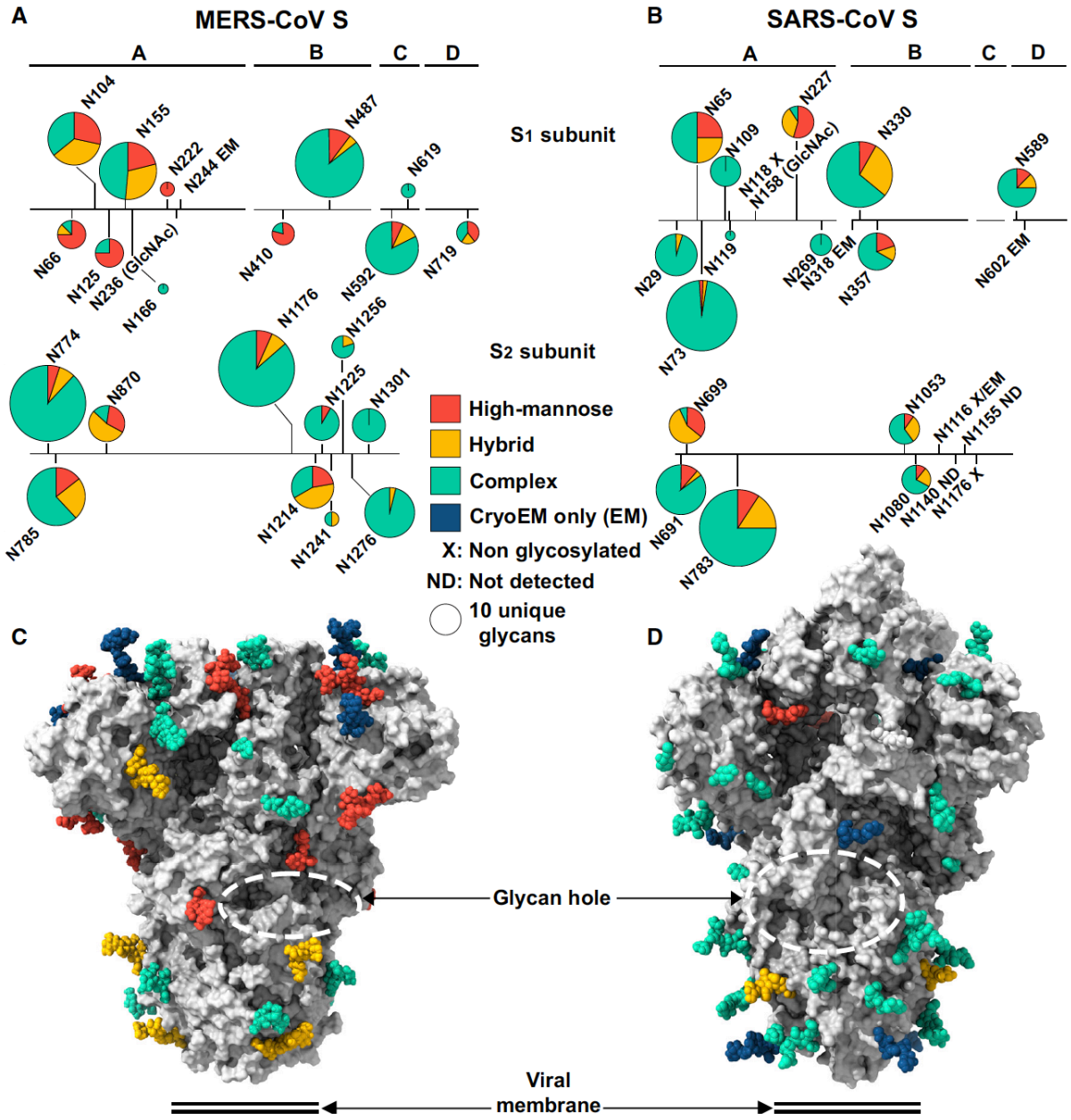
Figure IV.I: Characterization of the Fine Molecular Structure of the MERS-CoV and SARS-CoV S Glycan Shields



(A-B) Quantification of the molecular weight fraction contributed by the N-linked carbohydrates decorating the MERS-CoV (A) and SARS-CoV (B) S glycoproteins expressed using HEK293F cells. Molecular weight estimates were determined using size-

exclusion chromatography coupled with multi-angle light scattering, refractometry and UV spectrophotometry (Veesler et al. 2009). (C-D) LC-MS/MS analysis of the fucosylation (C) and sialylation (D) content of the MERS-CoV S 2Pand SARS-CoV S 2P ectodomain glycans (in-solution digests). Each N-linked glycosylation site is represented by a pie chart colored according to the number of fucose or sialic acid moieties detected and for which the diameter is scaled based on the number of unique glycopeptides identified. NeuAc: N-acetyl-neu-raminic acid.

Figure IV.II: Characterization of the S glycan shield of MERS-CoV and SARS-CoV S glycoproteins



(A and B) Analysis of the glycans N-linked to MERS-CoV S 2P (A) and SARS-CoV S 2P (B) ectodomain trimers expressed using HEK293F cells. Each site is represented by a pie chart colored according to the processing state detected by LC-MS/MS and for which the diameter is scaled based on the number of unique glycopeptides identified. Glycan assignment was performed as follows: high-mannose, 2 HexNAc; hybrid, 3 HexNAc; complex, R4 HexNAc; GlcNAc, N-acetyl glucosamine (detected for endo-H-treated samples).

(C and D) Surface representation of the fully glycosylated MERS-CoV S 2P (C) and SARS-CoV S 2P (D) ectodomain trimers (gray) with N-linked glycans rendered as spheres colored according to the processing state (i.e., corresponding to the largest number of unique N-linked glycans identified by LC-MS/MS, as in A and B). Domains A–D are labeled in (A) and (B). (C) and (D) were generated using the MERS-CoV S2P/LCA60 and SARS-CoV S2P/S230 structures for which the Fabs were removed for clarity.

Figure IV.III: Site specific glycan identifications of the MERS-CoV S 2P ectodomain trimer (in solution digest). Refer to https://doi.org/10.1016/j.cell.2018.12.028 for entire table

residue #	sequon	All comb-ined	Tryp	Tryp-HF	Chym	Chym-HF	Tryp-Chym	aLP	AspN	Un-glyco-sylated
66	NITI	9	5	1	5	3	5	6	-	-
104	NYSQ	30	23	19	14	9	19	12	1	x
125	NSTG	9	8	2	1	1	7	-	-	x
155	NFSD	33	27	21	-	1	24	-	-	x
166	NHTL	3	-	1	-	1	-	2	-	-
222	NASL	3	-	1	-	1	1	1	-	-
236	NCTF	2	-	-	-	1	-	1	-	-
244	NITE	-	-	-	-	-	-	-	-	-
410	NLTK	6	2	1	-	1	3	1	-	-
487	NLTT	48	-	2	12	31	32	31	-	x
592	NDTK	29	17	18	-	-	21	7	-	x
619	NCTA	3	-	2	-	-	2	-	-	-
719	NSSL	7	-	-	2	5	3	-	-	-
774	NSSY	59	47	37	17	10	20	7	-	x
785	NFSF	35	14	14	26	10	27	23	-	x
870	NLTL	12	-	-	2	1	1	10	1	-
1176	NNTR	59	-	1	41	21	32	35	-	-
1214	NIST	19	-	3	-	-	-	18	-	-
1225	NSTG	11	-	-	-	-	-	11	1	-
1241	NVST	2	-	-	-	-	-	1	1	-
1256	NTTL	5	-	-	1	-	-	4	-	-
1276	NESYI	26	13	24	-	-	13	2	-	x
1301	NYT	12	5	15	-	-	2	-	-	x
		TOTAL	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	
		422	161	162	121	96	212	172	4	

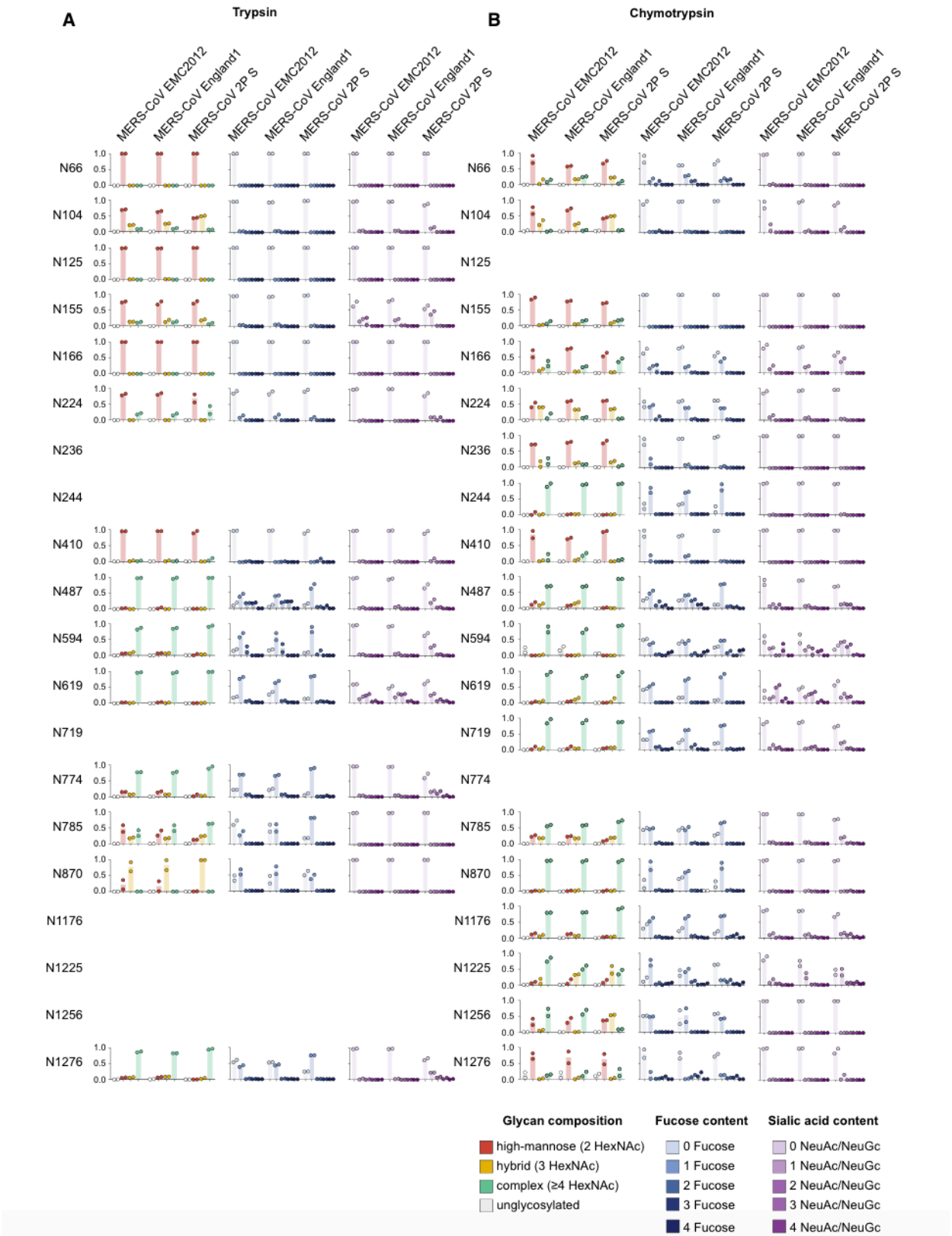
Figure IV.IV: Site specific glycan identifications of the SARS-CoV S 2P ectodomain trimer (in solution digest) (two pages). Refer to https://doi.org/10.1016/j.cell.2018.12.028 for entire table

residue #	sequon	All combined	Tryp	Tryp+	Tryp-HF	Chym	Chym+
29	NYTQ	19	-	-	3	-	-
65	NVTG	29	-	-	-	14	10
73	NHTF	54	-	-	-	42	27
109	NKSQ	10	-	-	-	-	-
118	NNST	X	-	-	-	-	-
119	NSTN	1	-	-	-	-	-
158	NCTF	1	-	-	-	-	-
227	NITN	12	10	5	-	8	6
269	NGTI	6	-	-	5	-	-
318	NITN	-	-	-	-	-	-
330	NATK	47	-	-	-	46	36
357	NSTF	16	-	-	-	9	5
589	NASS	16	-	-	-	15	7
602	NCTD	-	-	-	-	-	-
691	NNTI	27	-	-	-	1	-
699	NFSI	15	-	-	-	-	-
783	NFSQ	64	58	23	-	22	6
1056	NFTT	10	-	-	-	-	-
1080	NGTS	9	-	-	-	1	1
1116	NNTV	X	-	-	-	-	-
1140	NHTS	-	-	-	-	-	-
1155	NASV	-	-	-	-	-	-
1176	NESL	X	-	-	-	-	-
		TOTAL	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**
		336	68	28	8	158	98

Continued Figure IV.IV:

residue #	sequon	Chym-HF	Tryp-Chym	aLP	aLP+	AspN	Unglycosylated
29	NYTQ	-	-	9	8	7	X
65	NVTG	5	6	25	8	-	X
73	NHTF	-	17	25	3	-	X
109	NKSQ	-	-	-	13	-	-
118	NNST	-	-	-	-	-	X
119	NSTN	-	-	1	1	-	-
158	NCTF	1	-	-	-	-	-
227	NITN	1	6	5	4	-	-
269	NGTI	-	-	1	1	-	X (deaminated)
318	NITN	-	-	-	-	-	-
330	NATK	5	19	3	-	-	-
357	NSTF	4	15	-	-	-	-
589	NASS	2	6	3	2	-	-
602	NCTD	-	-	-	-	-	-
691	NNTI	6	-	16	20	-	-
699	NFSI	2	-	12	10	-	-
783	NFSQ	-	17	53	4	-	X
1056	NFTT	-	1	7	4	-	-
1080	NGTS	-	6	2	-	-	X (deaminated)
1116	NNTV	-	-	-	-	-	X
1140	NHTS	-	-	-	-	-	-
1155	NASV	-	-	-	-	-	-
1176	NESL	-	-	-	-	-	X
		TOTAL	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	
		26	93	162	78	7	

Figure IV.V: Comparison of glycans N-linked to full length MERS-CoV S incorporated in authentic virions and to a purified ectodomain

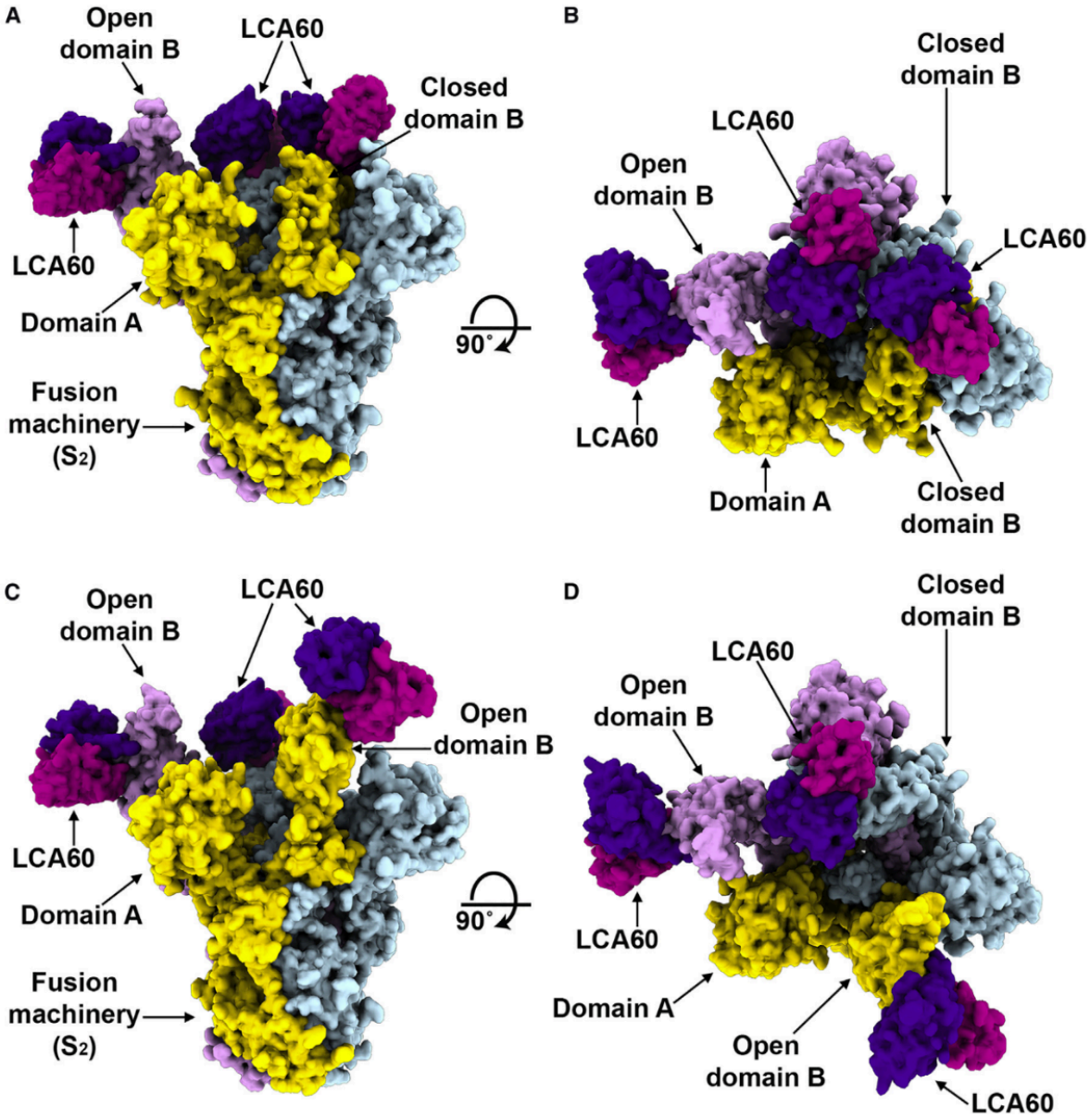


(A-B) LC-MS/MS analysis of S oligosaccharides from MERS-CoV England1 and EMC/2012 virions as well as MERS-CoV S 2P ectodomain trimer (in-gel digests). Colored dots represent individual technical replicate values and histograms represent the experimental mean signal of the relative integrated MS1 peak areas foreach glycopeptide upon trypsin (A) or chymotrypsin digestion (B). NeuAc: N-acetyl-neuraminic acid, NeuGc: N-glycolyl-neuraminic acid.

Figure IV.VI: Site-Specific Glycan Identifications of the Full-Length S Trimer from Authentic MERS-CoV England1 and EMC/2012 virions, as well as the MERS-CoV S 2P Ectodomain Trimer (In-Gel Digests). Refer to https://doi.org/10.1016/j.cell.2018.12.028 for entire table

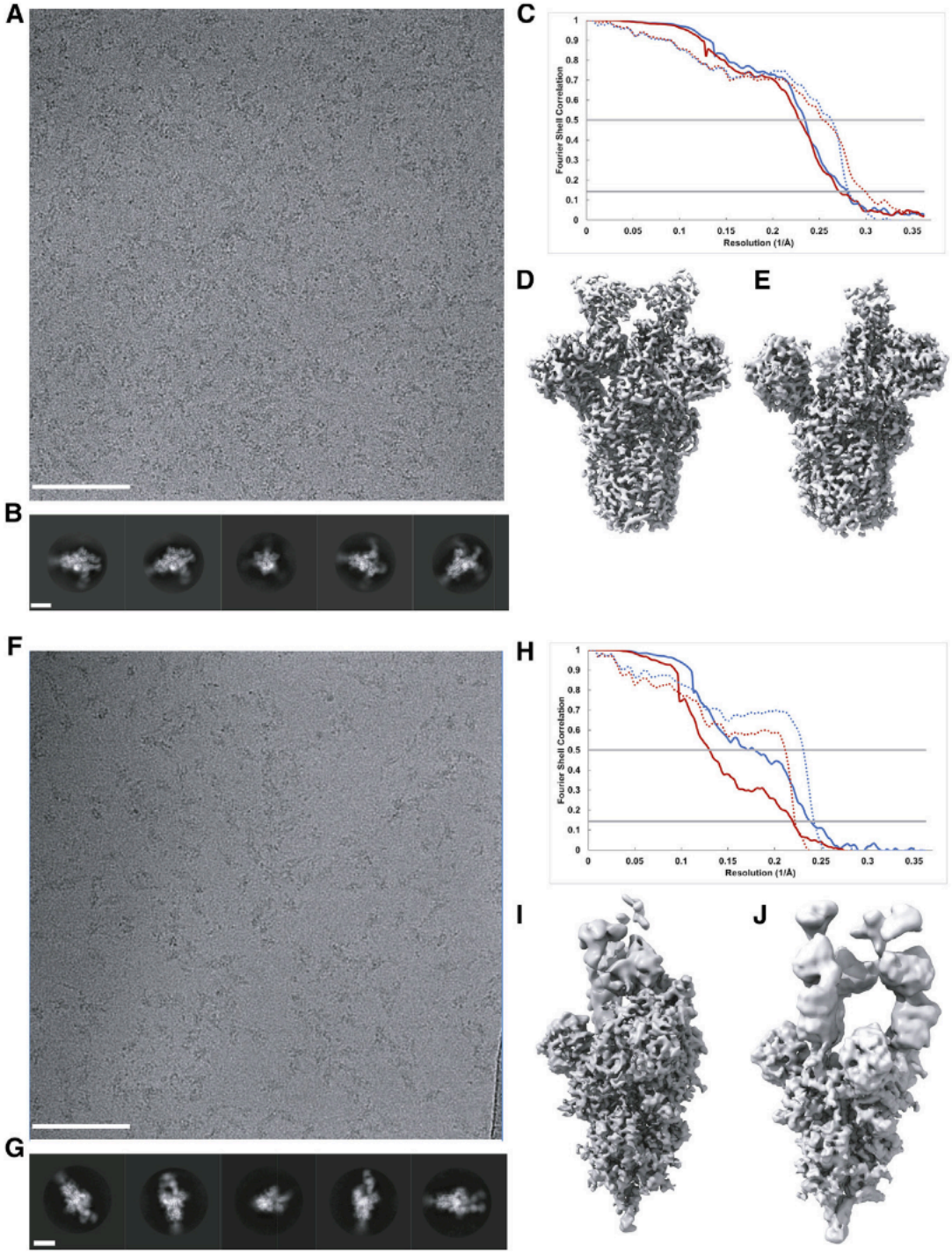
residue #	sequon	EMC T	EMC C	England T	England C	M2P T	M2P C
66	NITI	2	5	1	-	2	2
104	NYSQ	11	6	10	-	12	2
125	NSTG	4	-	1	-	7	-
155	NFSD	17	2	10	-	11	-
166	NHTL	2	-	2	-	2	1
222	NASL	3	-	2	-	1	-
236	NCTF	-	4	-	1	-	3
244	NITE	-	1	-	-	-	6
410	NLTK	2	-	3	-	3	-
487	NLTT	-	6	-	-	10	3
592	NDTK	12	-	5	-	34	-
619	NCTA	4	-	3	-	28	1
719	NSSL	-	4	-	-	-	3
774	NSSY	16	-	7	-	45	1
785	NFSF	-	2	-	-	6	6
870	NLTL	-	-	-	-	-	-
1176	NNTR	-	7	-	-	-	8
1214	NIST	-	-	-	-	-	-
1225	NSTG	-	-	-	-	-	-
1241	NVST	-	1	-	-	-	-
1256	NTTL	-	-	-	-	-	-
1276	NESYI	4	-	3	-	41	-
1301	NYT	-	-	-	-	33	-
		TOTAL	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**	**TOTAL**
		77	38	47	1	235	36

Figure IV.VII: CryoEM Structures of the MERS-CoV S Glycoprotein in Complex with the LCA60 Neutralizing Antibody



(A and B) Orthogonal views of the state 1 structure with one open and two closed B domains.(C and D) Orthogonal views of the state 2 structure with two open and one closed B domains. The structures are rendered as molecular surfaces with different colors for each S protomer (light blue, plum, and gold) and the LCA60 Fab heavy (purple) and light(pink) chains (only the variable domains are shown). The open B domains/Fabs are only included for visualization and were omitted from the final models.

Figure IV.VIII: CryoEM characterization of the MERS-CoV S glycoprotein in complex with the LCA60 fab fragment (A-E) and the SARS-CoV S glycoprotein in complex with the S230 fab fragment (F-J)



(A) Representative micrograph. Scale bar is 100 nm. (B) Reference-free 2D class averages. Scale bar is 100 A˚. (C) Gold standard (solid lines) and map/model (dotted

lines) Fourier shell correlation curves for the state 1 (blue lines) or state 2 (red lines) reconstructions. (D-E) Reconstructions for state1 (D) and state 2 (E). (F)Representative micrograph. Scale bar is 100 nm. (G) Reference-free 2D class averages. Scale bar is 100 A˚. (H) Gold standard (solid lines) and map/model (dottedlines) Fourier shell correlation curves for the state 1 (blue lines) or state 2 (red lines) reconstructions. (I-J) Reconstructions for state 1 (I) and state 2 (J)

Figure IV.IX: CryoEM data collection and refinement statistics

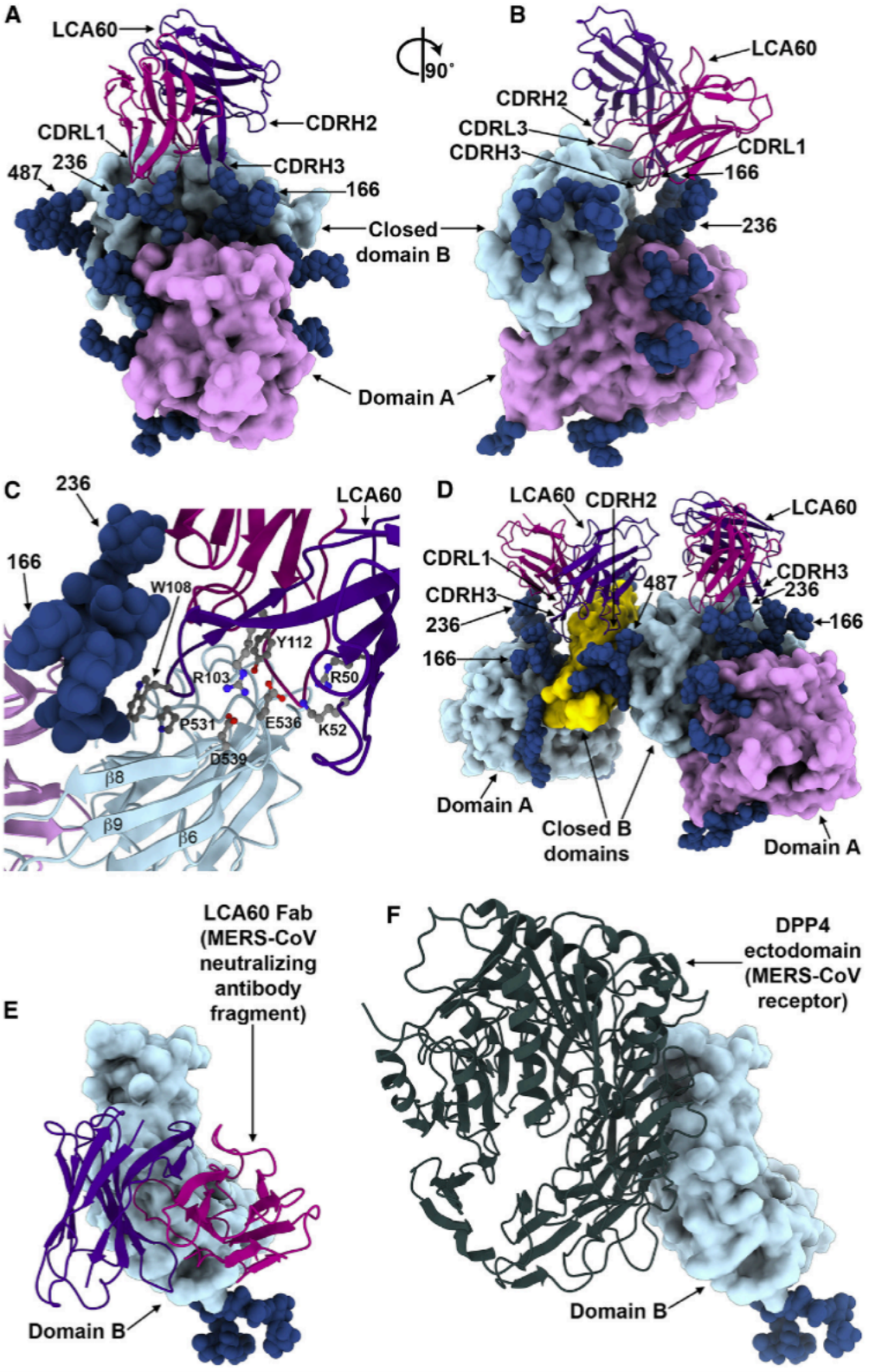
Data Collection	**MERS-LCA60 state 1**	**MERS-LCA60 state 2**	**SARS-S230 state 1**	**SARS-S230 state 2**
PDB ID	6NB3	6NB4	6NB6	6NB7
EMDB ID	EMD-0401	EMD-0402	EMD-0403	EMD-0404
Number of Particles	54,071	78,719	23,071	19,986
Pixel Size (Å)	1.37	1.37	1.37	1.37
Defocus Range (μm)	1.0-3.2	1.0-3.2	0.8-3.0	0.8-3.0
Voltage (kV)	300	300	300	300
Electron dose (e^-/Å^2)	40	40	40	40
Refinement				
Resolution	3.5	3.6	4.2	4.5
Map sharpening B factor (Å^2)	-65	-75	-37	-50
Model Validation				
Molprobity score	0.97	0.94	1.05	1.06
All-atom clashscore	1.07	0.98	1.14	1.2
Poor rotamers (%)	0.31	0.32	0.37	0.51
Ramachandran favored (%)	97.1	97.2	96.5	96.6
Ramachandran allowed (%)	99.81	99.82	99.8	99.8
Ramachandran outliers (%)	0.19	0.18	0.2	0.2

Figure IV.X: X-ray crystallography data collection and refinement statistics

	LCA60	S230
PDB ID	6NB5	6NB8
Data collection		
Space group	$P2_1$	$P2_12_12_1$
Cell constants		
a,b,c (Å)	70.72, 69.57, 93.84	37.8, 104.4, 108.7
α,β,γ (˚)	90.0, 107.7, 90.0	90, 90, 90
Wavelength (Å)	0.9774	0.9774
Resolution (Å)	39.0-3.0(3.2-3.0)	38.0-1.5 (1.58-1.50)
Rmerge (%)	24.0(83.0)	5.9 (64.1)
I/σ(I)	3.2 (1.0)	16.2 (2.5)
CC (1/2)	84(43)	99.7(59.8)
Completeness (%)	99.6(99.6)	99.6(97.7)
Redundancy	3.1	6.3
Refinement		
Resolution (Å)	39 - 3.0	38 - 1.5
Unique reflections	17587	69756
Rwork/Rfree (%)	22.8/28.3	14.9/16.7
Number of protein atoms	6413	3511
Number of water atoms	0	528
R.m.s.d. bond lengths (Å)	0.007	0.004
R.m.s.d. bond angles (˚)	1.35	1.232
Ramachandran favored (%)	95.52	98.65
Ramachandran allowed (%)	99.88	100
Ramachandran outliers (%)	0.12	0

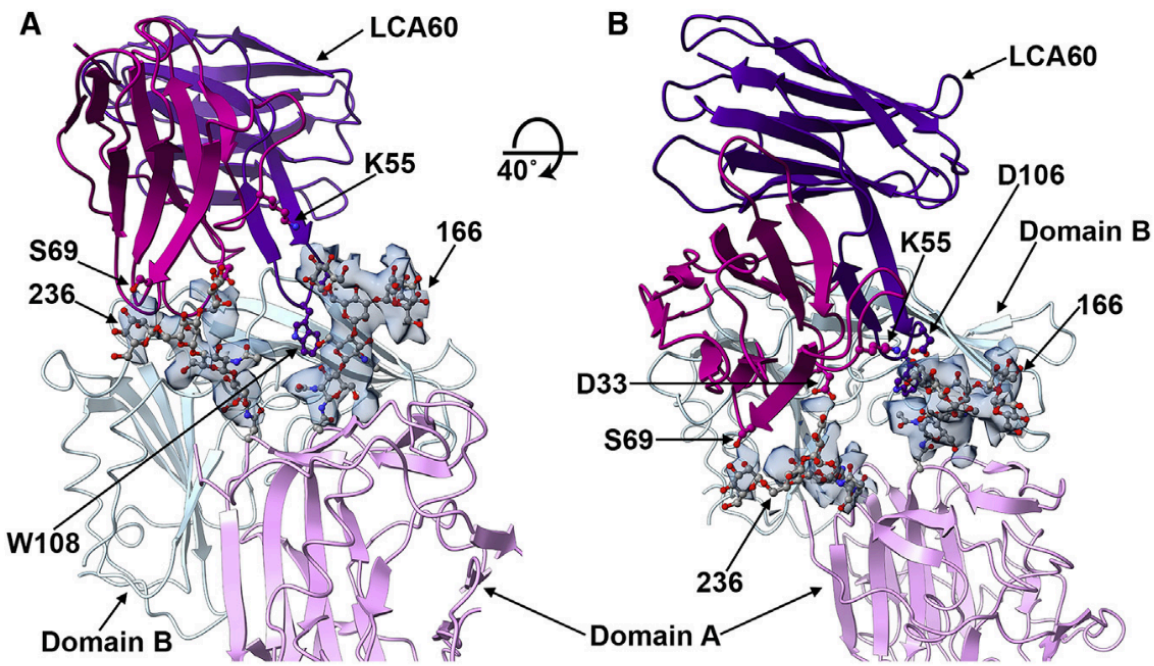
[a]Numbers in parentheses refer to outer resolution shell

Figure IV.XI: LCA60 interactions with the MERS-CoV S receptor binding domain



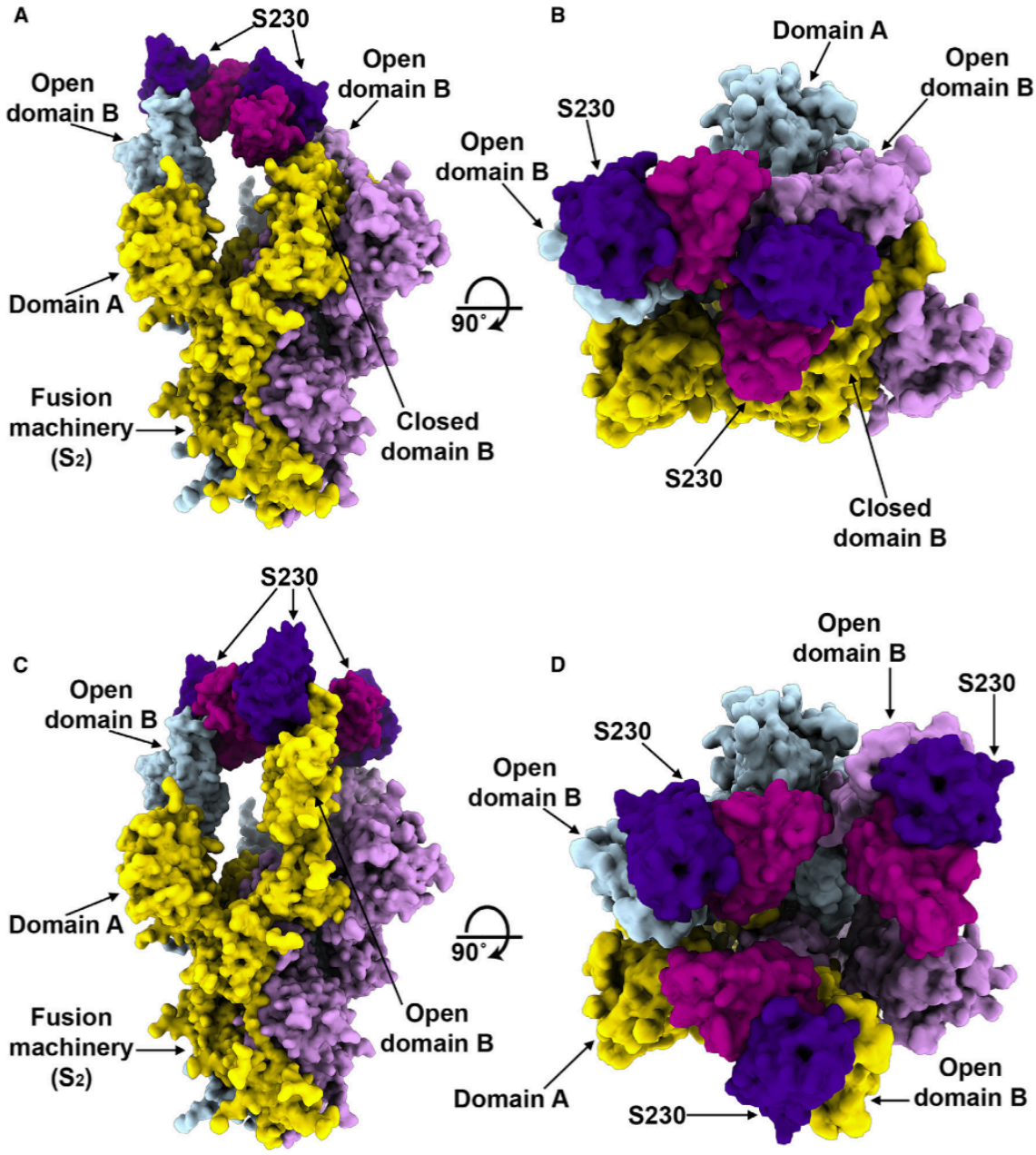
(A and B) Quasi-orthogonal views of the LCA60Fab binding to a closed B domain. The A domain from a neighboring protomer is also shown. (C) Zoomed-in view of the LCA60 CDRH3-mediated contacts. Residues involved in key interactions are shown in ball-and-stick representation colored by atom type (blue: nitrogen, red: oxygen, gray: carbon).(D) LCA60 Fabs bound to closed B domains in the state 1 structure. The glycan N-linked to N487 of one closed B domain interact with the LCA60 Fab bound to the neighboring closed B domain. (E and F) LCA60 (E) and DPP4 (F) would clash upon binding to MERS-CoV S and share partially over-lapping epitopes on the B domain. In (A), (B), (D), and (F), S is rendered as molecular surfaces, the Fabs as ribbon diagrams (only the variable domains are shown), and the glycans as blue spheres. In (C), S is rendered as a ribbon diagram. The color scheme is identical to Figure IV.VII. Selected glycans are labeled based on the N-linked glycosylation sequon numbering

Figure IV.XII: LCA60 interactions with MERS-CoV S involved N-linked glycans



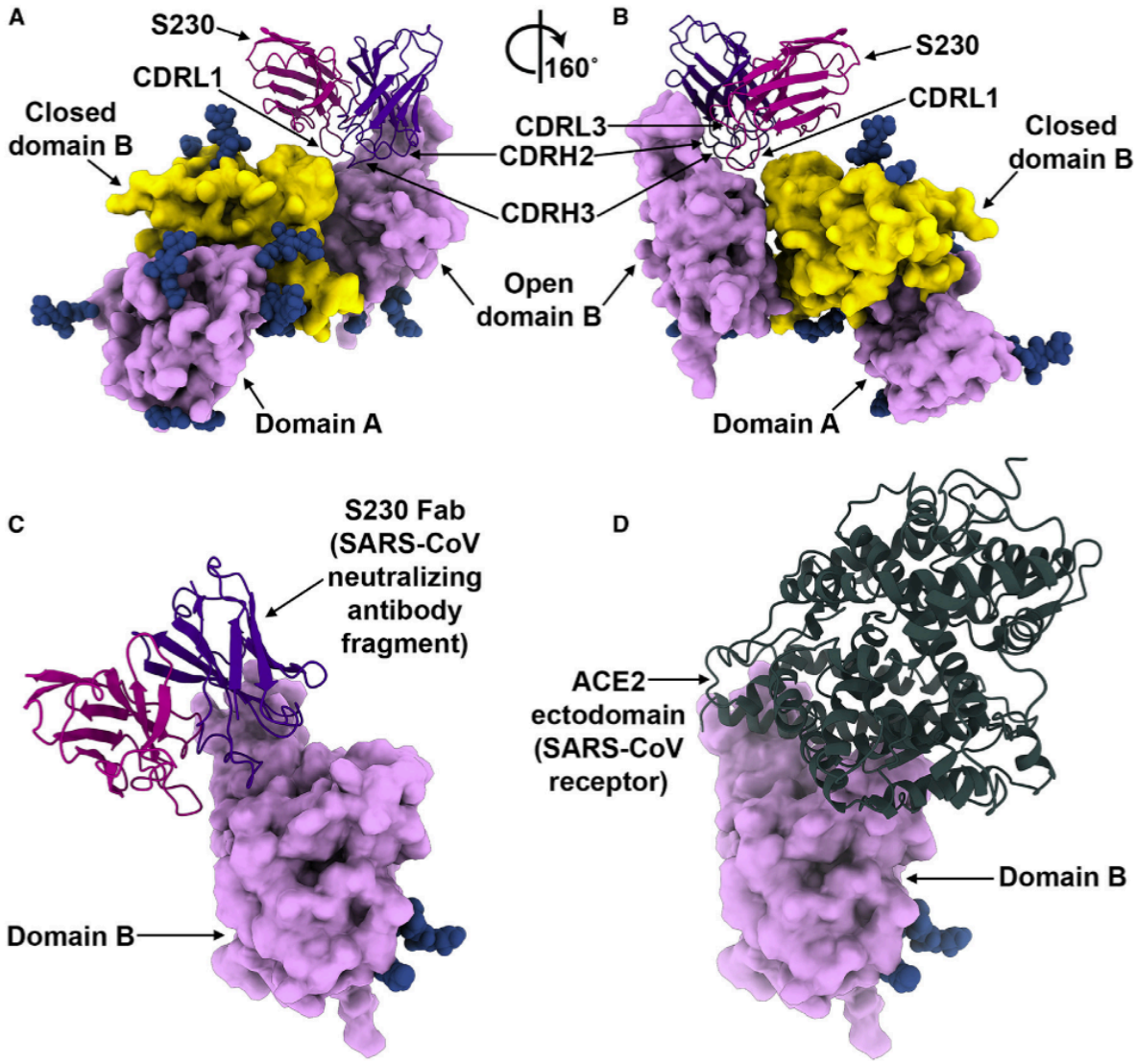
(A-B) Ribbon diagrams of LCA60 binding to a closed B domain. The A domain from a neighboring protomer is also shown. LCA60 residues interacting or putatively interacting with the glycans at positions N236 and N166 are shown in ball and stick representation (blue: nitrogen, red: oxygen, gray: carbon). LCA60atoms are colored identically except for carbon atoms that are pink/purple. The color scheme is identical to **Figure IV.VII**. Selected glycans are labeled based on the N-linked glycosylation sequon numbering. Only the LCA60 variable domains are shown. The cryoEM density corresponding to glycans at positions N166 and N236 is shown as a transparent blue surface.

Figure IV.XIII: CryoEM Structure of the SARS-CoV S Glycoprotein in Complex with the S230 Neutralizing Antibody



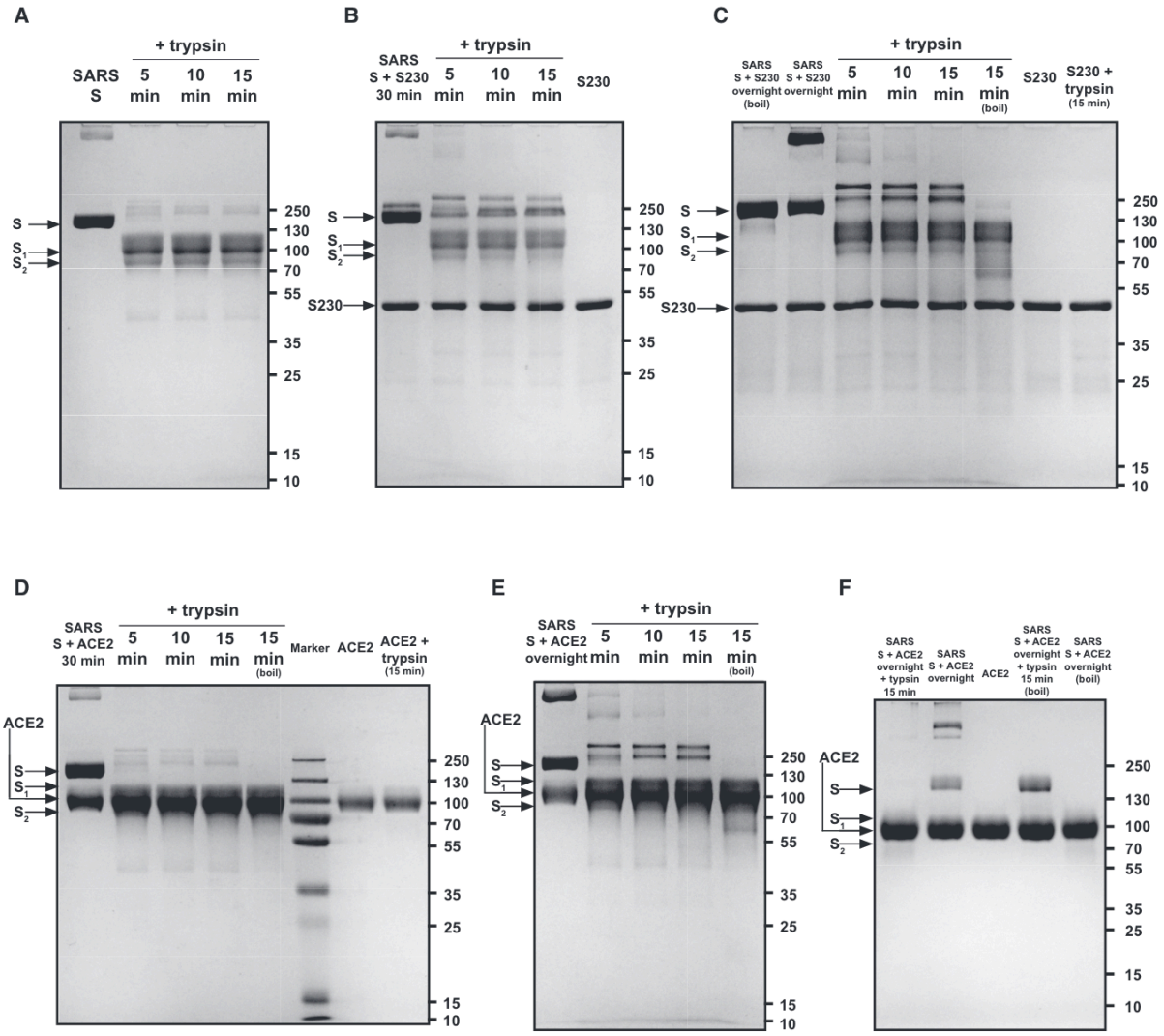
(A and B) Orthogonal views of the state 1 structure with one open, one partially open, and one closed B domain. (C and D) Orthogonal views of the state 2 structure with three open B domains that do not follow 3-fold symmetry. The structures are rendered as molecular surfaces with different colors for each S protomer (light blue, plum, and gold) and the LCA60 Fab heavy (purple) and light(pink) chains (only the variable domains are shown).

Figure IV.XIV: S230 interactions with the SARS-CoV S receptor binding domain



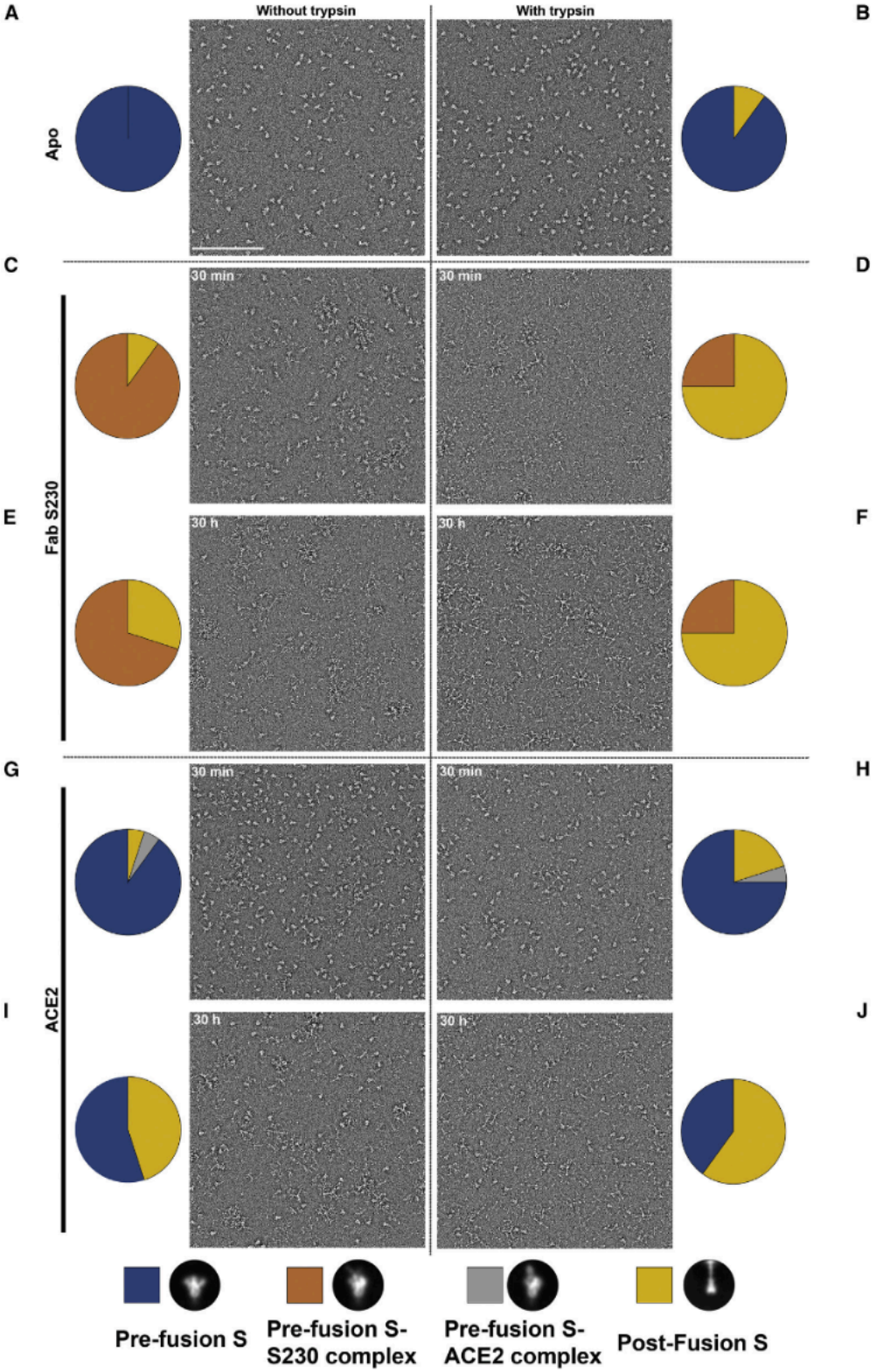
(A and B) Two views of the S230 Fab binding to an open B domain. The A domain of the same protomer and B domain from a neighboring protomer are also shown. (C and D) S230 (C) and ACE2 (D) would clash upon binding to SARS-CoV S and share partially overlapping epitopes on the B domain. In (A)–(D), S is rendered as molecular surfaces, the Fab as ribbon diagrams (only the variable domains are shown), and the glycans as blue spheres. The color scheme is identical to **Figure IV.XIII**. Selected glycans are labeled based on the N-linked glycosylation sequon numbering.

Figure IV. XV: SDS-PAGE analysis of SARS-CoV S cleavage by trypsin, under limited proteolysis conditions, in the presence and absence of S230 or ACE2



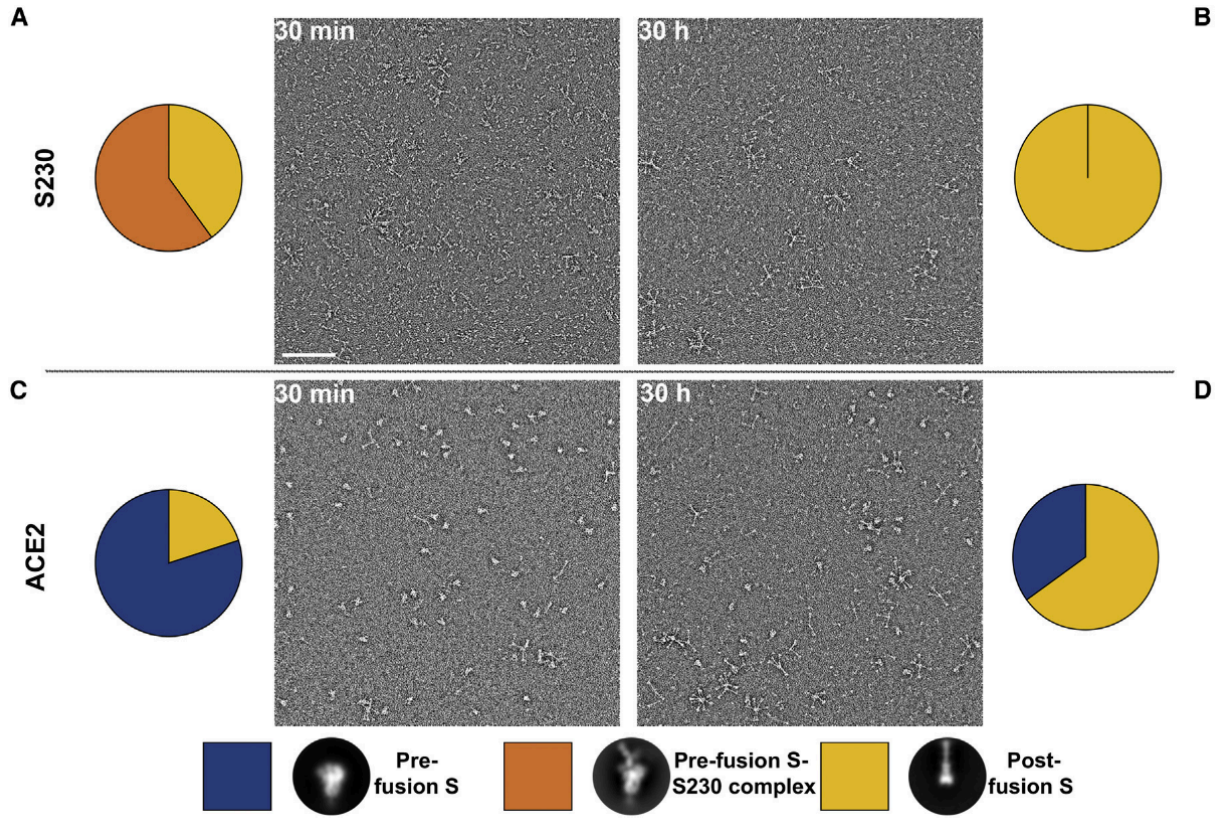
(A-C) Wild-type SARS-CoV S ectodomain trimer at 7.5mM (protomer concentration) was cleaved with 1.6mg/mL of trypsin either directly (A), or after incubation with an equimolar concentration of Fab S230 for 30 min (B) or 30 hr (C). (D-F) Wild-type SARS-CoV S ectodomain trimer at 7.5mM (protomer concentration) was cleaved with 1.6mg/mL of trypsin after incubation with an equimolar concentration of the ACE2 ectodomain for 30 min (D) or 30 hr (E) or after incubation with 30mMof the ACE2 ectodomain for 30 hr (F). Boil indicates that the sample was boiled for one minute in SDS-PAGE loading buffer prior to migration.

Figure IV. XVI: S230 or ACE2 mediated refolding of SARS-CoV S visualized by negative staining EM



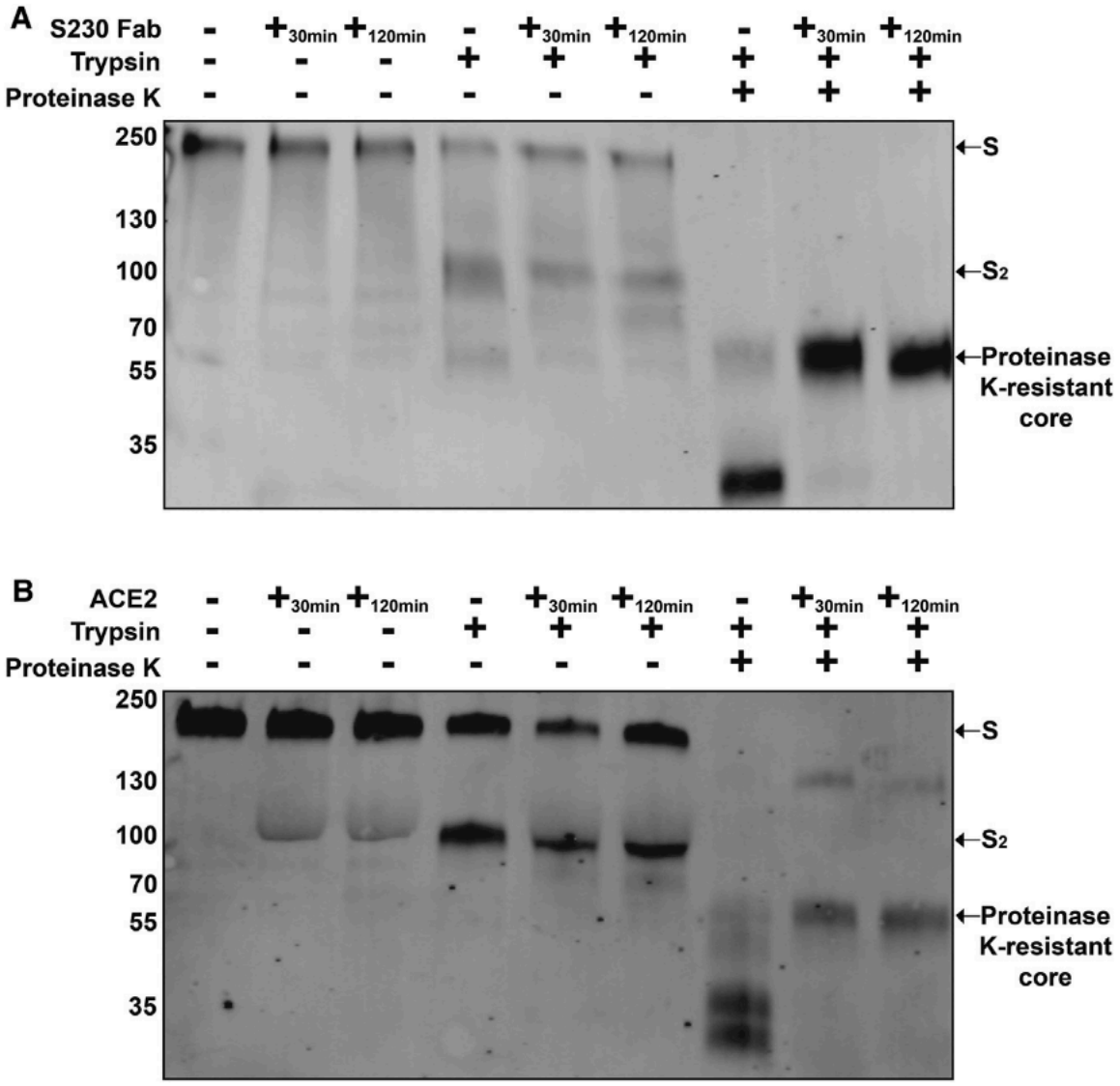
(A–J) The effect of Fab S230 (C–F) or of the ACE2 ectodomain (G–J) on the conformational state of the wild-type SARS-CoV S ectodomain trimer was analyzed using single-particle EM of negatively stained samples. Estimates of the fraction of particles corresponding to each state are represented as pie charts based on the number of particle images clustered in each group after reference-free 2D classification. Only pre-fusion and postfusion conformations were included in the calculations, and the results are rendered using bins with a minimal width of 5%. 1.6mg/mL trypsin was used for (B), (D), (F), (H), and (J). Time points shown at 30 min (C–D and G–H) or 30 h (E, F, I, and J). Representative 2D class averages corresponding to the different states discussed are shown at the bottom

Figure IV.XVII: Effect of S230 or ACE2 on Fusogenic Conformational Changes of Pre-cleaved SARS-CoV S



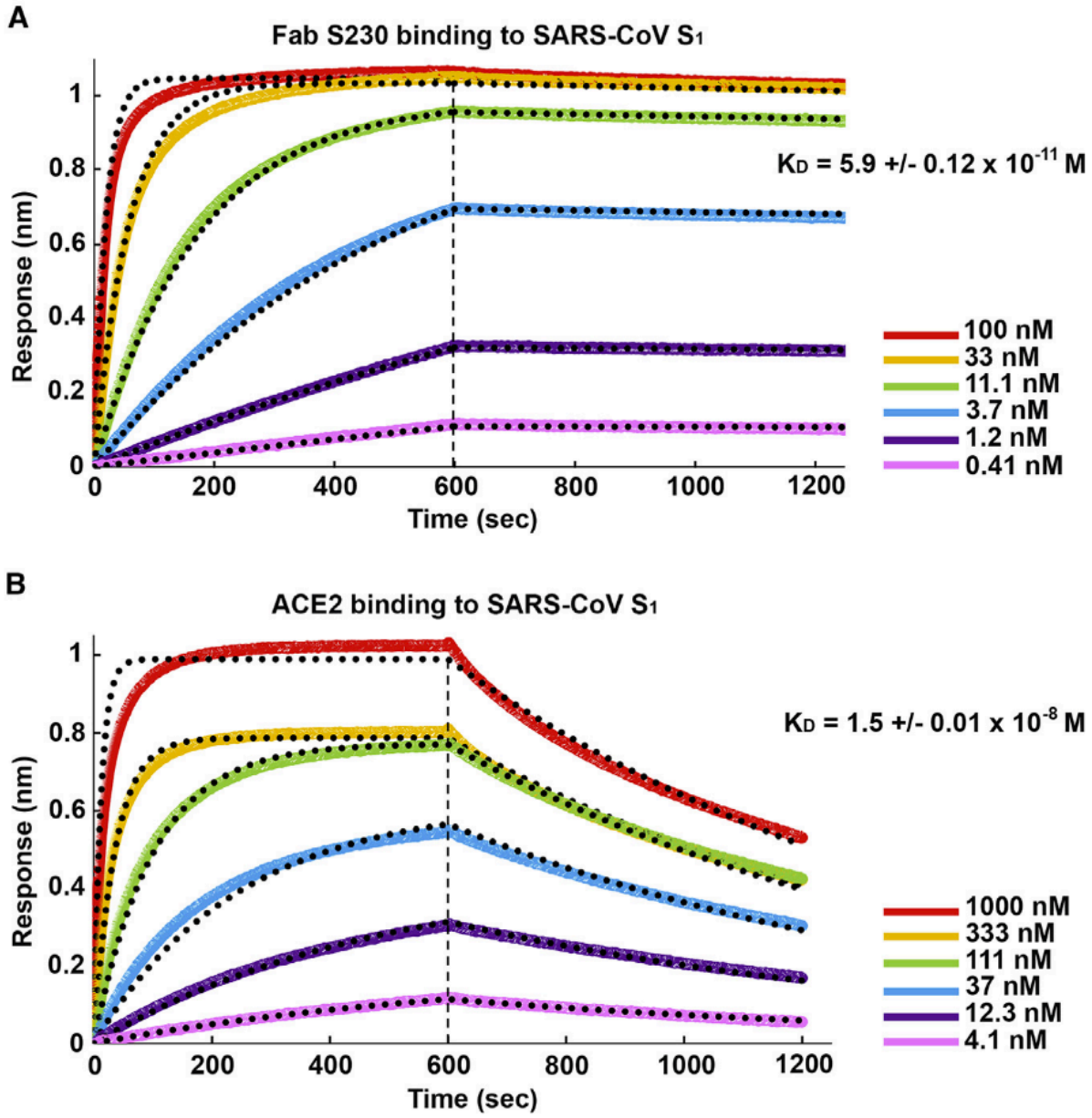
The effect of Fab S230 (A-B) or of the ACE2 ectodomain (C-D) on the conformational state of the wild-type SARS-CoV S ectodomain trimer was analyzed using single-particle electron microscopy of negatively stained samples. (A-D) Estimates of the fraction of particles corresponding to each state are represented as pie charts based on the number of particle images clustered in each group after reference-free 2D classification. Only prefusion and postfusion conformations were included in the calculations and the results are rendered using bins with a minimal width of 5%.

Figure IV.XVIII: ACE2 or S230 mediated refolding of the full-length membrane embedded SARS-CoV S trimer visualized by western blotting



(A and B) The effect of 0.5mM S230 Fab (A) or of10mM ACE2 ectodomain (B) on the conformational state of the full-length SARS-CoV S trimer embedded in the membrane of infectious MLV pseudovirus was probed by western blotting using an anti-SARS-CoV S2polyclonal antibody. Refolding to the postfusion conformation was detected by the appearance of a proteinase-K-resistant band migrating at approximately 55 kDa (Matsuyama and Taguchi, 2009). Trypsin was used at 5mg/mL and proteinase K at 10mg/mL. Digestion experiments and western blots were performed in triplicates, and a representative result is shown for each of them.

Figure IV.XIX: Biolayer Interferometry Analysis of S230 and ACE2 Binding to the SARS-CoV S_1 Subunit



The SARS-CoV S_1 subunit was immobilized using amine coupling on ARG2 biosensors and binding of the S230 Fab (A) or the ACE2 ectodomain (B) at multiple concentrations were monitored. Equilibrium constants of dissociation and estimated errors using a global fit are indicated. The concentrations of S230 Fab orACE2 ectodomain injected are indicated on each panel. Global fit curves are shown as black dotted lines. The vertical dashed lines indicate the transition between association and dissociation phases.

Chapter V- Coronavirus Spike proteins in complex with sialic acid receptors

Previously we had shown that structure and specifically CryoEM could be used to identify small molecule ligands on CoV S proteins (OC43 S in complex with 9-O-acetyl-sialic acid). This work was fleshed out by Tortorici et al. *NSMB.* 2019. The structure was obtained by a soaking method, by incubating a low-affinity ligand at incredibly high concentration (100mM) for a long time (more than 12 hours) and then preparing standard CryoEM grids and data collection methods. There also was previous work showing that MERS-CoV S could similarly bind sialic acids (Li et al. 2017) but with different specificity than OC43 S (Tortorici et al. 2019). The specificity seemed to be contributing to the viral tropism, but a structure was lacking since MERS-CoV S did not have the same pocket as OC43 S in the A domain. To uncover this, we combined CryoEM for structural analysis of five different sialic acid derivatives to identify and characterize the specificity seen in MERS-CoV tropism and mutational scanning to confirm the importance of the pocket for viral entry.

Adapted from: Park, Y.J., Walls, A.C., Wang, Z.Q., Sauer, M.M., Li, W., Tortorici, M.A., Bosch, J.J., DiMao, F., Veesler, D. *NSMB*. 2019. doi:10.1038/s41594-019-0334-7

V.I MERS-CoV S structure in complex with sialic acid

The Middle East respiratory syndrome coronavirus (MERS-CoV) was discovered in 2012 on the Arabian Peninsula as the causative agent of acute respiratory distress (and renal failure), with an unusually high fatality rate of 35% (Zaki et al. 2012). The vast majority of the 2,428 confirmed cases were reported in the Middle East although the virus was introduced into 27 countries. 163 new MERS-CoV cases were reported in 2019 and

there is currently no vaccine or cure for MERS or any coronavirus-induced diseases in humans. Zoonotic MERS-CoV transmission from dromedary camels, which is the host reservoir for this virus, along with human-to-human transmission, are the two known modes of infection (Haagmans et al. 2014; Sabir et al. 2016). Future emergence of (zoonotic) coronaviruses is likely to occur, given the large reservoir found in bats and birds, underscoring the importance of studying these pathogens (Gui et al. 2017; Hu et al. 2017; Menachery et al. 2015; Woo et al. 2006) .

MERS-CoV is an enveloped Nidovirus decorated with homotrimers of the spike (S) glycoprotein that mediates entry into host cells. S is the major antigen present at the viral surface and is the target of neutralizing antibodies during infection as well as the focus of vaccine design. Recent cryo-electron microscopy (cryoEM) structures of MERS-CoV and related coronavirus S ectodomain trimers provided snapshots of this key protein in prefusion (Gui et al. 2017; Kirchdoerfer et al. 2016; Shang et al. 2018; Walls et al. 2016; Walls et al. 2016a; Xiong et al. 2018; Yuan et al. 2017) and postfusion (Walls et al. 2017) conformations, receptor-bound states (Kirchdoerfer et al. 2018; Song et al. 2018; Tortorici et al. 2019), and in complex with neutralizing antibodies (Pallesen et al. 2017; Walls et al. 2019). MERS-CoV S is comprised of an N-terminal S_1 subunit which is folded as 4 domains (named A through D) and mediates attachment to dipeptidyl-peptidase 4 (DPP4, the host receptor) (Raj et al. 2013), and a C-terminal S_2 subunit that merges the viral and cellular membranes to initiate infection. MERS-CoV S is cleaved at the junction between the S_1 and S_2 subunits either during viral biogenesis or upon encounter of target cells (Earnest et al. 2017; Millet & Whittaker, 2014; Park et al. 2016). A second cleavage site, designated S_2', is found upstream of the fusion peptide in the S_2 subunit and is processed

at the onset of membrane fusion. Both cleavage sites participate in enhancing viral entry and modulating host range and cell tropism (Millet & Whittaker, 2014; Park et al. 2016; Yang et al. 2015).

Sialic acids (derivatives of neuraminic acid) are ubiquitous carbohydrates found as terminal residues on glycoproteins and glycolipids decorating the surface of eukaryotic cells (Stencel-Baerenwald et al. 2014). Neuraminic acid modifications along with formation of specific glycosidic linkages result in a wide chemical variety of sialoglycoconjugates across cell types, tissues and animal species. As a result, differential sialoside recognition can profoundly impact zoonotic transmission, tropism and virulence of many viruses. For instance, few amino acid substitutions in influenza virus hemagglutinins account for the preference switch from avian enteric tract to human respiratory tract sialylated receptors (Liu, J. et al. 2009; Xiong et al. 2013; Xiong et al. 2013a) .

MERS-CoV primarily infects human lung epithelial cells upon interacting with DPP4 (Raj et al. 2013). Crystal structures of the MERS-CoV S domain B in complex with the DPP4 ectodomain (Lu et al. 2013; Wang, N. et al. 2013) with cryoEM structures of the MERS-CoV S ectodomain trimer (Pallesen et al. 2017; Walls et al. 2019; Yuan et al. 2017), furthered our understanding of the mechanism of DPP4 engagement. In addition to attachment to DPP4, we recently showed that MERS-CoV infection of human airway epithelial cells involves low-affinity interactions with sialosides, using the S glycoprotein domain A, as depletion of sialic acid from the cell surface dampened viral entry (Li et al. 2017). A binding preference was found for α2,3-linked over α2,6-linked sialosides and these interactions were hindered by 9-O-acetylation or 5-N-glycolylation of the terminal

neuraminic acid (Li et al. 2017). Furthermore, we found that MERS-CoV S sialylated receptors are abundant in the camel nasal respiratory epithelium and the human lung alveoli, which coincides with DPP4 expression and the sites of MERS-CoV replication in these mammals (Li et al. 2017; Widagdo et al. 2016). As sialoside modifications, linkages and distribution vary among and within host species, the selectivity of MERS-CoV S for certain sialoside glycotopes may provide a determinant of host and tissue tropism of this zoonotic pathogen. These findings provided a plausible rational for explaining that equine DPP4 could support MERS-CoV infection of cultured cells although horses were resistant to experimental MERS-CoV infection (despite the presence of DPP4 in their respiratory tract (Barlan et al. 2014; Meyer et al. 2015; Vergara-Alert et al.2017). Although sialoside attachment appears to be a key step modulating MERS-CoV infection, the location of the sialoside-binding site, the interactions involved in ligand recognition, and the molecular basis for receptor specificity remain unknown. This information is crucial for understanding viral tropism and infectivity, assessing the zoonotic potential of MERS-CoV and related coronaviruses, and providing a blueprint for the design of coronavirus inhibitors.

To understand the structural basis of MERS-CoV attachment to and specificity for host sialosides, we determined cryoEM structures of the S glycoprotein ectodomain trimer in complex with 5-N-acetyl neuraminic acid (Neu5Ac), 5-N-glycolyl neuraminic acid (Neu5Gc), sialyl-Lewis^X (sLeX), α2,3-sialyl-N-acetyl-lactosamine (2,3-SLN) and α2,6-sialyl-N-acetyl-lactosamine (2,6-SLN) at 2.7 Å, 3.0 Å, 2.7 Å, 2.7 Å and 2.9 Å resolution, respectively. We demonstrate the receptor binds in a groove located at the surface of domain A that is distinct from the 9-O-acetyl sialoside-engagement site identified for

HCoV-OC43 and related β-1 coronavirus S glycoproteins (Hulswit et al. 2019; Tortorici et al. 2019). We further show the residues involved in Neu5Ac recognition are conserved across MERS-CoV isolates and essential for MERS-CoV S-mediated hemagglutination of human erythrocytes and entry into human airway epithelial cells. Our data rationalize MERS-CoV S attachment to neuraminic acids that are not 9-O-acetylated or 5-N-glycolylated and suggest an explanation for the selectivity for α2,3-linked over α2,6-linked sialosides. This study provides a structural framework for understanding MERS-CoV S sialoglycan receptor engagement and specificity, illuminates host range and cell tropism, and identifies a site of vulnerability accessible to neutralizing antibodies and small molecule inhibitors.

To identify the MERS-CoV S sialoside-recognition site, we determined a cryoEM structure of the MERS-CoV S 2P stabilized ectodomain trimer (Pallesen et al. 2017) incubated with 100 mM Neu5Ac at 2.7 Å resolution **(Figures V.I, V.II, V.III)**. The presence of the sialoside stabilized a large fraction of MERS-CoV S trimers in the closed state, with all three DPP4-binding B domains adopting a down conformation, obeying 3-fold symmetry. We did not detect major conformational changes compared to closed protomers of the LCA60 Fab-bound MERS-CoV S structure (Walls et al. 2019) (Cα r.m.s.d. of 1.35 Å for 1,145 aligned residues). The resolution estimate of our map is supported by the detection of ordered water molecules (Campbell et al. 2015) interacting with the S glycoprotein **(Figure V.II)** and by the observation of folic acid bound to S, as previously identified using X-ray crystallography and mass-spectrometry of isolated domain A (Pallesen et al. 2017). The density is equally-well resolved for the ligand and the surrounding S amino acid residues allowing unambiguous identification of the binding

site and docking of Neu5Ac in the map **(Figure V.II)**. The ligand interacts with a groove located at the periphery of domain A that is distinct from the HCoV-OC43 S 9-O-acetyl sialoside-binding site (Tortorici et al. 2019) **(Figures V.I, V.IV)**. The absence of N-linked glycans in the immediate vicinity of the binding groove likely facilitates unobstructed engagement of sialoglycoconjugates at the surface of target cells (**Figure V.IV)**. The ligand-binding site is located ~50 and ~75Å away from the DPP4-binding site of a neighboring B domain in the closed or open state, respectively (**Figure V.I)**.

The ligand C1-carboxylate electrostatically interacts with the Ser133 side chain hydroxyl, and the 5-nitrogen atom is hydrogen-bonded to the Ile132 backbone carbonyl **(Figure V.I)**. The Neu5Ac hydroxyl group at position 8 is hydrogen-bonded to the Arg307 side chain guanidium whereas the hydroxyl group at position 9 interacts with the Arg307 guanidium as well as the Ala92 backbone amide and carbonyl oxygen via hydrogen-bonding **(Figure V.I)**. The ligand N-acetyl methyl inserts into a hydrophobic depression defined by residues Phe39, Phe101, Ile131 and Ile132 whereas the N-acetyl carbonyl oxygen is hydrogen bonded to the Gln36 side chain amide **(Figure V.I)**. The sialoside buries 288 $Å^2$ of it surface upon MERS-CoV S engagement, corresponding to 61% of the ligand total accessible surface area, with most interactions occurring on the same side of the ligand. Bioinformatics analysis of MERS-CoV isolates show that the residues contacting the sialoside or participating in the formation of the binding groove are conserved. Specifically, Gln36, Phe39, Phe101, Ile132 and R307 are strictly conserved in the 284 MERS-CoV isolates sequenced to date. The H91Y substitution is observed in 4 MERS-CoV isolates (Human/Korea/Seoul/168-1-2015 GB: ALB08311, Human/Korea/Seoul/168-2-2015 GB: ALB08322, Human/KOR/Seoul/014-2015 GB:

ANC28634.1 and Human/Riyadh_1337/KSA/2014 GB: AMW90853.1), the S133N and S133R substitutions are present in two isolates (Human/England/4/2013 GB: AJD81440.1 and Hu/Quaseem-KSA-18012872/2018 GB: QBF80608.1, respectively) and the Q304K substitution is observed in 1 isolate (Camel/UAE/D1164.11/2014 GB: AJG44080.1).

The architecture of the binding groove suggests 9-*O*-acetylation of the neuraminic acid could not be sterically accommodated **(Figure V.V)**, in agreement with the previously reported distinct specificity of MERS-CoV S compared to HCoV-OC43 S (Li et al. 2017). To assess if 5-*N*-glycolylation of the neuraminic acid could be tolerated, we determined a cryoEM structure of MERS-CoV S 2P incubated with 100 mM Neu5Gc at 3.0 Å resolution **(Figures V.II, V.III, V.V, V.VI).**

Comparison of this map with the MERS-CoV S/Neu5Ac map shows the Neu5Gc density is weaker than that of the Neu5Ac and the glycolyl hydroxyl group could not be resolved, even after extensive focused 3D classification of the data **(Figure V.V)**. This observation is in line with the reported MERS-CoV S binding preference to Neu5Ac compared to Neu5Gc based on the lack of hemagglutination of (Neu5Gc-rich) horse erythrocytes and absence of binding to Neu5Gc-containing sialosides on a glycan array (Li et al. 2017). The structure suggests the MERS-CoV S hydrophobic pocket defined by Phe39, Phe101, Ile131 and Ile132 could sterically and/or chemically disfavor accommodation of the 5-N-glycolyl hydroxyl group in the conformation observed in virtually all Neu5Gc-containing protein complex structures **(Figure V.V).** The high concentrations of sialosides used for the cryoEM experiments, however, allowed to visualize Neu5Gc bound to MERS-CoV S, much in the same way that crystal structures

of influenza virus hemagglutinins in complex with mismatched receptor analogues or of rhesus rotavirus VP8* bound to Neu5Gc were obtained in spite of very weak or lack of detectable interactions in solution (Xiong et al. 2013; Xiong et al. 2013a; Yu et al. 2012). Since the equine respiratory tract is abundantly decorated with Neu5Gc-containing oligosaccharides (Suzuki et al. 2000), these data collectively suggest a plausible explanation for the resistance of horses to experimental MERS-CoV infection despite the fact MERS-CoV S can use equine DPP4 for promoting entry into target cells *in vitro* (Li et al. 2017; Meyer et al. 2015; Vergara-Alert et al. 2017) .

To evaluate the importance of the interactions identified by cryoEM for sialoside recognition, we substituted with alanine individual MERS-CoV S amino acid residues interacting with Neu5Ac and assessed the effect on domain A-mediated hemagglutination of human erythrocytes. As previously described, the wild-type MERS-CoV S domain A fused to human immunoglobulin Fc only promoted hemagglutination of human erythrocytes when multimerized on lumazine synthase nanoparticles (Li et al. 2017), indicative of enhancement of low-affinity interactions via avidity **(Figures V.V, V.VII, V.VIII)**. Complete inhibition of hemagglutination was observed, however, for each binding-site mutant tested, i.e F39A, H91A, I132A, S133A and R307A **(Figures V.VII, V.VIII)**. These results validated the structural observations by demonstrating that recognition of Neu5Ac via these residues is necessary for MERS-CoV S-mediated attachment to sialosides.

To further assess the functional relevance of the identified interactions, we interrogated the impact of individual domain A substitutions on MERS-CoV S-mediated infectivity. We used the murine leukemia virus platform with a luciferase reporter (Millet &

Whittaker, 2016) for quantifying entry of particles pseudotyped with either wildtype or mutant MERS-CoV S into human airway Calu-3 cells. The S133A and R307A substitutions abrogated entry of pseudotyped particles due to disruption of the aforementioned electrostatic interactions involving the Neu5Ac C1 carboxylate and glycerol side chain, respectively (**Figures V.II, V.III)**. The F39A and H91A mutants showed reduced infectivity of the pseudotyped particles by more than 90%, likely due to disruption of the ligand-binding groove and loss of favorable Van der Waals or electrostatic interactions (**Figures V.II, V.III)**. These findings demonstrate the key role of the identified residues for interacting with Neu5Ac in the context of a full-length, membrane-embedded MERS-CoV S glycoprotein and show that attachment to sialoside receptors using the binding site identified by cryoEM is essential for promoting MERS-CoV S-mediated entry into human airway Calu-3 cells. Furthermore, as sialidase treatment of Calu-3 cells (Li et al. 2017) or mutations of the S sialoside-recognition site abrogated MERS-CoV S-mediated entry, it is likely that sialoside attachment either precedes DPP4 engagement or occurs simultaneously.

We previously showed that the MERS-CoV S domain A preferentially interacts with α2,3-linked, compared to α2,6-linked sialosides, with sulfated sLeX being the preferred binder (Li et al. 2017). To understand the architectural principles underlying this specificity, we determined three cryoEM structures of MERS-CoV S 2P in the presence of 50 mM sLeX **(Figures V.II, V.IX, V.III),** 70 mM 2,3-SLN **(Figures V.II, V.III, V.X)** or 70 mM 2,6-SLN **(Figures V.II, V.III, V.XI)** at 2.7 Å, 2.7 Å and 2.9 Å resolution, respectively.

The cryoEM reconstruction obtained for MERS-CoV S in complex with sLeX resolves the terminal Neu5Ac linked to galactose, via a 2,3-glycosidic bond adopting a

cis configuration, as well as the N-acetyl glucosamine (GlcNAc) and fucose rings at lower contour level due to conformational heterogeneity of the ligand **(Figure V.XII).** Engagement of the Neu5Ac moiety of the sLeX involves identical interactions to those observed in our structure of MERS-CoV S in complex with Neu5Ac **(Figure V.I)**. This finding and the detection of a galactose moiety linked to the oxygen at position 2 of Neu5Ac validate our docking of the ligand in the cryoEM density. The sLeX trisaccharide core (Neu5Ac(α2-3)galactose(β1-4)GlcNAc) adopts a roughly linear conformation when bound to MERS-CoV S **(Figure V.XII)**, as is the case for influenza virus hemagglutinins in complex with α2,3-sialosides (Liu et al. 2009; Xiong, X. et al. 2013; Xiong et al. 2013a). The MERS-CoV S structure in complex with 2,3-SLN shows a very similar pose of the ligand to that observed for sLeX although the GlcNAc is even more conformationally heterogeneous in the trisaccharide compared to sLeX **(Figure V.XII)**.

The cryoEM map obtained for MERS-CoV S bound to 2,6-SLN resolves the Neu5Ac and galactose moieties, linked via a 2,6-glycosidic bond adopting a cis configuration, and the GlcNAc ring at lower contour level due to conformational variability of the ligand **(Figure V.XII)**. The 2,6-SLN has a U-shaped conformation, comparable to reports on influenza virus hemagglutinins (Xiong et al. 2013; Xiong et al. 2013a), that positions the GlcNAc moiety away from the binding-groove and the rest of the S glycoprotein **(Figure V.XII)**.

The cryoEM structures suggest an explanation for the finding that sulfated sLeX is the preferred MERS-CoV S sialoside ligand (Li et al. 2017). Although the sLeX conformational dynamic limited its resolvability, the galactose and the GlcNAc moieties appear to be positioned within contact distance of the MERS-CoV S domain A due to the

ligand conformation **(Figure V.XII)**. Furthermore, sulfation of the GlcNAc hydroxyl group at position 6 could putatively allow formation of electrostatic interactions with the Gln304 side chain or neighboring residues **(Figure V.XII)**, as described for influenza virus A/Vietnam/1194/2004 hemagglutinin (Xiong et al. 2013b). The bent architecture of 2,6-SLN, however, orients the GlcNAc away from the binding site, hindering contacts with S (which appear limited to the Neu5Ac and galactose moeities). We therefore hypothesize that the trend of preferential attachment to α2,3- over α2,6-sialosides results from enhanced interactions with the former type of ligands, leading MERS-CoV to target cell types presenting such sialylated receptors at their surface.

We show here that sialosides bind to MERS-CoV S using a distinct site than the one observed for HCoV-OC43 S in complex with 9-O-acetyl sialylated receptors (Hulswit et al. 2019; Tortorici et al. 2019), despite the fact the bound ligands are only separated by a few angstroms from each other when superimposing the two structures. Although the β-sandwich architecture of domain A is conserved among all coronaviruses, distinct variable loops are grafted onto it, and neither the MERS-CoV S nor the HCoV-OC43 S sialoside-binding sites are conserved in the infectious bronchitis virus S glycoprotein (Shang et al. 2018), which is known to engage α2,3-sialosides at the surface of avian host cells (Wickramasinghe et al. 2011). These findings suggest coronavirus S glycoproteins have independently evolved the ability to recognize sialosides using domain A, similarly to their ability to engage different proteinaceous receptors via modifications of domain B. Our results therefore set coronavirus S glycoproteins apart from influenza virus A hemagglutinins, which recognize the Neu5Ac moiety through a largely conserved binding site (Xiong et al. 2013; Xiong et al. 2013). We hypothesize that coronaviruses that

utilize both an entry receptor and an attachment receptor for infection, such as MERS-CoV or transmissible gastroenteritis virus, have more evolutionary freedom with respect to their ability to attach to sialoglycans than viruses only using an entry receptor. This could have led to the independent loss and/or acquisition of a sialoside-recognition function as an adaption to new hosts upon cross-species transmission of distinct viruses. The ability to interact with sialoglycoconjugates might be a recent MERS-CoV S acquisition, as neither HKU4 S_1 nor HKU5 S_1 could hemagglutinate human erythrocytes (Li et al. 2017). This modular evolutionary plasticity underlies coronavirus cross-species transmission, switching of host cell types and viral fitness.

We previously suggested MERS-CoV S interacts with sialoside receptors through lower affinity interactions than influenza A/California/04/2009 hemagglutinin (containing the T200A and E227A substitutions), as multimerization with a lumazine synthase nanoparticle was required to detect binding of a dimeric MERS-CoV S domain A-Fc construct but not of a dimeric hemagglutinin ectodomain-Fc construct (Li et al. 2017). Since influenza A virus utilizes neuraminidase to promote release of progeny virions from host cells, off-target cells or decoys, the enhanced affinity for sialosides is balanced with a receptor-destroying enzyme. Similarly, HCoV-OC43 possesses a hemagglutinin esterase with receptor-destroying activity (Bakkers et al. 2017) to achieve the same function. Since MERS-CoV does not encode an equivalent enzyme, the binding dynamic of viral particles to sialoside receptors is entirely based on the energetics of these reversible interactions (Li et al. 2017). The strong dependence on interactions with sialosides for MERS-CoV entry into human airway Calu-3 cells is therefore likely a result of the high concentration of sialylated receptors (Li et al. 2017), favoring virus adsorption

through avidity and potentially increasing the likelihood of binding to DPP4, leading to viral entry. Similar roles have been proposed for carcinoembryonic antigen-related cell adhesion molecule 5 (Chan et al. 2016) and binding immunoglobulin protein (Chu et al. 2018).

Although MERS-CoV S is densely decorated with N-linked glycans, the sialoside-recognition site is unobstructed, a possible requirement for receptor engagement. This finding is in marked contrast with the conformational masking and/or glycan shielding of the domain B entry receptor-binding motifs observed for several coronavirus S glycoproteins, including MERS-CoV S (Pallesen et al. 2017; Song et al. 2018; Walls et al. 2016; Walls et al. 2019; Yuan et al. 2017). We recently identified an antibody blocking domain A-mediated hemagglutination of erythrocytes with *in vitro* neutralizing activity of MERS-CoV S pseudovirions and partial *in vivo* prophylactic protection of mice challenged with a lethal dose of MERS-CoV (Widjaja et al. 2019). These findings, along with our mutagenesis data, indicate the sialoside-binding groove identified here represents a key site of vulnerability to inhibitors of MERS-CoV infection that could be targeted for the future development of therapeutics.

V.II Methods for coronavirus spike proteins in complex with sialic acid receptors

Cell lines. HEK293F is a cell line adapted to grow in suspension (Life Technologies). HEK293T/17 cells (ATCC, catalog number: CRL-11268) and human hepatic Huh-7 cells (National Institutes of Biomedical Innovation, Health and Nutrition, catalog number: JCRB0403) are adherent cells. Cell lines were not authenticated or tested for mycoplasma contamination.

Protein expression and purification. The MERS-CoV S 2P ectodomain was produced in 500mL HEK293F cells grown in suspension using FreeStyle 293 expression medium (Life technologies) at 37°C in a humidified 8% CO2 incubator rotating at 130 r.p.m. The cultures were transfected using 293fectin (ThermoFisher Scientific) with cells grown to a density of 1 million cells per mL and cultivated for three days. The supernatants were harvested and cells resuspended for another three days, yielding two harvests. Clarified supernatants were purified using a 5mL Cobalt affinity column (Takara). Purified protein was concentrated and flash frozen in Tris-saline (50 mM Tris pH 8.0, 150 mM NaCl) prior to cryo-EM analysis.

Protein A-fused lumazine synthase nanoparticles (a self-assembling 60-meric lumazine synthase nanoparticle, N-terminally extended with the immunoglobulin Fc-binding domain of the *Staphylococcus aureus* protein A) and Fc-tagged domain A (residues 19–357) or Fc-tagged domain B (residues 358-588) of the MERS-CoV S protein (GB: YP_009047204.1) were expressed and purified as previously described[34]. Purified proteins were analyzed on a 12% SDS-PAGE gel under reducing conditions and stained with GelCodeBlue stain reagent (Thermo Scientific).

CryoEM sample preparation and data collection. Incubation of 1 μM MERS-CoV S 2P stabilized ectodomain trimer with 100 mM Neu5Ac (Millipore 110138), 100 mM Neu5Gc (Miilipore 362000), 50 mM sLeX (Carbosynth OS04058), 70 mM 2,3-SLN (Carbosynth OS06484) or 70 mM 2,6-SLN (Carbosynth OS09313) was performed overnight at 4ºC. We used 100 mM Neu5Ac based on our previous work on HCoV-OC43 S (Tortorici et al. 2019) and work on influenza virus (Xiong et al. 2013) whereas the lower concentrations used for larger ligands were determined empirically based on the contrast

of cryoEM micrographs. 3uL of 0.16 mg/mL MERS S-sialoglycan complex was loaded onto a freshly glow discharged (30 s at 20 mA) lacey carbon grid with a thin layer of evaporated continuous carbon prior to plunge freezing using a vitrobot MarkIV (ThermoFisher Scientific) using a blot force of -1 and 2.5 second blot time at 100% humidity and 25°C. 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 (Suloway et al. 2005) at a nominal magnification of 130,000x with a pixel size of 0.525 Å. The dose rate was adjusted to 8 counts/pixel/s, and each movie was acquired in super-resolution mode fractionated in 50 frames of 200 ms.

CryoEM data processing. Movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking and extraction were carried out using Warp (Tegunov & Cramer, 2019). Particle images were extracted with a box size of 800 binned to 400 yielding a pixel size of 1.05 Å. For each data set two rounds of reference-free 2D classification were performed using Relion (Zivanov et al. 2018) to select well-defined particle images. Subsequently, two rounds of 3D classification with 50 iterations each (angular sampling 7.5˚ for 25 iterations and 1.8˚ with local search for 25 iterations) were carried out using Relion without imposing symmetry to select closed trimers with the three B domains in the down conformation. 3D refinements imposing C3 symmetry were carried out using non-uniform refinement in cryoSPARC (Punjani et al. 2017) using an initial model generated by converting a modified version of PDB ID 6NB3 (Walls et al. 2019) with all three B domains closed to a map low-pass filtered to 30Å

resolution. Particle images were subjected to the Bayesian polishing procedure implemented in Relion3.0 (Zivanov et al. 2018) before performing another round of 3D refinement in Relion followed by per-particle defocus refinement and again 3D refinement in Relion. The particles were then subjected to 3D classification without refining angles and shifts using a soft mask delineating the ligand-binding site and with a tau value of 60. Final 3D refinements were carried out using non-uniform refinement in cryoSPARC. 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 and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution (Scheres & Chen, 2012).

Model building. We used UCSF Chimera (Goddard et al. 2007) to fit the MERS-CoV S atomic model (PDB ID 6NB3) into the cryoEM map followed by subsequent manually rebuilding using Coot (Brown et al. 2015; Emsley et al. 2010). N-linked glycans were hand-built into the density where visible, and then refined and relaxed using Rosetta (DiMaio et al. 2011; DiMaio et al. 2015; Frenz et al. 2017; Frenz et al. 2019; Walls et al. 2017; Wang et al. 2015). Glycan refinement relied on a dedicated Rosetta protocol, which uses physically realistic geometries based on prior knowledge of saccharide chemical properties (Frenz et al. 2019), and was aided by using both sharpened and unsharpened maps. The final models were analyzed using MolProbity (Chen et al. 2010), EMRinger (Barad et al. 2015), Privateer (Agirre et al. 2015) and PISA (Krissinel & Henrick, 2007). Figures were generated using UCSF ChimeraX (Goddard et al. 2018) .

Pseudovirus entry assay. MLV-based MERS-CoV S-pseudoviruses were prepared as previously described (Millet & Whittaker, 2016). HEK293T/17 cells were

cotransfected with a full-length MERS-CoV S or MERS-CoV S mutant encoding-plasmid, an MLV Gag-Pol packaging construct and the MLV transfer vector encoding a luciferase reporter using the Lipofectamine 2000 or 3000 transfection reagent (Life Technologies) according to the manufacturer's instructions. Cells were incubated for 5-12 hours at 37 °C with 8% CO_2 with transfection medium. Cells were then washed with DMEM 2X and DMEM containing 10% FBS was added for 72 hours. The supernatants were then harvested and filtered through 0.45-μm membranes prior to concentration with a 30kDa centrifugal concentrator membrane (Amicon).

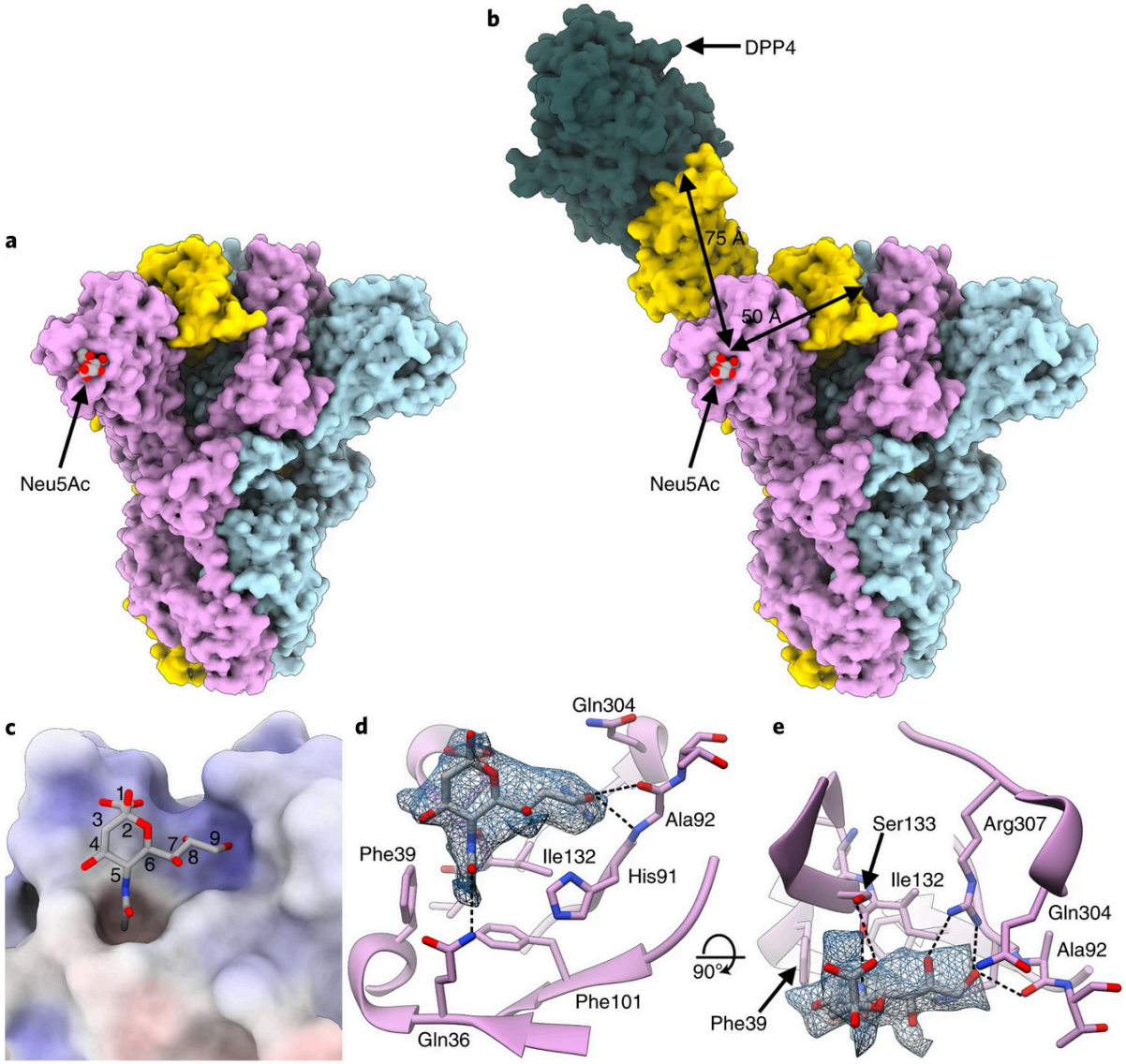
Calu3 cells grown in DMEM containing 10% FBS and 1% penstrep were plated into 96-well plates following trypsinization at a density of ~$3.2x10^4$ cells/mL and allowed to grow for 48 hours at 37 °C with 8% CO_2. These cells were washed 3X with DMEM prior to addition of 40μL of MERS-CoV S or MERS-CoV S mutant pseudovirus for 2 hours at 37 °C with 8% CO_2. Following the two-hour incubation, 40μL of 20%FBS and 2%PS was added to the wells and the plate was incubated for 72 hours. ONE-Glo-EX was added in equivolume (~75μL to account for evaporation) and the luciferase counts were read on a Varioskan plate reader.

Western Blotting. 4X SDS-PAGE loading buffer was added to all concentrated pseudovirus samples. The samples were run on a 4-15% gradient Tris-Glycine Gel (BioRad) and transferred to PVDF membranes. An anti-S_2 MERS-CoV S purified 1.6c7 primary antibody[49] (1:500 dilution) and an Alexa Fluor 680-conjugated goat anti-human secondary antibody (1:40,000 dilution, Jackson Laboratory) were used for Western-blotting. A LI-COR processor was used to develop images.

Hemagglutination assay. High-sensitivity nanoparticle-based hemagglutination assays were performed as previously described (Li et al. 2017). Briefly, wild-type or mutants MERS-CoV S domain A -Fc (starting at 6.77 μg per well) were complexed with protein A-fused lumazine synthase nanoparticles at a 0.6:1 molar ratio for 30 min on ice. Mixtures were subsequently serially diluted (2-fold dilutions) and mixed 1:1 with human erythrocytes (0.5% in phosphate buffered saline) in V-bottom, 96-well plates. Hemagglutination was scored after 2 h of incubation at 4 °C.

Figures Chapter V: V.I-V.XII: Adapted from Park et al. 2019

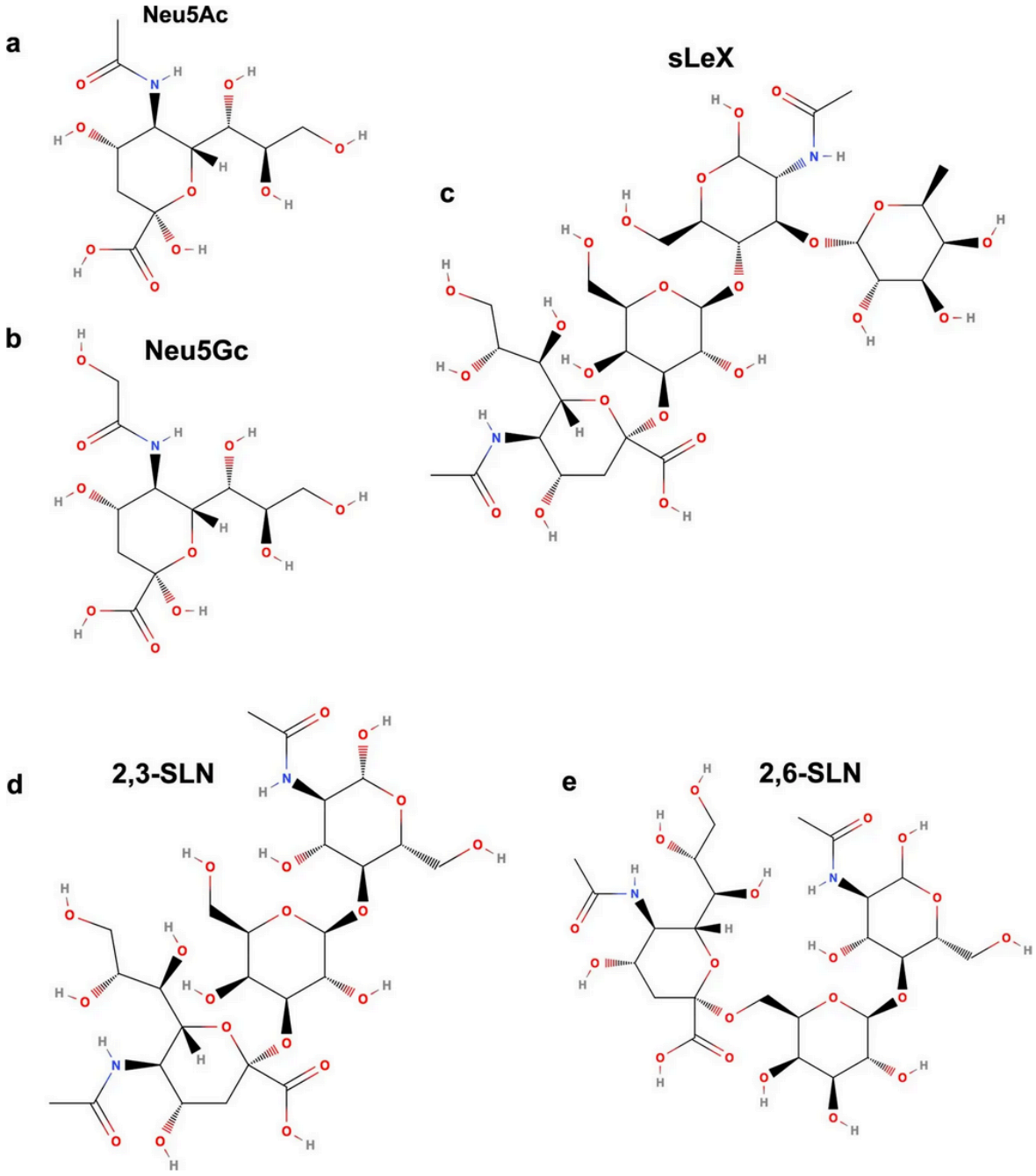
Figure V.I: Cryo-EM identification of a sialoside-binding site in the MERS-CoV S glycoprotein



a, Molecular surface representation of the MERS-CoV S ectodomain trimer, with each protomer in a different color. The Neu5Ac ligands are rendered as spheres. **b**, Spatial relationships between the attachment (sialoside) and entry (DDP4) receptor-binding sites are illustrated using a composite model obtained from the cryo-EM structure shown in **a**, the cryo-EM structure of MERS-CoV S bound to the LCA60 neutralizing antibody Fab fragment (PDB 6NB3)[20] and a crystal structure of the MERS-CoV S domain B bound to DPP4 (PDB 4KR0)[32]. The sialoside-binding site is located ~50 and ~75 Å away from the DPP4-binding site of a neighboring B domain in the closed or open state, respectively. **c**, Surface representation of the ligand-binding site colored by electrostatic potential from −12 (red) to +10 (blue) $k_B T/e_c$. Neu5Ac is shown in stick representation. **d**, **e**, Two

orthogonal views of the sialoside-binding site rendered as a ribbon diagram with the side chains of key surrounding residues shown as sticks. Neu5Ac is rendered as sticks and the corresponding region of cryo-EM density is shown as a blue mesh contoured at 5.5σ. Dashed lines show selected electrostatic interactions formed between MERS-CoV S amino acid residues and the ligand. In all panels, nitrogen and oxygen atoms are colored blue and red, respectively, while carbon atoms are colored gray (Neu5Ac) or pink (MERS-CoV S). Glycans in MERS-CoV S have been omitted for clarity.

Figure V.II: Chemical structures of the sialosides used for structural studies



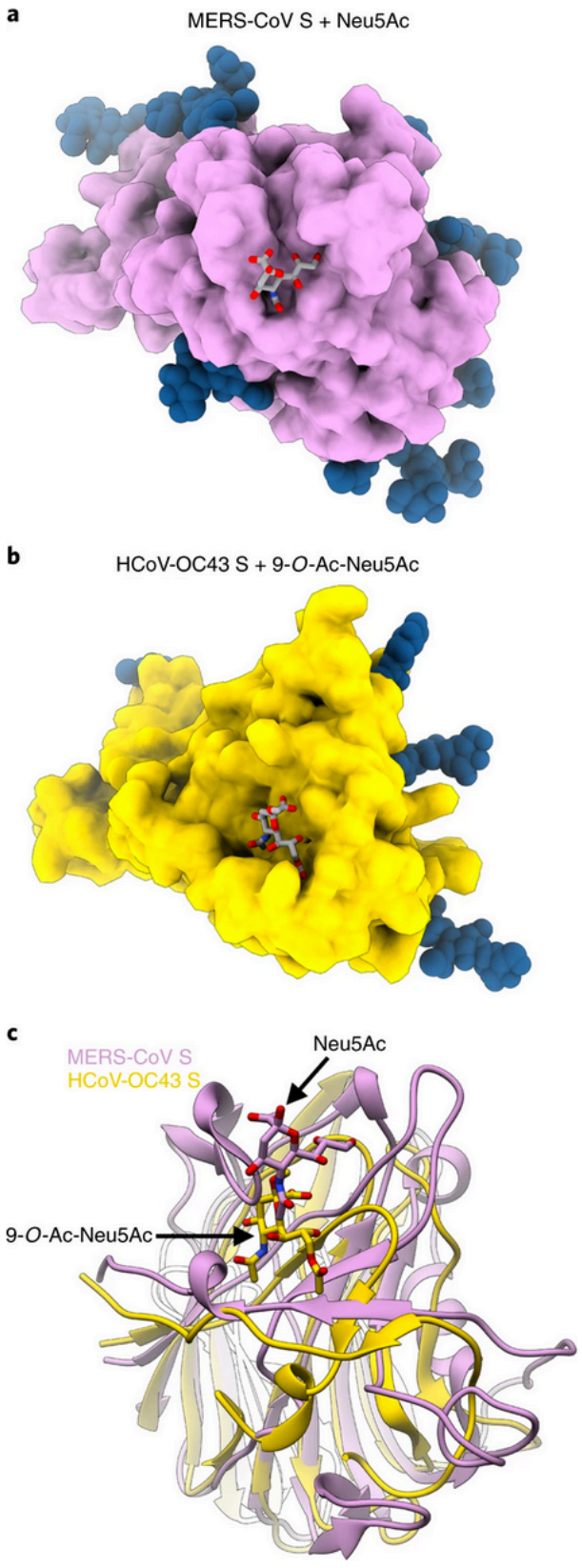
a, Neu5Ac. **b**, Neu5Gc. **c**, sLeX. **d**, 2,3-SLN. **e**, 2,6-SLN.

Figure V.III: Cryo-EM data collection and refinement statistics

	MERS-CoV S + Neu5Ac (EMD-20542, PDB 6Q04)	**MERS-CoV S + Neu5Gc (EMD-20829)**	**MERS-CoV S + sLeX (EMD-20543, PDB 6Q05)**	**MERS-CoV S + 2,3-SLN (EMD-20544, PDB 6Q06)**	**MERS-CoV S + 2,6-SLN (EMD-20545, PDB 6Q07)**
Data collection and Processing					
Magnification	130,000	130,000	130,000	130,000	130,000
Voltage (kV)	300	300	300	300	300
Electron exposure ($e^-/Å^{-2}$)	70	70	70	70	70
Defocus range (μm)	0.5-3.0	0.5-2.0	0.4-2.5	0.5-2.2	0.5-3.0
Pixel size (Å)	0.525	0.525	0.525	0.525	0.525
Symmetry Imposed	C3	C3	C3	C3	C3
Initial particle images	380,000	592,000	790,000	347,000	251,000
Final particle images	42,000	23,000	57,000	49,000	23,000
Refinement					
Map Resolution (Å)	2.7	3.0	2.7	2.7	2.9
FSC Threshold	0.143	0.143	0.143	0.143	0.143
Map sharpening B factor ($Å^2$)	-61	-57	-62	-59	-54
Model Composition					
Protein Residues	3,477		3,477	3,477	3,477
Ligands	6		6	6	6
Water molecules B factors ($Å^2$)	234		195	216	60
Protein	12.1		11.1	12.0	16.2
Ligands	19		16.7	18.9	24.6
Water Molecules	9.6		8.9	9.1	14
R.m.s. deviations					
Bond Lengths (Å)	0.026		0.026	0.026	0.026

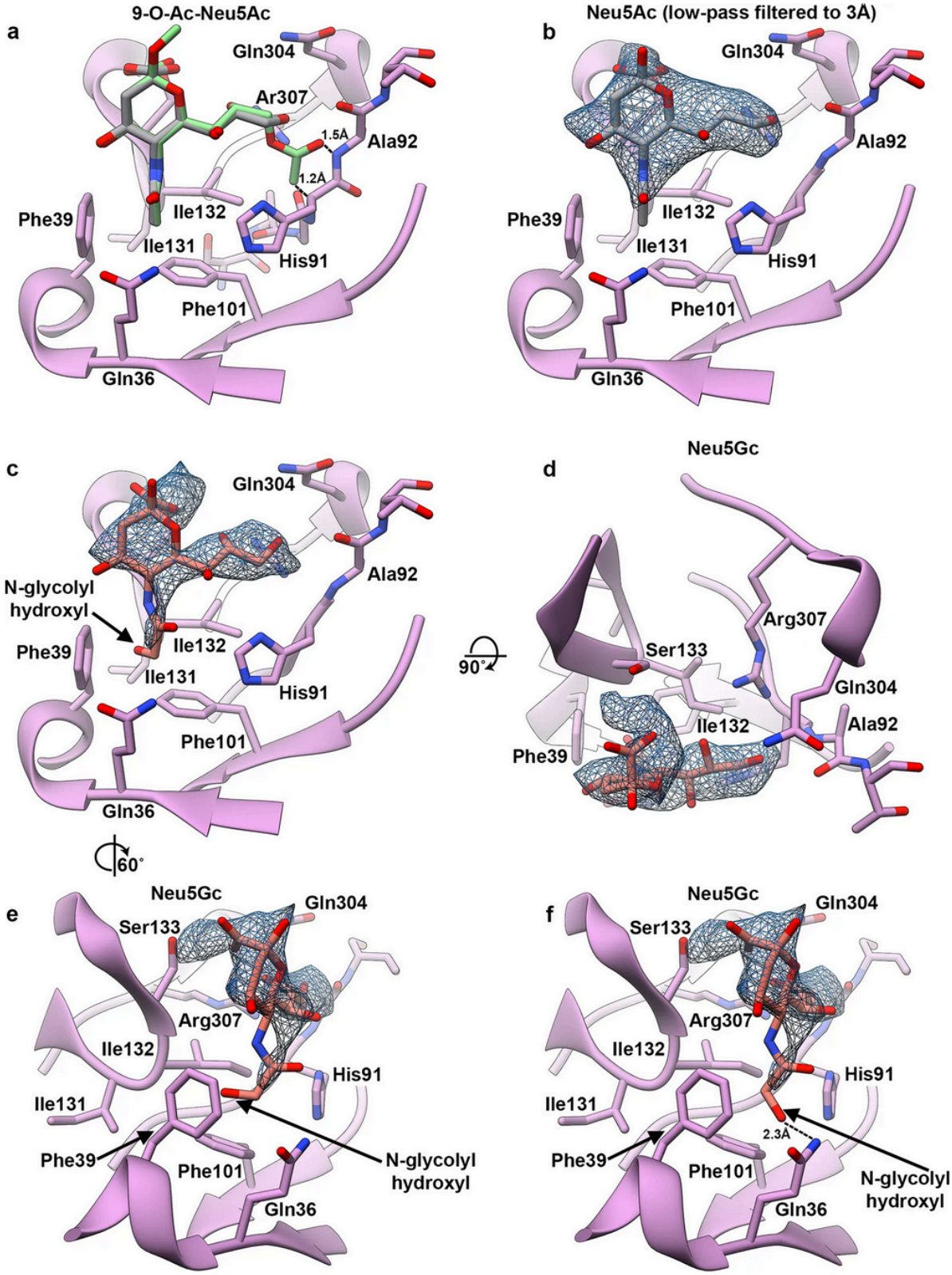
Bond Angles (°)	2.0		2.0	2.0	2.0
Validation					
MolProbity Score	0.98		1.04	1.01	0.99
Clashscore	2.1		2.54	2.28	2.19
Poor rotamers (%)	0.2			0.2	0.2
Favored (%)	98.35		98.26	98.35	98.35
Allowed (%)	99.91		99.91	99.91	99.91
Disallowed (%)	0.09		0.09	0.09	0.09
EMRinger Score	4.43		4.27	4.5	4.66

Figure V.IV: MERS-CoV S and HCoV-OC43 S glycoproteins interact with sialosides, using distinct binding grooves.



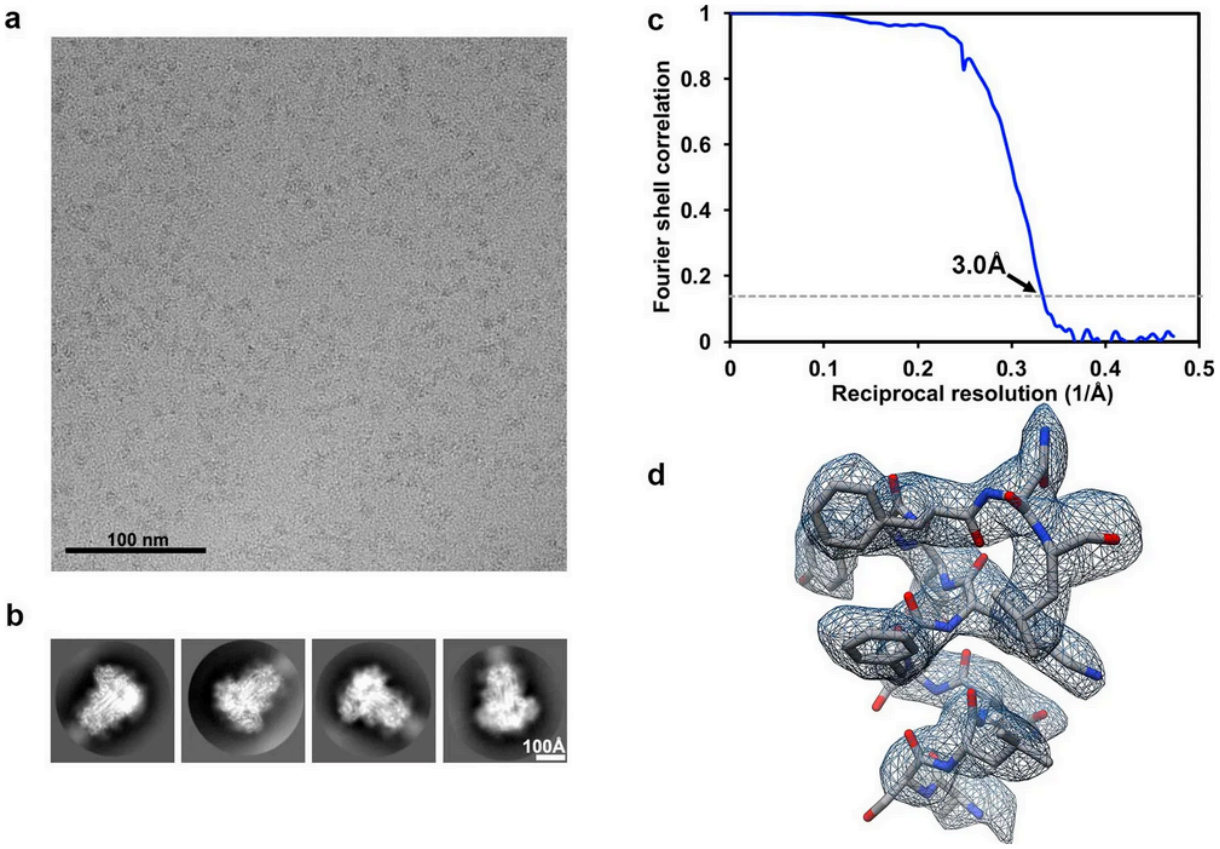
a,**b**, Molecular surface representation of the MERS-CoV S (**a**) and HCoV-OC43 S (**b**) A domains bound to Neu5Ac and 5-*N*-acetyl,9-*O*-acetyl neuraminic acid α-methyl glycoside (9-*O*-Ac-Neu5Ac, PDB 6NZK[19]), respectively, and oriented identically. The sialosides are rendered as sticks with carbon, nitrogen and oxygen atoms colored gray, blue and red, respectively. N-linked glycans observed in the cryo-EM reconstructions are rendered as dark blue spheres. **c**, Ribbon diagrams of superimposed MERS-CoV S (pink) and HCoV-OC43 S (gold) A domains with bound sialosides. The sialosides are rendered as sticks with carbon, nitrogen and oxygen atoms colored pink (MERS-CoV S) or gold (HCoV-OC43 S), blue and red, respectively.

Figure V.V: Structural basis for the MERS-CoV S selectivity for neuraminic acids that are not 9-O-acetylated or 5-N-glycolylated.



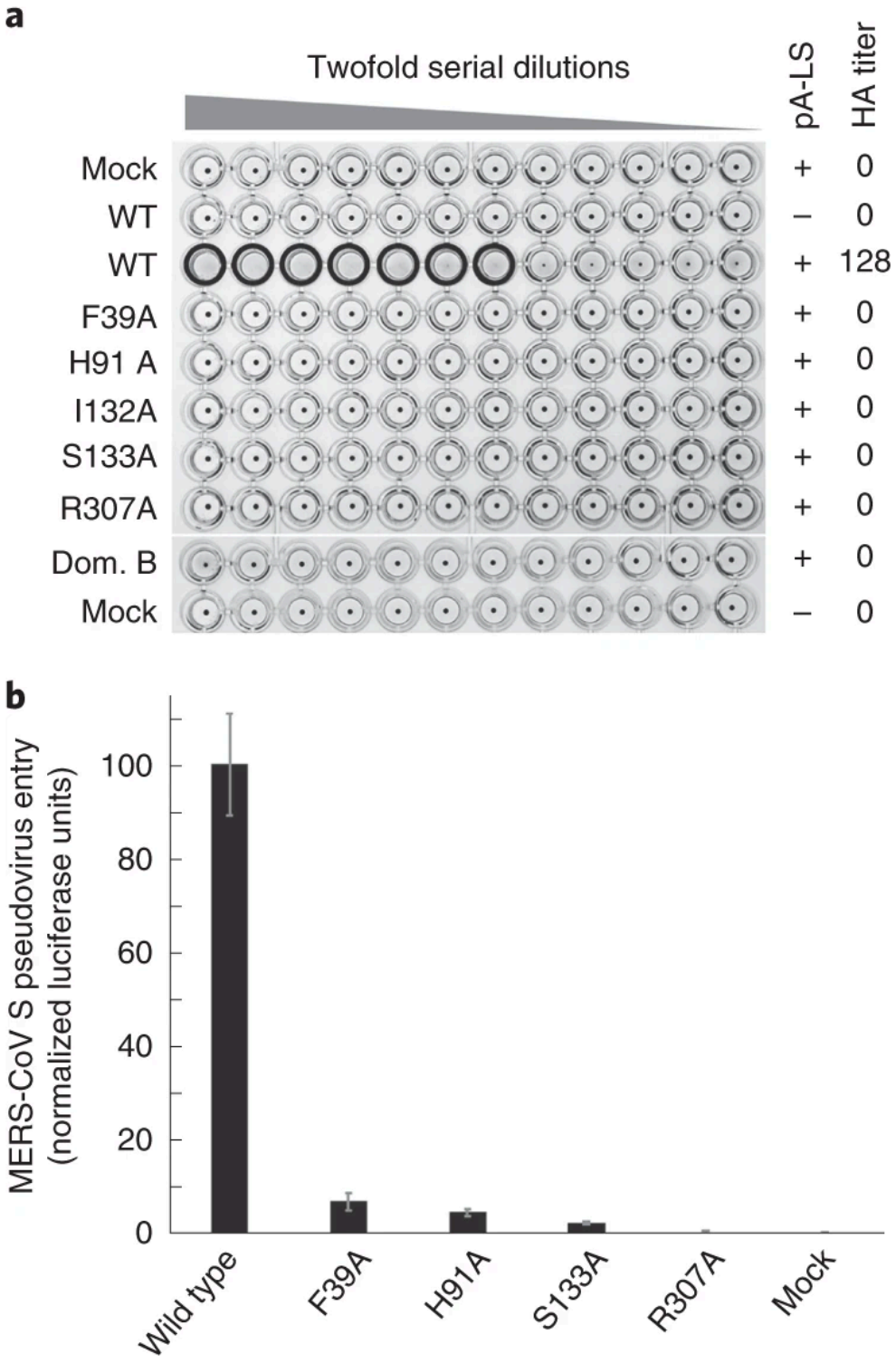
a, Superimposition of the 5-N-acetyl,9-*O*-acetyl neuraminic acid α-methyl glycoside (9-O-Ac-Neu5Ac, green) from the HCoV-OC43 S holo structure (PDB 6NZK[19]) to the bound Neu5Ac from the MERS-CoV S/Neu5Ac complex structure suggests the MERS-CoV S sialoside-binding site could not sterically accommodate the 9-O-acetyl group in this conformation. The distances between the 9-O-acetyl group and His91 or Ala92 are indicated. **b**, The MERS-CoV S sialoside-binding site rendered as a ribbon diagram with the side chains of key surrounding residues shown. Neu5Ac is rendered with the corresponding region of cryoEM density from the MERS-CoV S/Neu5Ac complex structure (low-pass filtered to 3Å and scaled to the MERS-CoV S/Neu5Gc map) shown as a blue mesh contoured at 6σ. **c**–**f**, The cryoEM structure of MERS-CoV S in complex with Neu5Gc shows weaker density for the sialoside (blue mesh contoured at 6σ), indicating poor steric and/or chemical accommodation of the 5-N-glycolyl hydroxyl group in the hydrophobic pocket defined by Phe39, Phe101, Ile131 and Ile132. In panels (**c-e**) Neu5Gc was modeled based on the porcine rotavirus CRW-8 VP8* holo structure (PDB 3TAY[41]) which corresponds to the conformation observed in 1 out of 8 structures available in the protein data bank with a Neu5Gc analogue. In panel **f**, Neu5Gc was modeled based on the rhesus rotavirus VP8* holo structure (PDB 3TB0[41]) which is the only structure in the protein data bank featuring Neu5Gc with the 5-N-glycolyl group in an alternate orientation. The distance between the 5-N-glycolyl hydroxyl group and Gln36 is indicated.

Figure V.VI: CryoEM analysis of the MERS-CoV S glycoprotein in complex with Neu5Gc at 3.0 Å resolution



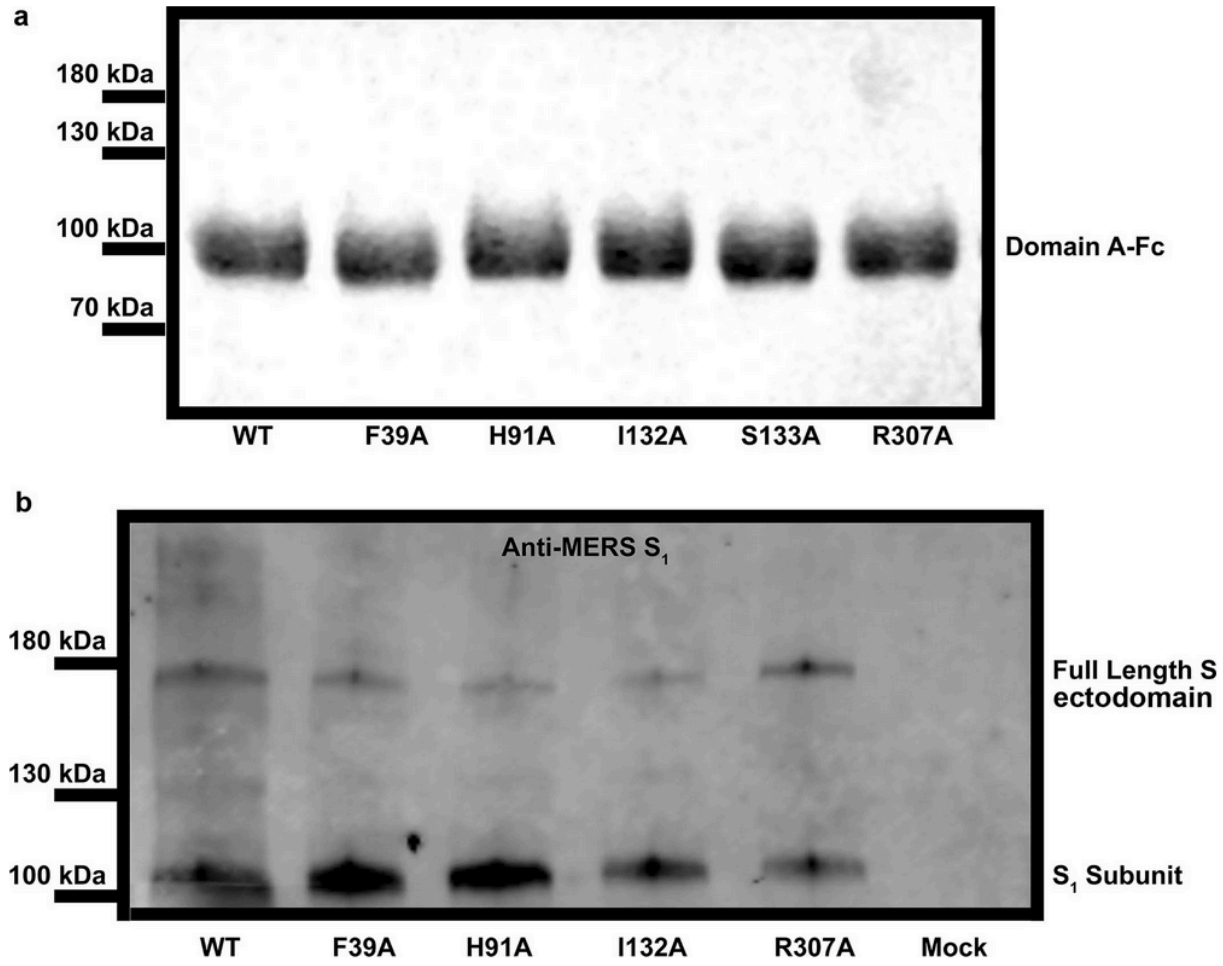
a,**b**, Representative electron micrograph (**a**) and class averages (**b**) for the Neu5Gc-bound MERS-CoV S structure. **c**, Gold-standard Fourier shell correlation curve. The 0.143 cut-off is indicated by horizontal dashed lines. **d**, Representative density shown as blue mesh with the corresponding atomic model rendered as sticks colored grey, blue and red for carbon, nitrogen and oxygen atoms, respectively.

Figure V.VII: The ligand-binding site is required for MERS-CoV S mediated attachment to sialosides and entry into human airway epithelial cells



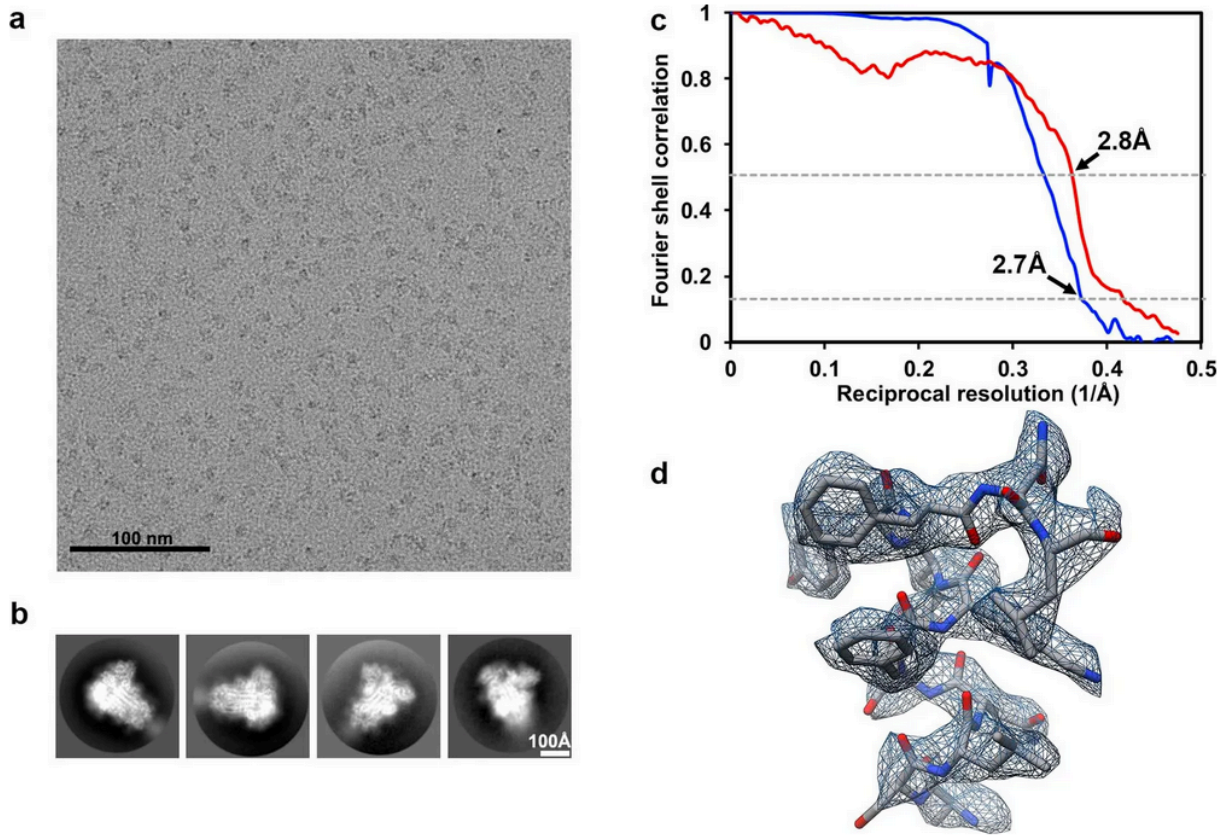
a, Hemagglutination assay using MERS-CoV S wild type (WT) or mutants. Wells positive for hemagglutination (HA) are encircled. F39A, H91A, I132A, S133A or R307A substitutions completely abrogated domain A-Fc-mediated hemagglutination of human erythrocytes. Mock-treated erythrocytes, absence of the protein A-fused lumazine synthase (pA-LS) or replacement of domain A with domain B (Dom. B) were used as negative controls. The assays were performed four times and a representative experiment is shown. **b**, MERS-CoV S F39A, H91A, S133A or R307A substitutions inhibited entry of pseudotyped murine leukemia viral particles into human airway Calu-3 cells. Data are normalized relative to wild type and shown as mean and s.d. of $n = 3$ pseudovirus experiments (technical replicates).

Figure V.VIII: SDS-PAGE and Western blot analyses



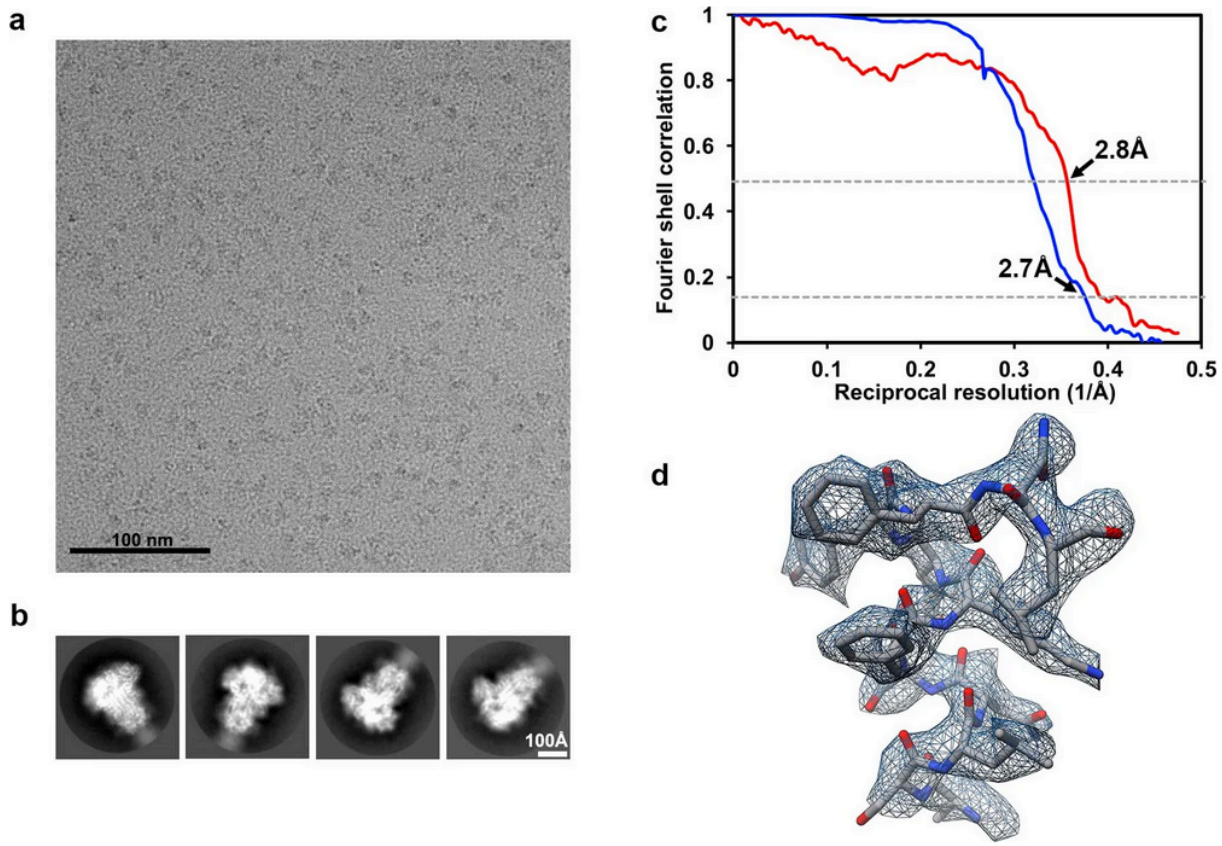
a, SDS-PAGE analysis of purified wild type or mutants MERS-CoV S domain A fused to human immunoglobulin Fc. Two micrograms of each protein was loaded. b, Western-blot analysis of murine leukemia viral particles pseudotyped with wild type or mutants MERS-CoV S using an anti-MERS-CoV S_1 polyclonal antibody. Uncropped blot image is available as source data.

Figure V.IX: CryoEM analysis of the MERS-CoV S glycoprotein in complex with sLeX at 2.7 Å resolution



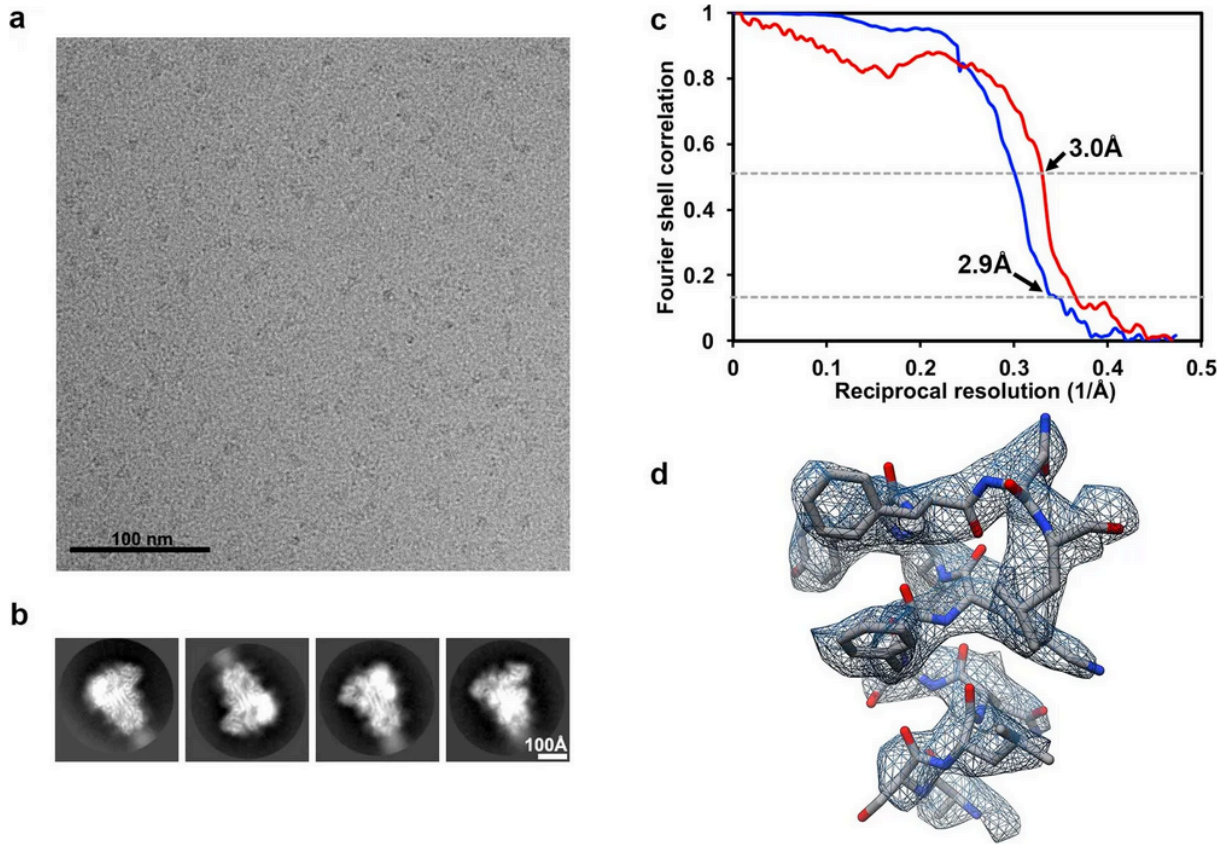
a,b, Representative electron micrograph (a) and class averages (b) for the sLeX-bound MERS-CoV S structure. c, Gold-standard (blue) and map/model (red) Fourier shell correlation curves. The 0.143 and 0.5 cutoffs are indicated by horizontal dashed lines. d, Representative density shown as blue mesh with the corresponding atomic model rendered as sticks colored grey, blue and red for carbon, nitrogen and oxygen atoms, respectively.

Figure V.X: CryoEM analysis of the MERS-CoV S glycoprotein in complex with 2,3-SLN at 2.7 Å resolution



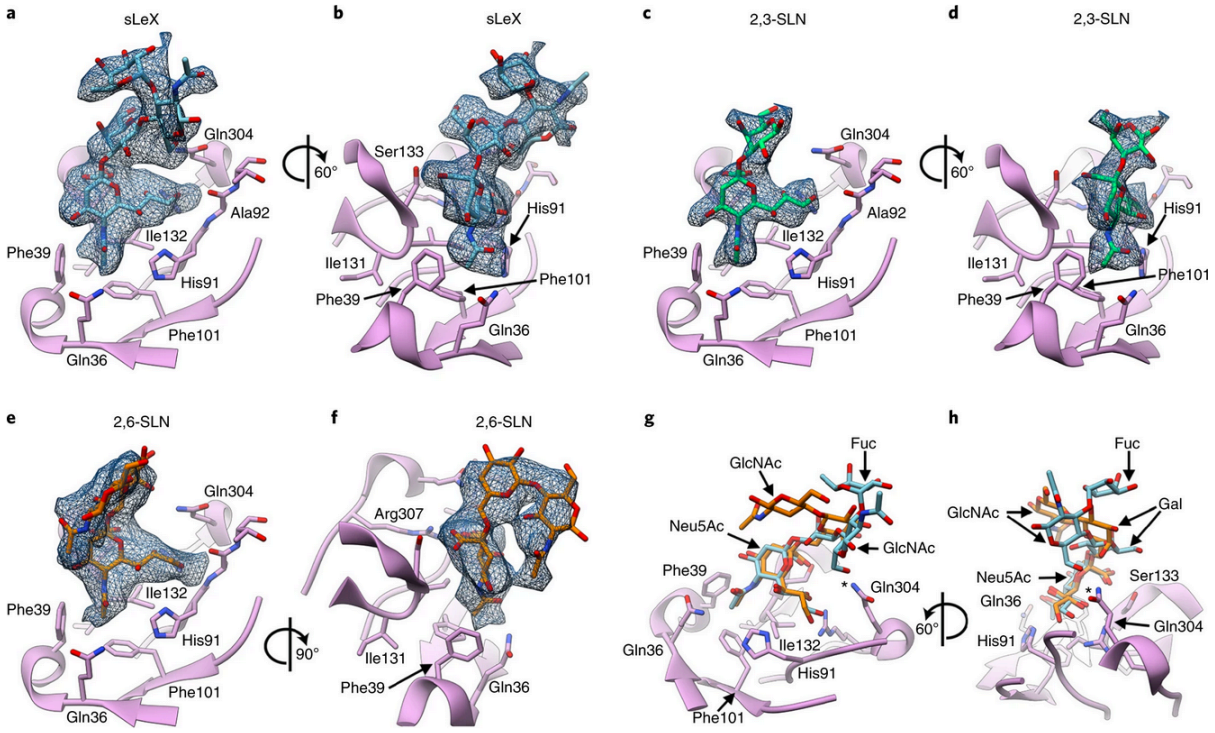
a,b, Representative electron micrograph (a) and class averages (b) for the 2,3-SLN-bound MERS-CoV S structure. c, Gold-standard (blue) and map/model (red) Fourier shell correlation curves. The 0.143 and 0.5 cutoffs are indicated by horizontal dashed lines. d, Representative density shown as blue mesh with the corresponding atomic model rendered as sticks colored grey, blue and red for carbon, nitrogen and oxygen atoms, respectively.

Figure V.XI: CryoEM analysis of the MERS-CoV S glycoprotein in complex with 2,6-SLN at 2.9 Å resolution



a,b, Representative electron micrograph (a) and class averages (b) for the 2,6-SLN-bound MERS-CoV S structure. c, Gold-standard (blue) and map/model (red) Fourier shell correlation curves. The 0.143 and 0.5 cutoffs are indicated by horizontal dashed lines. d, Representative density shown as blue mesh with the corresponding atomic model rendered as sticks colored grey, blue and red for carbon, nitrogen and oxygen atoms, respectively.

Figure V.XII: Structural basis of MERS-CoV S selectivity for α2,3-linked sialoside receptors.



a–f, Enlarged views of the MERS-CoV S structures in complex with sLeX (a,b), 2,3-SLN (c,d) and 2,6-SLN (e,f). The cryo-EM densities corresponding to the ligands are shown as a blue mesh contoured at 3.7σ (a), 3.2σ (b), 5.8σ (c,d) and 4.5σ (e,f). g,h, The ligand-binding site with the sLeX and 2,6-SLN-bound MERS-CoV S structures superimposed, showing the distinct orientation of the two ligands relative to domain A. The asterisks indicate the proximity of the sLeX GlcNAc to the binding site, whereas the 2,6-SLN GlcNAc points away towards the solvent. In a–d, the A domain is rendered as a ribbon diagram with selected side chains of residues surrounding the sialoside shown as sticks. The ligands are rendered as sticks with atoms colored by elements (carbon, cyan (sLeX), green (2,3-SLN) or orange (2,6-SLN); nitrogen, blue; oxygen, red). Neu5Ac, 5-*N*-acetyl neuraminic acid; GlcNAc, *N*-acetyl glucosamine; Fuc, fucose; Gal, galactose.

Concluding remarks:

Through my PhD I have solved the structure of the MHV S (Walls et al. 2016, Walls et al. 2017a) protein, blasting open the CoV S protein structural field. With this structure we learned the architecture of CoV S proteins in general, knowing the connectivity and organization. There were still many questions left open ended by this structure that I next tried to address. One of those questions was, if we solve an evolutionarily distant spike, i.e. not a beta-CoV S protein, will the organization be at all different? The answer, for alpha (Walls et al. 2016a) and delta CoVs (Xiong, et al 2018) at least, was that the S_1 organization is different and these structural differences account for alpha-CoV spikes inability to bind their native receptor. Another question was, how does the pre-fusion differ from the post-fusion and are these transitions going to look similar to other class I fusion proteins? I solved the structure of the S_2 subunit of the MHV S protein in the post-fusion conformation (Walls et al. 2017) and identified similarities to the RSV F pre to postfusion transition. We also showed here key regions that could be important in the transition process and identifying how fusion occurs. What started as an easy, low ball project to identify antibody-antigen interactions turned into a multi-year multi-scientist innovation that led us not only to two novel structures of MERS-CoV and SARS-CoV S in complex with potent, neutralizing antibodies from humans (Walls et al. 2019), but also let us peak into the fusion transition we had previously been skirting around with static structures. While these structures were dynamic and dealt with the opening and closing of the receptor binding domains, the final story is of MERS-CoV S in complex with sialoside receptors which had a stabilizing effect on these domains.

When I began this project, CryoEM was still in its infancy and there were many concerns and limitations in the field with no clear directions or solutions. Five years later, and the changes of staying in two fields in their infancy are enormous. The year we solved the structure of the MHV S protein, the first coronavirus spike protein solved, there were only 1,324 total structures solved by CryoEM deposited in the PDB. Just three years later when this dissertation was written there already have been more than 1,324 deposited structures to date in 2019, more than total submissions prior to 2016 in just one year. This is not the only measure of increased output of the CryoEM field. I once had to travel to other microscopes around the country to use the top of the line Krios, but luckily the Beckman foundation grant allowed the UW department of biochemistry to purchase our own Krios which has significantly increased our throughput and lowered my stress levels. Polishing, a necessary process for obtaining high resolution structures, took two weeks and 18Tb of storage space with the MHV S structure. Now, we routinely polish datasets in just a matter of hours. The MHV S structure at the time of publication was considered a difficult problem due to the small size of the protein and *de novo* nature of the model building problem. At the time of dissertation writing, I have routinely heard that my 450kDa protein is a large and easy CryoEM target. It is not just the CryoEm aspect of my project that has improved and exploded, but since our MHV S publication, the number of CoV S protein structures solved has also exploded. We the CoV S field now have alpha, beta, delta, and gamma CoV S structures, pre and post-fusion conformation structures, and structures in complex with protein receptors, sugar receptors, and neutralizing antibodys. It's been fun to know that in as short of time as one single PhD, I have seen my project go from 'difficult', 'never been done before', or 'insurmountable' to 'low hanging fruit',

‘simple’, or ‘routine. I cannot help but smile knowing that I have contributed to this change through the many hours I have spent working in lab late, tinkering on the microscope in the early mornings, or spending my weekend on the computer processing data. Looking back and knowing that my body of work is someone else’s introductory slides is a satisfying way to end my PhD. This work is the foundation for all future structural and hypothesis driven CoV S projects and that is pretty amazing when you step back and think on it.

Thank you for reading!

References:

Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G. J., Wilson, K. S., & Cowtan, K. D. (2015). Privateer: Software for the conformational validation of carbohydrate structures. *Nature Structural & Molecular Biology, 22*(11), 833-834. doi:10.1038/nsmb.3115

Bai, X., Fernandez, I. S., McMullan, G., & Scheres, S. H. W. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife, 2*, e00461. doi:10.7554/eLife.00461

Bai, X., Rajendra, E., Yang, G., Shi, Y., & Scheres, S. H. W. (2015). Sampling the conformational space of the catalytic subunit of human γ-secretase. *eLife, 4* doi:10.7554/eLife.11182

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America,98*(18), 10037-10041. doi:10.1073/pnas.181342398

Bakkers, M. J. G., Lang, Y., Feitsma, L. J., Hulswit, R. J. G., de Poot, Stefanie A. H., van Vliet, Arno L. W., . . . de Groot, R. J. (2017). Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. *Cell Host & Microbe, 21*(3), 356-366. doi:10.1016/j.chom.2017.02.008

Bakkers, M. J. G., Zeng, Q., Feitsma, L. J., Hulswit, R. J. G., Li, Z., Westerbeke, A., . . . de Groot, R. J. (2016). Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proceedings of the National Academy of Sciences, 113*(22), E3111-E3119.

Barad, B. A., Echols, N., Wang, R. Y., Cheng, Y., DiMaio, F., Adams, P. D., & Fraser, J. S. (2015). EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods, 12*(10), 943-946. doi:10.1038/nmeth.3541

Barlan, A., Zhao, J., Sarkar, M. K., Li, K., McCray, P. B., Perlman, S., & Gallagher, T. (2014). Receptor variation and susceptibility to middle east respiratory syndrome coronavirus infection. *Journal of Virology, 88*(9), 4953-4961. doi:10.1128/JVI.00161-14

Belouzard, S., Chu, V. C., & Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences of the United States of America, 106*(14), 5871-5876. doi:10.1073/pnas.0809524106

Beniac, D. R., Andonov, A., Grudeski, E., & Booth, T. F. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology, 13*(8), 751-752. doi:10.1038/nsmb1123

Bern, M., Kil, Y. J., & Becker, C. (2012). Byonic: Advanced peptide and protein identification software. *Current Protocols in Bioinformatics, Chapter 13*, Unit13.20. doi:10.1002/0471250953.bi1320s40

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., . . . Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research, 35*(6), e45. doi:10.1093/nar/gkm047

Booth, T. F., Grudeski, E., Beniac, D. R., & Andonov, A. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology, 13*(8), 751-752. doi:10.1038/nsmb1123

Bosch, B.J. Ruurd van der Zee, de Haan, C.A.M., & Rottier, P. J.M. (2003). The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *Journal of Virology, 77*(16), 8801-8811. doi:10.1128/JVI.77.16.8801-8811.2003

Bosch, B. J., Bartelink, W., & Rottier, P. J. M. (2008). Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *Journal of Virology, 82*(17), 8887-8890. doi:10.1128/JVI.00415-08

Brown, A., Long, F., Nicholls, R. A., Toots, J., Emsley, P., & Murshudov, G. (2015). Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallographica Section D: Biological Crystallography, 71*(Pt 1), 136-153. doi:10.1107/S1399004714021683

Bullough, P. A., Hughson, F. M., Skehel, J. J., & Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature, 371*(6492), 37-43. doi:10.1038/371037a0

Burkard, C., Verheije, M. H., Wicht, O., van Kasteren, S. I., van Kuppeveld, F. J., Haagmans, B. L., . . . de Haan, Cornelis A. M. (2014). Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathogens, 10*(11), e1004502. doi:10.1371/journal.ppat.1004502

Buzon, V., Natrajan, G., Schibli, D., Campelo, F., Kozlov, M. M., & Weissenhorn, W. (2010). Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. *PLoS Pathogens, 6*(5), e1000880. doi:10.1371/journal.ppat.1000880

Campbell, M. G., Cheng, A., Brilot, A. F., Moeller, A., Lyumkis, D., Veesler, D., . . . Grigorieff, N. (2012). Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure, 20*(11), 1823-1828. doi:10.1016/j.str.2012.08.026

Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S., & Carragher, B. (2015). 2.8 Å resolution reconstruction of the thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *eLife, 4* doi:10.7554/eLife.06380

Chan, C., Chu, H., Wang, Y., Wong, B. H., Zhao, X., Zhou, J., . . . Yuen, K. (2016). Carcinoembryonic antigen-related cell adhesion molecule 5 is an important surface attachment factor that facilitates entry of middle east respiratory syndrome coronavirus. *Journal of Virology, 90*(20), 9114-9127. doi:10.1128/JVI.01133-16

Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D. C., & Harrison, S. C. (2005). Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature, 433*(7028), 834-841. doi:10.1038/nature03327

Chen, L., Gorman, J. J., McKimm-Breschkin, J., Lawrence, L. J., Tulloch, P. A., Smith, B. J., . . . Lawrence, M. C. (2001). The structure of the fusion glycoprotein of newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion. *Structure (London, England: 1993), 9*(3), 255-266. doi:10.1016/s0969-2126(01)00581-0

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W., & Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy, 135*, 24-35. doi:10.1016/j.ultramic.2013.06.004

Chen VB, Arendall WB, Headd JJ et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 1):12-21. doi: 10.1107/S0907444909042073.

Cheng, Y., Grigorieff, N., Penczek, P. A., & Walz, T. (2015). A primer to single-particle cryo-electron microscopy. *Cell*. 161(3), 438-449.

Cheng, Y. (2015). Single-particle cryo-EM at crystallographic resolution. *Cell, 161*(3), 450-457. doi:10.1016/j.cell.2015.03.049

Chiu, S. S., Chan, K. H., Chu, K. W., Kwan, S. W., Guan, Y., Poon, L. L. M., & Peiris, J. S. M. (2005). Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in hong kong, china. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America, 40*(12), 1721-1729. doi:10.1086/430301

Chu, H., Chan, C., Zhang, X., Wang, Y., Yuan, S., Zhou, J., . . . Yuen, K. (2018). Middle east respiratory syndrome coronavirus and bat coronavirus HKU9 both can utilize GRP78 for attachment onto host cells. *The Journal of Biological Chemistry, 293*(30), 11709-11726. doi:10.1074/jbc.RA118.001897

Chua, K. B., Crameri, G., Hyatt, A., Yu, M., Tompang, M. R., Rosli, J., . . . Wang, L. (2007). A previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proceedings of the National Academy of Sciences, USA, 104*(27), 11424-11429.

Coleman, C. M., & Frieman, M. B. (2014). Coronaviruses: Important emerging human pathogens. *Journal of Virology, 88*(10), 5209-5212. doi:10.1128/JVI.03488-13

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research, 16*(22), 10881-10890. doi:10.1093/nar/16.22.10881

Corti, D., Zhao, J., Pedotti, M., Simonelli, L., Agnihothram, S., Fett, C., Fernandez-Rodriguez, B., Foglierini, M., Agatic, G., Vanzetta, F. *et al.* (2015). Prophylactic and postexposure efficacy of a potent human monoclonal antibody against MERS coronavirus. Proceedings of the National Academy of Sciences of the United States of America *112*, 10473-10478.

Daniel, C., Anderson, R., Buchmeier, M. J., Fleming, J. O., Spaan, W. J., Wege, H., & Talbot, P. J. (1993). 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. *Journal of Virology, 67*(3), 1185-1194.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., & André, I. (2011). Modeling symmetric macromolecular structures in Rosetta3. *PloS One, 6*(6), e20450. doi:10.1371/journal.pone.0020450

DiMaio, F., Song, Y., Li, X., Brunner, M. J., Xu, C., Conticello, V., . . . Baker, D. (2015). Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods, 12*(4), 361-365. doi:10.1038/nmeth.3286

DiMaio, F., Zhang, J., Chiu, W., & Baker, D. (2013). Cryo-EM model validation using independent map reconstructions. *Protein Science: A Publication of the Protein Society, 22*(6), 865-868. doi:10.1002/pro.2267

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic Acids Research, 32*, 665. doi:10.1093/nar/gkh381

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H., Becker, S., . . . Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England Journal of Medicine, 348*(20), 1967-1976. doi:10.1056/NEJMoa030747

Du, L., Tai, W., Yang, Y., Zhao, G., Zhu, Q., Sun, S., Liu, C., Tao, X., Tseng, C.K., Perlman, S. *et al.* (2016). Introduction of neutralizing immunogenicity index to the rational design of MERS coronavirus subunit vaccines. Nat Commun *7*, 13473.

Duquerroy, S., Vigouroux, A., Rottier, P. J. M., Rey, F. A., & Bosch, B. J. (2005). Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology, 335*(2), 276-285. doi:10.1016/j.virol.2005.02.022

Earnest, J. T., Hantak, M. P., Li, K., McCray, P. B., Perlman, S., & Gallagher, T. (2017). The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. *PLoS Pathogens, 13*(7), e1006546. doi:10.1371/journal.ppat.1006546

Eckert, D. M., Malashkevich, V. N., & Kim, P. S. (1998). Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *Journal of Molecular Biology, 284*(4), 859-865. doi:10.1006/jmbi.1998.2214

Egelman, E. H. (2014). Ambiguities in helical reconstruction. *eLife, 3* doi:10.7554/eLife.04969

Elshabrawy, H. A., Coughlin, M. M., Baker, S. C., & Prabhakar, B. S. (2012). Human monoclonal antibodies against highly conserved HR1 and HR2 domains of the SARS-CoV spike protein are more broadly neutralizing. *PloS One, 7*(11), e50366. doi:10.1371/journal.pone.0050366

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of coot. *Acta Crystallographica. Section D, Biological Crystallography, 66*(Pt 4), 486-501. doi:10.1107/S0907444910007493

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics, 13*(1), 22-24. doi:10.1002/pmic.201200439

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204-1214.

Falkowska, E., Kajumo, F., Garcia, E., Reinus, J., & Dragic, T. (2007). Hepatitis C virus envelope glycoprotein E2 glycans modulate entry, CD81 binding, and neutralization. *Journal of Virology, 81*(15), 8072-8079. doi:10.1128/JVI.00459-07

Fouchier, R. A. M., Hartwig, N. G., Bestebroer, T. M., Niemeyer, B., de Jong, J. C., Simon, J. H., & Osterhaus, Albert D. M. E. (2004). A previously undescribed coronavirus associated with respiratory disease in humans. *Proceedings of the National Academy of Sciences of the United States of America, 101*(16), 6212-6216. doi:10.1073/pnas.0400762101

Frenz, B., Rämisch, S., Borst, A. J., Walls, A. C., Adolf-Bryfogle, J., Schief, W. R., . . . DiMaio, F. (2019). Automatically fixing errors in glycoprotein structures with rosetta. *Structure, 27*(1), 134-139.e3. doi:10.1016/j.str.2018.09.006

Frenz, B., Walls, A. C., Egelman, E. H., Veesler, D., & DiMaio, F. (2017). RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nature Methods, 14*(8), 797-800. doi:10.1038/nmeth.4340

Frese, C. K., Zhou, H., Taus, T., Altelaar, A. F. M., Mechtler, K., Heck, A. J. R., & Mohammed, S. (2013). Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *Journal of Proteome Research, 12*(3), 1520-1525. doi:10.1021/pr301130k

Gao, J., Lu, G., Qi, J., Li, Y., Wu, Y., Deng, Y., . . . Gao, G. F. (2013). 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. *Journal of Virology, 87*(24), 13134-13140. doi:10.1128/JVI.02433-13

Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C. *et al.* (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature *503*, 535-538.

Glaeser, R. M., Typke, D., Tiemeijer, P. C., Pulokas, J., & Cheng, A. (2011). Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. *Journal of Structural Biology, 174*(1), 1-10. doi:10.1016/j.jsb.2010.12.005

Goddard, T. D., Huang, C. C., & Ferrin, T. E. (2007). Visualizing density maps with UCSF chimera. *Journal of Structural Biology, 157*(1), 281-287. doi:10.1016/j.jsb.2006.06.010

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science: A Publication of the Protein Society, 27*(1), 14-25. doi:10.1002/pro.3235

Graham, R. L., Donaldson, E. F., & Baric, R. S. (2013). A decade after SARS: Strategies for controlling emerging coronaviruses. *Nature Reviews. Microbiology, 11*(12), 836-848. doi:10.1038/nrmicro3143

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., & Wang, X. (2017). Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Research, 27*(1), 119-129. doi:10.1038/cr.2016.152

Haagmans, B.L., Al Dhahiry, S.H., Reusken, C.B., Raj, V.S., Galiano, M., Myers, R., Godeke, G.J., Jonges, M., Farag, E., Diab, A. *et al.* (2014). Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect Dis *14*, 140-145.

Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology, 15*(7), 690-698. doi:10.1038/nsmb.1456

Hoek, L. v. d., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J. M., Wolthers, K. C., . . . Berkhout, B. (2004). Identification of a new human coronavirus. *Nature Medicine, 10*(4), 368-373. doi:10.1038/nm1024

Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B., & Pöhlmann, S. (2005). Human coronavirus NL63 employs the severe acute respiratory syndrome

coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences of the United States of America, 102*(22), 7988-7993. doi:10.1073/pnas.0409465102
Holm, L., & Rosenström, P. (2010). Dali server: Conservation mapping in 3D. *Nucleic Acids Research, 38*(Web Server issue), 545. doi:10.1093/nar/gkq366
Hu, B., Zeng, L.P., Yang, X.L., Ge, X.Y., Zhang, W., Li, B., Xie, J.Z., Shen, X.R., Zhang, Y.Z., Wang, N. *et al.* (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog *13*, e1006698.
Huang, C., Tang, M., Zhang, M., Majeed, S., Montabana, E., Stanfield, R. L., . . . Kwong, P. D. (2005). Structure of a V3-containing HIV-1 gp120 core. *Science (New York, N.Y.), 310*(5750), 1025-1028. doi:10.1126/science.1118398
Hulswit, R. J. G., Lang, Y., Bakkers, M. J. G., Li, W., Li, Z., Schouten, A., . . . de Groot, R. J. (2019). Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proceedings of the National Academy of Sciences of the United States of America, 116*(7), 2681-2690. doi:10.1073/pnas.1809667116
Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., . . . Liddington, R. C. (2006). Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *The Journal of Biological Chemistry, 281*(45), 34610-34616. doi:10.1074/jbc.M603275200
Igonet, S., & Rey, F. A. (2012). SnapShot: Viral and eukaryotic protein fusogens. *Cell, 151*(7), 1634-1634.e1. doi:10.1016/j.cell.2012.11.041
Ingallinella, P., Bianchi, E., Finotto, M., Cantoni, G., Eckert, D. M., Supekar, V. M., . . . Pessi, A. (2004). Structural characterization of the fusion-active complex of severe acute respiratory syndrome (SARS) coronavirus. *Proceedings of the National Academy of Sciences of the United States of America, 101*(23), 8709-8714. doi:10.1073/pnas.0402753101
Jeffers, S. A., Hemmila, E. M., & Holmes, K. V. (2006). Human coronavirus 229E can use CD209L (L-SIGN) to enter cells. *Advances in Experimental Medicine and Biology, 581*, 265-269. doi:10.1007/978-0-387-33012-9_44
Jeffers, S. A., Tusell, S. M., Gillim-Ross, L., Hemmila, E. M., Achenbach, J. E., Babcock, G. J., . . . Holmes, K. V. (2004). CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proceedings of the National Academy of Sciences of the United States of America, 101*(44), 15748-15753. doi:10.1073/pnas.0403812101
Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.
Kaledhonkar, S., Fu, Z., Caban, K., Li, W., Chen, B., Sun, M., . . . Frank, J. (2019). Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. *Nature, 570*(7761), 400-404. doi:10.1038/s41586-019-1249-5
Kausche, G.A., E. Pfankuch H. Ruska.Die sichtbarmachung von pflanzlichem virus im übermikroskop. 1939.
Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F., & Matsuyama, S. (2012). Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *Journal of Virology, 86*(12), 6537-6545. doi:10.1128/JVI.00094-12

Kimanius, D., Forsberg, B.O., Scheres, S.H., and Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. eLife *5*.
Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., . . . Ward, A. B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature, 531*(7592), 118-121. doi:10.1038/nature17200
Kirchdoerfer, R.N., Wang, N., Pallesen, J., Wrapp, D., Turner, H.L., Cottrell, C.A., Corbett, K.S., Graham, B.S., McLellan, J.S., and Ward, A.B. (2018). Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Sci Rep *8*, 15701.
Krempl, C., Schultze, B., Laude, H., & Herrler, G. (1997). Point mutations in the S protein connect the sialic acid binding activity with the enteropathogenicity of transmissible gastroenteritis coronavirus. *Journal of Virology, 71*(4), 3285-3287.
Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology, 372*(3), 774-797. doi:10.1016/j.jmb.2007.05.022
Krokhin, O., Li, Y., Andonov, A., Feldmann, H., Flick, R., Jones, S., Stroeher, U., Bastien, N., Dasuri, K.V., Cheng, K. *et al.* (2003). Mass spectrometric characterization of proteins from the SARS virus: a preliminary report. Mol Cell Proteomics *2*, 346-356.
Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R., & Gallagher, T. M. (2001). Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *Journal of Virology, 75*(6), 2792-2802. doi:10.1128/JVI.75.6.2792-2802.2001
Kucukelbir, A., Sigworth, F. J., & Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nature Methods, 11*(1), 63-65. doi:10.1038/nmeth.2727
Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., . . . Carragher, B. (2009). Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology, 166*(1), 95-102.
Lee, B., & Richards, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology, 55*(3), 379-400. doi:10.1016/0022-2836(71)90324-x
Li, F. (2012). Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *Journal of Virology, 86*(5), 2856-2858. doi:10.1128/JVI.06882-11
Li, F., Berardi, M., Li, W., Farzan, M., Dormitzer, P. R., & Harrison, S. C. (2006). Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *Journal of Virology, 80*(14), 6794-6800. doi:10.1128/JVI.02744-05
Li, F., Li, W., Farzan, M., & Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science, 309*(5742), 1864-1868. doi:10.1126/science.1116480
Li, W., Hulswit, R. J. G., Widjaja, I., Raj, V. S., McBride, R., Peng, W., . . . Bosch, B. (2017). Identification of sialic acid-binding function for the middle east respiratory syndrome coronavirus spike glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America, 114*(40), E8508-E8517. doi:10.1073/pnas.1712592114

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C. *et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450-454.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H. *et al.* (2005b). Bats are natural reservoirs of SARS-like coronaviruses. Science *310*, 676-679.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N. *et al.* (2005c). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J *24*, 1634-1643.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., . . . Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods, 10*(6), 584-590. doi:10.1038/nmeth.2472

Licitra, B. N., Millet, J. K., Regan, A. D., Hamilton, B. S., Rinaldi, V. D., Duhamel, G. E., & Whittaker, G. R. (2013). Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. *Emerging Infectious Diseases, 19*(7), 1066-1073. doi:10.3201/eid1907.121094

Liu, J., Stevens, D. J., Haire, L. F., Walker, P. A., Coombs, P. J., Russell, R. J., . . . Skehel, J. J. (2009). Structures of receptor complexes formed by hemagglutinins from the asian influenza pandemic of 1957. *Proceedings of the National Academy of Sciences of the United States of America, 106*(40), 17175-17180. doi:10.1073/pnas.0906849106

Liu, Q., Stone, J. A., Bradel-Tretheway, B., Dabundo, J., Benavides Montano, J. A., Santos-Montanez, J., . . . Aguilar, H. C. (2013). Unraveling a three-step spatiotemporal mechanism of triggering of receptor-induced nipah virus fusion and cell entry. *PLoS Pathogens, 9*(11), e1003770. doi:10.1371/journal.ppat.1003770

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., . . . Gao, G. F. (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature, 500*(7461), 227-231. doi:10.1038/nature12328

Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. *The Journal of Biological Chemistry, 294*(13), 5181-5197. doi:10.1074/jbc.REV118.005602

Lyumkis, D., Brilot, A. F., Theobald, D. L., & Grigorieff, N. (2013). Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of Structural Biology, 183*(3), 377-388. doi:10.1016/j.jsb.2013.07.005

Lyumkis, D., Julien, J., de Val, N., Cupo, A., Potter, C. S., Klasse, P., . . . Ward, A. B. (2013a). Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science (New York, N.Y.), 342*(6165), 1484-1490. doi:10.1126/science.1245627

Lyumkis, D., Vinterbo, S., Potter, C. S., & Carragher, B. (2013b). Optimod--an automated approach for constructing and optimizing initial models for single-particle electron microscopy. *Journal of Structural Biology, 184*(3), 417-426. doi:10.1016/j.jsb.2013.10.009

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open

source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966-968.
Madu, I. G., Roth, S. L., Belouzard, S., & Whittaker, G. R. (2009). Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *Journal of Virology, 83*(15), 7411-7421. doi:10.1128/JVI.00079-09
Matsuyama, S., and Taguchi, F. (2009). Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. Journal of virology *83*, 11133-11141.
McCoy, L.E., van Gils, M.J., Ozorowski, G., Messmer, T., Briney, B., Voss, J.E., Kulp, D.W., Macauley, M.S., Sok, D., Pauthner, M. *et al.* (2016). Holes in the Glycan Shield of the Native HIV Envelope Are a Target of Trimer-Elicited Neutralizing Antibodies. Cell Rep *16*, 2327-2338.
McLellan, J. S., Chen, M., Joyce, M. G., Sastry, M., Stewart-Jones, G. B. E., Yang, Y., . . . Kwong, P. D. (2013). Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science, 342*(6158), 592-598. doi:10.1126/science.1243283
McLellan, J. S., Chen, M., Leung, S., Graepel, K. W., Du, X., Yang, Y., . . . Graham, B. S. (2013a). Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science, 340*(6136), 1113-1117. doi:10.1126/science.1234914
McLellan, J. S., Yang, Y., Graham, B. S., & Kwong, P. D. (2011). Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. *Journal of Virology, 85*(15), 7788-7796. doi:10.1128/JVI.00555-11
Memish, Z.A., Mishra, N., Olival, K.J., Fagbo, S.F., Kapoor, V., Epstein, J.H., Alhakeem, R., Durosinloun, A., Al Asmari, M., Islam, A. *et al.* (2013). Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. Emerg Infect Dis *19*, 1819-1823.
Menachery, V.D., Yount, B.L., Jr., Debbink, K., Agnihothram, S., Gralinski, L.E., Plante, J.A., Graham, R.L., Scobey, T., Ge, X.Y., Donaldson, E.F. *et al.* (2015). A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med *21*, 1508-1513.
Menachery, V.D., Yount, B.L., Jr., Sims, A.C., Debbink, K., Agnihothram, S.S., Gralinski, L.E., Graham, R.L., Scobey, T., Plante, J.A., Royal, S.R. *et al.* (2016). SARS-like WIV1-CoV poised for human emergence. Proceedings of the National Academy of Sciences of the United States of America *113*, 3048-3053.
Meyer, B., García-Bocanegra, I., Wernery, U., Wernery, R., Sieberg, A., Müller, M. A., ... Eckerle, I. (2015). Serologic assessment of possibility for MERS-CoV infection in equids. *Emerging Infectious Diseases, 21*(1), 181-182. doi:10.3201/eid2101.141342
Michielssens, S., Peters, J.H., Ban, D., Pratihar, S., Seeliger, D., Sharma, M., Giller, K., Sabo, T.M., Becker, S., Lee, D. *et al.* (2014). A designed conformational shift to control protein binding specificity. Angewandte Chemie *53*, 10367-10371.
Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, Pyrc K. Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. *J Virol*. 2014;88(22):13221-13230. . doi: 10.1128/JVI.02078-14.

Millet, J. K., & Whittaker, G. R. (2015). Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Research, 202*, 120-134. doi:10.1016/j.virusres.2014.11.021

Millet, J. K., & Whittaker, G. R. (2014). Host cell entry of middle east respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proceedings of the National Academy of Sciences of the United States of America, 111*(42), 15214-15219. doi:10.1073/pnas.1407087111

Millet, J. K., & Whittaker, G. R. (2016). Murine leukemia virus (MLV)-based coronavirus spike-pseudotyped particle production and infection. *Bio-Protocol, 6*(23) doi:10.21769/BioProtoc.2035

Mindell, J. A., & Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology, 142*(3), 334-347.

Mole, B. (2013). Deadly pig virus slips through US borders. *Nature, 499*(7459), 388. doi:10.1038/499388a

Mou, H., Raj, V. S., van Kuppeveld, Frank J. M., Rottier, P. J. M., Haagmans, B. L., & Bosch, B. J. (2013). The receptor binding domain of the new middle east respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *Journal of Virology, 87*(16), 9379-9383. doi:10.1128/JVI.01277-13

Neuman, B. W., Adair, B. D., Yoshioka, C., Quispe, J. D., Orca, G., Kuhn, P., . . . Buchmeier, M. J. (2006). Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *Journal of Virology, 80*(16), 7918-7928.

Ng, M.L., Tan, S.H., See, E.E., Ooi, E.E., and Ling, A.E. (2003). Proliferative growth of SARS coronavirus in Vero E6 cells. J Gen Virol *84*, 3291-3303.

Nkolola, J.P., Bricault, C.A., Cheung, A., Shields, J., Perry, J., Kovacs, J.M., Giorgi, E., van Winsen, M., Apetri, A., Brinkman-van der Linden, E.C. *et al.* (2014). Characterization and immunogenicity of a novel mosaic M HIV-1 gp140 trimer. Journal of virology *88*, 9538-9552.

Pallesen, J., Wang, N., Corbett, K. S., Wrapp, D., Kirchdoerfer, R. N., Turner, H. L., . . . McLellan, J. S. (2017). Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences of the United States of America, 114*(35), E7348-E7357. doi:10.1073/pnas.1707304114

Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., . . . Kwong, P. D. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 env. *Nature, 514*(7523), 455-461. doi:10.1038/nature13808

Park, J.E., Li, K., Barlan, A., Fehr, A.R., Perlman, S., McCray, P.B., Jr., and Gallagher, T. (2016). Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. Proceedings of the National Academy of Sciences of the United States of America *113*, 12262-12267.

Pascal, K. E., Coleman, C. M., Mujica, A. O., Kamat, V., Badithe, A., Fairhurst, J., . . . Kyratsous, C. A. (2015). Pre- and postexposure efficacy of fully human antibodies against spike protein in a novel humanized mouse model of MERS-CoV infection. *Proceedings of the National Academy of Sciences of the United States of America, 112*(28), 8738-8743. doi:10.1073/pnas.1510830112

Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., & Li, F. (2011). Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proceedings of the National Academy of Sciences of the United States of America, 108*(26), 10696-10701. doi:10.1073/pnas.1104306108

Peng, G., Xu, L., Lin, Y., Chen, L., Pasquarella, J. R., Holmes, K. V., & Li, F. (2012). Crystal structure of bovine coronavirus spike protein lectin domain. *The Journal of Biological Chemistry, 287*(50), 41931-41938. doi:10.1074/jbc.M112.418210

Prabakaran, P., Gan, J., Feng, Y., Zhu, Z., Choudhry, V., Xiao, X., . . . Dimitrov, D. S. (2006). Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *The Journal of Biological Chemistry, 281*(23), 15829-15836. doi:10.1074/jbc.M600697200

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods, 14*(3), 290-296. doi:10.1038/nmeth.4169

Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H. W., Müller, M. A., Dijkman, R., . . . Haagmans, B. L. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature, 495*(7440), 251-254. doi:10.1038/nature12005

Raj, V.S., Okba, N. M. A., Gutierrez-Alvarez, J., Drabek, D., van Dieren, B., Widagdo, W., . . . Haagmans, B. L. (2018). Chimeric camel/human heavy-chain antibodies protect against MERS-CoV infection. *Science Advances, 4*(8), eaas9667. doi:10.1126/sciadv.aas9667

Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L., & Casasnovas, J. M. (2012). Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathogens, 8*(8), e1002859. doi:10.1371/journal.ppat.1002859

Ritchie, G., Harvey, D. J., Feldmann, F., Stroeher, U., Feldmann, H., Royle, L., . . . Rudd, P. M. (2010). Identification of N-linked carbohydrates from severe acute respiratory syndrome (SARS) spike glycoprotein. *Virology, 399*(2), 257-269. doi:10.1016/j.virol.2009.12.020

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research, 42*(Web Server issue), 320. doi:10.1093/nar/gku316

Rockx, B., Corti, D., Donaldson, E., Sheahan, T., Stadler, K., Lanzavecchia, A., & Baric, R. (2008). Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. *Journal of Virology, 82*(7), 3220-3235. doi:10.1128/JVI.02377-07

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology, 192*(2), 216-221. doi:10.1016/j.jsb.2015.08.008

Ron A. M. Fouchier, Nico G. Hartwig, Theo M. Bestebroer, Berend Niemeyer, Jan C. de Jong, James H. Simon, . . . Peter Palese. (2004). A previously undescribed coronavirus associated with respiratory disease in humans. *Proceedings of the National Academy of Sciences of the United States of America, 101*(16), 6212-6216. doi:10.1073/pnas.0400762101

Roseman, A. M. (2004). FindEM--a fast, efficient program for automatic selection of particles from electron micrographs. *Journal of Structural Biology, 145*(1-2), 91-99.
Rosenthal, P. B., & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology, 333*(4), 721-745. doi:10.1016/j.jmb.2003.07.013
Rossen, J.W., de Beer, R., Godeke, G.J., Raamsman, M.J., Horzinek, M.C., Vennema, H., and Rottier, P.J. (1998). The viral spike protein is not involved in the polarized sorting of coronaviruses in epithelial cells. Journal of virology *72*, 497-503.
Routledge, E., Stauber, R., Pfleiderer, M., & Siddell, S. G. (1991). Analysis of murine coronavirus surface glycoprotein functions by using monoclonal antibodies. *Journal of Virology, 65*(1), 254-262.
Russo, C. J., & Passmore, L. A. (2014). Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science (New York, N.Y.), 346*(6215), 1377-1380. doi:10.1126/science.1259530
Sabir, J. S. M., Lam, T. T., Ahmed, M. M. M., Li, L., Shen, Y., Abo-Aba, S. E. M., . . . Guan, Y. (2016). Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in saudi arabia. *Science (New York, N.Y.), 351*(6268), 81-84. doi:10.1126/science.aac8608
Schalk, A. F., Hawn, & M.C. (1931). An apparently new respiratory disease of chicks. *J. Am. Vet. Med. Assoc.,* (78), 413-422.
Scharf, L., Wang, H., Gao, H., Chen, S., McDowall, A.W., and Bjorkman, P.J. (2015). Broadly Neutralizing Antibody 8ANC195 Recognizes Closed and Open States of HIV-1 Env. Cell *162*, 1379-1390.
Scheres, S. H. W. (2012a). A bayesian view on cryo-EM structure determination. *Journal of Molecular Biology, 415*(2), 406-418. doi:10.1016/j.jmb.2011.11.010
Scheres, S. H. W. (2012b). RELION: Implementation of a bayesian approach to cryo-EM structure determination. *Journal of Structural Biology, 180*(3), 519-530. doi:10.1016/j.jsb.2012.09.006
Scheres, S. H. W., & Chen, S. (2012c). Prevention of overfitting in cryo-EM structure determination. *Nature Methods, 9*(9), 853-854. doi:10.1038/nmeth.2115
Scheres, S. H. W., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P. B., Frank, J., & Carazo, J. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods, 4*(1), 27-29. doi:10.1038/nmeth992
Scheres, S. H. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife, 3*, e03665. doi:10.7554/eLife.03665
Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Luo, C., Zhang, W., and Li, F. (2018). Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. PLoS Pathog *14*, e1007009.
Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Tai, W., Du, L., Zhou, Y., Zhang, W., and Li, F. (2017). Cryo-EM structure of porcine delta coronavirus spike protein in the pre-fusion state. Journal of virology.
Snijder, J., Borst, A. J., Dosey, A., Walls, A. C., Burrell, A., Reddy, V. S., . . . Veesler, D. (2017). Vitrification after multiple rounds of sample application and blotting

improves particle density on cryo-electron microscopy grids. *Journal of Structural Biology, 198*(1), 38-42. doi:10.1016/j.jsb.2017.02.008
Snijder, J., Ortego, M.S., Weidle, C., Stuart, A.B., Gray, M.D., McElrath, M.J., Pancera, M., Veesler, D., and McGuire, A.T. (2018). An Antibody Targeting the Fusion Machinery Neutralizes Dual-Tropic Infection and Defines a Site of Vulnerability on Epstein-Barr Virus. Immunity *48*, 799-811 e799.
Sommerstein, R., Flatz, L., Remy, M. M., Malinge, P., Magistrelli, G., Fischer, N., . . . Pinschewer, D. D. (2015). Arenavirus glycan shield promotes neutralizing antibody evasion and protracted infection. *PLoS Pathogens, 11*(11), e1005276. doi:10.1371/journal.ppat.1005276
Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog *14*, e1007236.
Song, Y., DiMaio, F., Wang, R. Y., Kim, D., Miles, C., Brunette, T., . . . Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure, 21*(10), 1735-1742. doi:10.1016/j.str.2013.08.005
Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T., & Dermody, T. S. (2014). The sweet spot: Defining virus-sialic acid interactions. *Nature Reviews. Microbiology, 12*(11), 739-749. doi:10.1038/nrmicro3346
Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., Weber, F., and Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. Virology *361*, 304-315.
Stewart-Jones, G. B. E., Soto, C., Lemmin, T., Chuang, G., Druz, A., Kong, R., . . . Kwong, P. D. (2016). Trimeric HIV-1-env structures define glycan shields from clades A, B, and G. *Cell, 165*(4), 813-826. doi:10.1016/j.cell.2016.04.010
Struwe, W.B., Chertova, E., Allen, J.D., Seabright, G.E., Watanabe, Y., Harvey, D.J., Medina-Ramirez, M., Roser, J.D., Smith, R., Westcott, D. *et al.* (2018). Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. Cell Rep *24*, 1958-1966 e1955.
Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L., & Egelman, E. H. (2016). Resolution advances in cryo-EM enable application to drug discovery. *Current Opinion in Structural Biology, 41*, 194-202. doi:10.1016/j.sbi.2016.07.009
Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L. J., Wong, S. K., . . . Marasco, W. A. (2004). Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proceedings of the National Academy of Sciences of the United States of America, 101*(8), 2536-2541. doi:10.1073/pnas.0307140101
Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., . . . Carragher, B. (2005). Automated molecular microscopy: The new leginon system. *Journal of Structural Biology, 151*(1), 41-60. doi:10.1016/j.jsb.2005.03.010
Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A., & Carfí, A. (2004). Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proceedings of the National Academy of Sciences of the United States of America, 101*(52), 17958-17963. doi:10.1073/pnas.0406128102

Swanson, K. A., Settembre, E. C., Shaw, C. A., Dey, A. K., Rappuoli, R., Mandl, C. W., . . . Carfi, A. (2011). Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers. *Proceedings of the National Academy of Sciences of the United States of America, 108*(23), 9619-9624. doi:10.1073/pnas.1106536108

Suzuki, Y., Ito, T., Suzuki, T., Holland, R. E., Chambers, T. M., Kiso, M., . . . Kawaoka, Y. (2000). Sialic acid species as a determinant of the host range of influenza A viruses. *Journal of Virology, 74*(24), 11825-11831. doi:10.1128/jvi.74.24.11825-11831.2000

Szakonyi, G., Klein, M. G., Hannan, J. P., Young, K. A., Ma, R. Z., Asokan, R., . . . Chen, X. S. (2006). Structure of the epstein-barr virus major envelope glycoprotein. *Nature Structural & Molecular Biology, 13*(11), 996-1001. doi:10.1038/nsmb1161

Taguchi, F., & Matsuyama, S. (2002). Soluble receptor potentiates receptor-independent infection by murine coronavirus. *Journal of Virology, 76*(3), 950-958. doi:10.1128/jvi.76.3.950-958.2002

Tegunov, D., & Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with warp. *Nature Methods, 16*(11), 1146-1152. doi:10.1038/s41592-019-0580-y

Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., . . . Veesler, D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural & Molecular Biology, 26*(6), 481-489. doi:10.1038/s41594-019-0233-y

Traggiai, E., Becker, S., Subbarao, K., Kolesnikova, L., Uematsu, Y., Gismondo, M.R., Murphy, B.R., Rappuoli, R., and Lanzavecchia, A. (2004). An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med *10*, 871-875.

Veesler, D., Blangy, S., Siponen, M., Vincentelli, R., Cambillau, C., & Sciara, G. (2009). Production and biophysical characterization of the CorA transporter from methanosarcina mazei. *Analytical Biochemistry, 388*(1), 115-121. doi:10.1016/j.ab.2009.02.018

Verbeke, E.J., Zhou, Y., Horton, A.P., Mallam, A.L., Taylor, D.W., Marcotte, E.M. (2019). Separating distinct structures of multiple macromolecular assemblies from cryo-EM projections. *BioRxiv.*

Vergara-Alert, J., Brand, Judith M. A. van den, Widagdo, W., Muñoz, M., Raj, S., Schipper, D., . . . Segalés, J. Livestock susceptibility to infection with middle east respiratory syndrome coronavirus. (2017). *Emerging Infectious Diseases Journal*, *23(2).* doi:10.3201/eid2302.161239

Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S., & Carragher, B. (2009). DoG picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *Journal of Structural Biology, 166*(2), 205-213.

Walls, A. C., Tortorici, M. A., Bosch, B., Frenz, B., Rottier, P. J. M., DiMaio, F., . . . Veesler, D. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature, 531*(7592), 114-117. doi:10.1038/nature16988

Walls, A. C., Tortorici, M. A., Frenz, B., Snijder, J., Li, W., Rey, F. A., . . . Veesler, D. (2016a). Glycan shield and epitope masking of a coronavirus spike protein

observed by cryo-electron microscopy. *Nature Structural & Molecular Biology, 23*(10), 899-905. doi:10.1038/nsmb.3293
Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B., Rey, F. A., & Veesler, D. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America, 114*(42), 11157-11162. doi:10.1073/pnas.1708727114
Walls, A., Tortorici, M. A., Bosch, B., Frenz, B., Rottier, P. J. M., DiMaio, F., . . . Veesler, D. (2017a). Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. *Protein Science, 26*(1), 113-121. doi:10.1002/pro.3048
Walls, A. C., Xiong, X., Park, Y., Tortorici, M. A., Snijder, J., Quispe, J., . . . Veesler, D. (2019). Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell, 176*(5), 1026-1039.e15. doi:10.1016/j.cell.2018.12.028
Walls, A., Tortorici, M. A., Bosch, B., Frenz, B., Rottier, P. J. M., DiMaio, F., . . . Veesler, D. (2017). Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. *Protein Science, 26*(1), 113-121. doi:10.1002/pro.3048
Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E. *et al.* (2005). SARS-CoV infection in a restaurant from palm civet. Emerg Infect Dis *11*, 1860-1865.
Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., . . . Wang, X. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Research, 23*(8), 986-993. doi:10.1038/cr.2013.92
Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., . . . Gao, G. (2014). Bat origins of MERS-CoV supported by bat coronavirus HKU4 Usage of human receptor CD26. *Cell Host & Microbe, 16*(3), 328-337. doi:10.1016/j.chom.2014.08.009
Wang, R. Y., Kudryashev, M., Li, X., Egelman, E. H., Basler, M., Cheng, Y., . . . DiMaio, F. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods, 12*(4), 335-338. doi:10.1038/nmeth.3287
Watanabe, Y., Raghwani, J., Allen, J.D., Seabright, G.E., Li, S., Moser, F., Huiskonen, J.T., Strecker, T., Bowden, T.A., and Crispin, M. (2018). Structure of the Lassa virus glycan shield provides a model for immunological resistance. Proceedings of the National Academy of Sciences of the United States of America *115*, 7320-7325.
Weissenhorn, W., Carfí, A., Lee, K. H., Skehel, J. J., & Wiley, D. C. (1998). Crystal structure of the ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Molecular Cell, 2*(5), 605-616. doi:10.1016/s1097-2765(00)80159-8
Wickramasinghe, I. N. A., de Vries, R. P., Gröne, A., de Haan, C. a. M., & Verheije, M. H. (2011). Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *Journal of Virology, 85*(17), 8903-8912. doi:10.1128/JVI.05112-11
Widagdo, W., Raj, V. S., Schipper, D., Kolijn, K., van Leenders, Geert J. L. H., Bosch, B. J., . . . Haagmans, B. L. (2016). Differential expression of the middle east respiratory syndrome coronavirus receptor in the upper respiratory tracts of

humans and dromedary camels. *Journal of Virology, 90*(9), 4838-4842. doi:10.1128/JVI.02994-15
Widjaja, I., Wang, C., van Haperen, R., Gutiérrez-Álvarez, J., van Dieren, B., Okba, N. M. A., . . . Bosch, B. (2019). Towards a solution to MERS: Protective human monoclonal antibodies targeting different domains and functions of the MERS-coronavirus spike glycoprotein. *Emerging Microbes & Infections, 8*(1), 516-530. doi:10.1080/22221751.2019.1597644
Woo, P. C. Y., Lau, S. K. P., Li, K. S. M., Poon, R. W. S., Wong, B. H. L., Tsoi, H., . . . Yuen, K. (2006). Molecular diversity of coronaviruses in bats. *Virology, 351*(1), 180-187. doi:10.1016/j.virol.2006.02.041
Wu, J., Yan, Z., Li, Z., Yan, C., Lu, S., Dong, M., & Yan, N. (2015). Structure of the voltage-gated calcium channel Cav1.1 complex. *Science (New York, N.Y.), 350*(6267), aad2395. doi:10.1126/science.aad2395
Wu, K., Li, W., Peng, G., & Li, F. (2009). Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proceedings of the National Academy of Sciences, 106*(47), 19970-19974.
Wu, S., Armache, J., & Cheng, Y. (2016). Single-particle cryo-EM data acquisition by using direct electron detection camera. *Microscopy (Oxford, England), 65*(1), 35-41. doi:10.1093/jmicro/dfv355
Xiong, X., Tuzikov, A., Coombs, P. J., Martin, S. R., Walker, P. A., Gamblin, S. J., . . . Skehel, J. J. (2013). Recognition of sulphated and fucosylated receptor sialosides by A/vietnam/1194/2004 (H5N1) influenza virus. *Virus Research, 178*(1), 12-14. doi:10.1016/j.virusres.2013.08.007
Xiong, X., Coombs, P. J., Martin, S. R., Liu, J., Xiao, H., McCauley, J. W., . . . Gamblin, S. J. (2013a). Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature, 497*(7449), 392-396. doi:10.1038/nature12144
Xiong, X., Martin, S. R., Haire, L. F., Wharton, S. A., Daniels, R. S., Bennett, M. S., . . . Gamblin, S. J. (2013b). Receptor binding by an H7N9 influenza virus from humans. *Nature, 499*(7459), 496-499. doi:10.1038/nature12372
Xiong, X., Tortorici, M. A., Snijder, J., Yoshioka, C., Walls, A. C., Li, W., . . . Veesler, D. (2018). Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections. *Journal of Virology, 92*(4) doi:10.1128/JVI.01628-17
Xu, Y., Liu, Y., Lou, Z., Qin, L., Li, X., Bai, Z., . . . Rao, Z. (2004). Structural basis for coronavirus-mediated membrane fusion. crystal structure of mouse hepatitis virus spike protein fusion core. *The Journal of Biological Chemistry, 279*(29), 30514-30522. doi:10.1074/jbc.M403760200
Yang, Y., Liu, C., Du, L., Jiang, S., Shi, Z., Baric, R. S., & Li, F. (2015). Two mutations were critical for bat-to-human transmission of middle east respiratory syndrome coronavirus. *Journal of Virology, 89*(17), 9119-9123. doi:10.1128/JVI.01279-15
Yin, H., Paterson, R. G., Wen, X., Lamb, R. A., & Jardetzky, T. S. (2005). Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proceedings of the National Academy of Sciences of the United States of America, 102*(26), 9288-9293. doi:10.1073/pnas.0503989102

Yin, H., Wen, X., Paterson, R. G., Lamb, R. A., & Jardetzky, T. S. (2006). Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature, 439*(7072), 38-44. doi:10.1038/nature04322

Ying, T., Prabakaran, P., Du, L., Shi, W., Feng, Y., Wang, Y., . . . Zhou, T. (2015). Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. *Nature Communications, 6*, 8223. doi:10.1038/ncomms9223

Yu, X., Dang, V. T., Fleming, F. E., von Itzstein, M., Coulson, B. S., & Blanchard, H. (2012). Structural basis of rotavirus strain preference toward N-acetyl- or N-glycolylneuraminic acid-containing receptors. *Journal of Virology, 86*(24), 13456-13466. doi:10.1128/JVI.06975-11

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., . . . Gao, G. F. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications, 8*, 15092. doi:10.1038/ncomms15092

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, Albert D. M. E., & Fouchier, R. A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in saudi arabia. *The New England Journal of Medicine, 367*(19), 1814-1820. doi:10.1056/NEJMoa1211721

Zeng, Q., Langereis, M. A., van Vliet, Arno L. W., Huizinga, E. G., & de Groot, R. J. (2008). Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proceedings of the National Academy of Sciences of the United States of America, 105*(26), 9065-9069. doi:10.1073/pnas.0800502105

Zhang, H., Wang, G., Li, J., Nie, Y., Shi, X., Lian, G., . . . Deng, H. (2004). Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies. *Journal of Virology, 78*(13), 6938-6945. doi:10.1128/JVI.78.13.6938-6945.2004

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology, 193*(1), 1-12. doi:10.1016/j.jsb.2015.11.003

Zheng, B., Liu, S., Du, L., Jiang, S., He, Y., & Zhou, Y. (2009). The spike protein of SARS-CoV - a target for vaccine and therapeutic development. *Nature Reviews Microbiology, 7*(3), 226-236. doi:10.1038/nrmicro2090

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature methods *14*, 331-332.

Zheng, Q., Deng, Y., Liu, J., van der Hoek, L., Berkhout, B., & Lu, M. (2006). Core structure of S2 from the human coronavirus NL63 spike glycoprotein. *Biochemistry, 45*(51), 15205-15215. doi:10.1021/bi061686w

Zhou, Y., Lu, K., Pfefferle, S., Bertram, S., Glowacka, I., Drosten, C., . . . Simmons, G. (2010). A single asparagine-linked glycosylation site of the severe acute respiratory syndrome coronavirus spike glycoprotein facilitates inhibition by mannose-binding lectin through multiple mechanisms. *Journal of Virology, 84*(17), 8753-8764. doi:10.1128/JVI.00554-10

Zhu, Z., Chakraborti, S., He, Y., Roberts, A., Sheahan, T., Xiao, X., . . . Dimitrov, D. S. (2007). Potent cross-reactive neutralization of SARS coronavirus isolates by human

monoclonal antibodies. *Proceedings of the National Academy of Sciences of the United States of America, 104*(29), 12123-12128. doi:10.1073/pnas.0701000104
Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., & Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife, 7* doi:10.7554/eLife.42166
Zivanov, J., Nakane, T., and Scheres, S.H.W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ *6*.