J. Virol. doi:10.1128/JVI.01628-17

Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections

Xiaoli Xiong[1], M. Alejandra Tortorici[2,3], Joost Snijder[1], Craig Yoshioka[4], Alexandra C. Walls[1], Wentao Li[5], Andrew T. McGuire[6], Félix A. Rey[2,3], Berend-Jan Bosch[5] and David Veesler[1]*.

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

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

[3]CNRS UMR 3569 Virologie, Paris, France.

[4]Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97201, USA.

[5]Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL, Utrecht, The Netherlands.

[6]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N. P.O. Box 19024 Seattle, WA 98109, USA

*Correspondence: dveesler@uw.edu

Abstract

Coronaviruses recently emerged as major human pathogens causing outbreaks of severe acute respiratory syndrome and Middle-East respiratory syndrome. They utilize the spike (S) glycoprotein anchored in the viral envelope to mediate host attachment and fusion of the viral and cellular membranes to initiate infection. The S protein is a major determinant of the zoonotic potential of coronaviruses and is also the main target of the host humoral immune response. We report here the 3.5 Å resolution cryo-electron microscopy structure of the S glycoprotein trimer from the pathogenic porcine deltacoronavirus (PDCoV), which belongs to the recently identified delta genus. Structural and glycoproteomics data indicate that the glycans of PDCoV S are topologically conserved when compared with the human respiratory coronavirus HCoV-NL63 S, resulting in similar surface areas being shielded from neutralizing antibodies and implying that both viruses are under comparable immune pressure in their respective hosts. The structure further reveals a shortened S_2' activation loop, containing a reduced number of basic amino acids, which participates to rendering the spike largely protease-resistant. This property distinguishes PDCoV S from recently characterized betacoronavirus S proteins and suggests that the S protein of enterotropic PDCoV has evolved to tolerate the protease-rich environment of the small intestine and to fine-tune its fusion activation to avoid premature triggering and reduction of infectivity.

Importance

Coronaviruses use transmembrane spike (S) glycoprotein trimers to promote host attachment and fusion of the viral and cellular membranes. We determined a near-atomic resolution cryo-electron microscopy structure of the S ectodomain trimer from the pathogenic porcine deltacoronavirus (PDCoV), which is responsible for diarrhea in piglets and has had devastating consequences for the swine industry worldwide. Structural and glycoproteomics data reveal that PDCoV S is decorated with 78 N-linked glycans obstructing the protein surface to limit accessibility to neutralizing antibodies in a way reminiscent of what has recently been described for a human respiratory coronavirus. PDCoV S is largely protease-resistant which distinguishes it from most other characterized coronavirus S glycoproteins and suggests that enteric coronaviruses have evolved to fine-tune fusion activation in the protease-rich environment of the small intestine of infected hosts.

Introduction

Coronaviruses are large enveloped viruses, with single-stranded positive-sense RNA genomes, classified in four genera (α, β, γ, and δ) based on their sequence similarity. Most recognized coronaviruses are animal viruses but four coronaviruses, namely HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1, are known to continuously circulate in the human population and are associated with up to 30% of respiratory tract infections(1). In addition, severe acute respiratory syndrome (SARS-CoV) and Middle-East respiratory syndrome (MERS-CoV) coronaviruses are zoonotic viruses causing deadly pneumonia in humans(2). SARS-CoV and MERS-CoV have resulted in more than 8,000 and 2,000 cases with fatality rates of 10 and 35%, respectively. No specific antiviral treatments or vaccines are approved for human coronaviruses and zoonosis remains a great pandemic threat.

The ability to recognize the appropriate receptor and to efficiently enter host cells are key requirements for cross-species spillover of zoonotic viruses such as influenza(3). For coronaviruses, these two functions are carried out by the spike (S) glycoprotein. Therefore, structural and functional studies of S glycoproteins can provide invaluable information to evaluate the cross-species transmission potential of these viruses. The coronavirus S protein is a class I viral fusion protein that forms homotrimers decorating the viral envelope. It is composed of an N-terminal S_1 subunit, responsible for receptor-binding, and a C-terminal S_2 subunit, which contains the fusion machinery. The combined activities of the two subunits promote coronavirus attachment to host cells

and subsequent fusion of the viral and cellular membranes, via irreversible conformational changes, initiating viral infection. Since it is the major surface protein, S is also the main target of neutralizing antibodies during infection and a focus of vaccine design.

The zoonotic potential of coronaviruses is determined by the receptor-binding properties of the S protein. For instance, SARS-CoV and MERS-CoV bind with high-affinity to their cognate human receptors, angiotensin-converting enzyme 2 (ACE2) and dipeptidyl peptidase 4 (DPP4), respectively(4, 5). Metagenomic data revealed that many MERS-CoV and SARS-CoV-like viruses exist in bats and one such virus, WIV-1, isolated from bat feces, shares 99.9% nucleotide sequence identity with SARS-CoV. The S protein encoded by WIV-1 binds human, bat and civet ACE2 orthologues allowing the virus to efficiently infect human cells expressing any of these three orthologues(6, 7). Similarly, HKU4-CoV and HKU5-CoV that are closely related to MERS-CoV have been identified in bats and HKU4-CoV can be adapted to bind human DPP4 by substituting three amino acids in the S receptor-binding domain(8, 9).

The zoonotic potential of coronaviruses is further determined by fusion activation which requires S processing by host proteases. Up to two cleavage sites are present in S glycoproteins: a site found at the boundary between the S_1 and S_2 subunits of some coronavirus S (the S_1/S_2 site) and a conserved site upstream from the fusion peptide (the S_2' site)(10).

For a subset of coronaviruses, such as MHV, SARS-CoV and MERS-CoV, the S glycoprotein is cleaved at the S_1/S_2 junction during biogenesis and viral egress(10-13). This proteolytic event, along with subsequent binding to the host receptor, enhances processing at the S_2' site and participates in MERS-CoV or SARS-CoV fusion activation(11, 13). Moreover, substitution of two residues at the boundary between the S_1 and S_2 subunits enables efficient processing by human proteases and allows the bat-infecting HKU4-CoV S protein to mediate entry into human cells(14).

Proteolysis at the conserved S_2' site is essential for fusion activation of all characterized coronavirus S proteins and it can occur at the host membrane or in internal cellular compartments. For instance, transmembrane protease/serine protease (TMPRSS) processing of SARS-CoV and MERS-CoV S at the cell membrane, furin-mediated processing of HCoV-NL63 and MERS-CoV S in the early endosomes, or endo-lysosomal protease-mediated triggering of SARS-CoV S (by cathepsin L) and MHV S are key events orchestrating spatial and temporal activation of fusion to ensure successful viral entry into host cells(12, 13, 15). Alternatively, porcine epidemic diarrhea coronavirus (PEDV), which replicates in the epithelial cells of the small intestine, undergoes S proteolytic activation by trypsin, which is highly abundant in the lumen of this organ(16). These examples illustrate how the availability of host proteases and the mechanism of proteolytic activation can directly restrict coronavirus activation, viral tropism, and pathogenesis.

138 One common pattern shared by both SARS and MERS outbreaks is that although they
139 both originated in bats, an intermediate host with closer physical proximity to humans
140 allowed for more efficient cross-species transmission. Palm civets and camels were the
141 most probable intermediate hosts for SARS-CoV and MERS-CoV, respectively(7, 17,
142 18). Due to their proximity with humans, pigs also acted as intermediate hosts for the
143 influenza pandemic (19) and for the emergence of Nipah virus in Malaysia(20). To date,
144 only α- and β-coronaviruses have been implicated in human diseases and several S
145 glycoproteins from viruses belonging to these two genera have been structurally
146 characterized(21-26). To the best of our knowledge, no porcine coronaviruses have
147 crossed the species barrier to infect humans, and their receptor usage appears to favor
148 porcine orthologues. Porcine epidemic diarrhea virus (PEDV), however, can infect pig,
149 human, monkey and bat cells, suggesting it has the potential to spillover to species
150 other than pig(27). As a result, cross-species transmission of coronaviruses poses an
151 imminent and long-term threat to human health which emphasizes the need for
152 surveying and studying these viruses to prevent and control infections.
153
154 The recently emerged porcine deltacoronavirus (PDCoV) is responsible for diarrhea in
155 piglets and has had devastating consequences for the swine industry worldwide(28, 29).
156 No vaccines or treatments are available for PDCoV. Here, we report the cryoEM
157 structure of the PDCoV S trimer revealing that it has a molecular architecture most
158 closely related to the S glycoproteins of the α-genus of coronaviruses. Integrating
159 structural and glycoproteomics data, we discovered that PDCoV S masks potential
160 epitopes with glycans in a way reminiscent of the human respiratory α-coronavirus

HCoV-NL63 S glycoprotein(22). These results support a relatedness between α- and δ-coronavirus S glycoproteins and suggest that the immune system of infected hosts exert comparable selection pressure on these viruses which has led to these adaptations. The structure also reveals the C-terminal S_2 fusion machinery of the PDCoV S protein features a short S_2' activation loop which appears to be largely resistant to proteolysis by trypsin/chymotrypsin. We conclude that PDCoV has evolved to be highly adapted to the protease-rich environment of the enteric tract to ensure proper spatial and temporal activation of fusion and prevent premature triggering which would significantly impact virus infectivity.

Results

Structure determination of the PDCoV S glycoprotein

PDCoV was first identified in Hong Kong in 2012(29) and it has since spread rapidly in the swine population across the globe(28, 29). Due to its recent emergence, relatively little is known about this virus compared to other swine coronaviruses. One feature that distinguishes PDCoV from other known coronaviruses is that it encodes one of the smallest S glycoproteins. We therefore set out to explore the architectural diversity of S proteins across coronavirus genera to understand shared and unique features of the structurally uncharacterized δ-genus.

We used *Drosophila* S2 cells to produce the PDCoV/USA/Illinois121/2014 S ectodomain (residues 1-1098) with a C-terminal fusion adding a GCN4 trimerization motif and a strep-tag(30). Following sample vitrification by triple blotting(31), data were

184 acquired on an FEI Titan Krios electron microscope equipped with a Gatan Quantum
185 GIF energy filter operated in zero-loss mode and a Gatan K2 Summit electron-counting
186 camera operated in super-resolution mode (Fig 1A-B). We determined a 3D
187 reconstruction at 3.5 Å resolution resolving most amino acid side chains, disulphide
188 bonds and N-linked glycans (Fig S1A). These features were used as fiducials to confirm
189 the sequence register during model building (Fig 1C-F and S1B-E Fig). Starting from the
190 HCoV-NL63 S structure(22), we obtained an atomic model of the PDCoV S trimer using
191 manual modeling in Coot(32) and Rosetta density-guided iterative refinement(33). The
192 final model comprises residue 52 to 1021 and 21 N-linked glycans (Table 1).
193
194 The PDCoV S protein assembles as a compact trimer with a height of ~145 Å and a
195 width of 115 Å (Fig 1C-D). The S_1 subunit has a modular organization comprising four
196 distinct domains, designated A, B, C and D, whereas the S_2 subunit adopts a mostly-
197 helical elongated architecture with a connector domain appended to its C-terminal
198 end(21, 22) (Fig 1E-F).
199
200 **The extensive PDCoV S glycan shield**
201 The unsharpened PDCoV S map resolves 21 N-linked glycans for each protomer that
202 form prominent protrusions extending from the protein surface (Fig 2A-B and Fig S1 F-
203 G). Using on-line reversed phased liquid chromatography with electron transfer/high-
204 energy collision-dissociation tandem mass-spectrometry(34), we detected 16 N-linked
205 glycosylation sites corresponding to those observed in the cryoEM map and confirmed 5
206 additional sites located in the structurally unresolved N and C-terminal parts of the

protein (Fig 2C and Table S1). Combining our structural and mass-spectrometry data, we found evidence for glycosylation at 26 out of 27 possible NXS/T glycosylation sequons. The intact glycopeptides detected by MS/MS for PDCoV S expressed in *Drosophila* S2 cells corresponded mostly to paucimannosidic glycans containing 3 mannose residues (with or without core fucosylation) and oligomannose glycans containing 4 to 9 mannose residues. We also detected complex glycans (with or without core fucosylation), which appears compatible with the accessibility and crowding of these carbohydrate chains that would permit processing(35, 36).

Overall, the glycan coverage of PDCoV S is dense and extensively decorates the accessible surface of the trimer. Although we detected substantially more N-linked glycans for HCoV-NL63 S(22) (34 sites per protomer), 6 validated glycans reside within the N-terminal domain 0, which is absent in PDCoV S and explains most of the discrepancy in the number of sites. Strikingly, numerous glycans identified in the PDCoV S structure overlap with glycans in the HCoV-NL63 S protein, either strictly or topologically, with most differences towards the viral membrane distal end of the molecule (Fig 2D-E). Transmission of zoonotic viruses into humans can result in drastic changes in glycosylation, as exemplified by the human influenza H3 hemagglutinin that has doubled its number of glycosylation sites since the 1968 pandemic although its amino acid sequence remains ~88% identical(37). There is considerable sequence divergence between the HCoV-NL63 and PDCoV S glycoproteins, which share 43% amino acid sequence identity. The observation that numerous glycosylation sites are conserved between the two proteins suggest that α- and δ-coronaviruses could face

similar immune pressure in their respective hosts, and that the areas that are masked by the conserved glycans are key to the function of these S glycoproteins. Based on the information gained from the HCoV-NL63 S structure(22), for which glycans appears to contribute to masking the receptor-binding loops from antibody recognition, we suggest that the glycan shield of PDCoV S and other coronavirus S glycoproteins could assist in immune evasion similarly to the well-characterized HIV-1 envelope trimer(35, 36).

Finally, coronavirus S glycans have previously been proposed to participate in host cell entry(38), since L-SIGN lectin can be used as an alternative receptor by SARS-CoV(39) and HCoV-229E(40), and it is conceivable they play a similar role for other S proteins.

Architecture of the S_1 receptor-binding subunit

The PDCoV and HCoV-NL63 S_1 subunits exhibit strikingly similar structures (r.m.s.d.=2.7 Å over 448 aligned Cα positions), except for the absence of the N-terminal domain 0 in the former glycoprotein (Fig 3A). Deletion of domain 0, which is responsible for attachment to sialoglycans, in the porcine transmissible gastro-enteritis virus (TGEV) S gene, gave rise to porcine respiratory coronavirus (PRCV) and in turn resulted in a loss of enteric tropism(41, 42). PDCoV and HCoV-NL63, however, exhibit opposite behavior, as they target the enteric or the respiratory tracts despite the absence or presence of a domain 0 in their S glycoproteins, respectively. We describe below the functionally-relevant similarities and differences detected in the PDCoV S structure relative to other coronavirus S structures.

253 Domain A is located at the viral membrane distal side and account for a large part of the
254 exposed surface area of the S_1 subunit. It folds as a galectin-like β-sandwich
255 supplemented with a helix on the viral membrane distal side and a three-stranded
256 antiparallel β-sheet plus a helix on the proximal side. The domain A surface is heavily
257 glycosylated and features 7 glycans for PDCoV (Fig 3D). We previously reported that
258 the HCoV-NL63 S glycan linked to Asn358 (domain A) points towards the receptor-
259 binding domain B, masking residues involved in receptor recognition. A marked
260 difference between the A domains of PDCoV S and HCoV-NL63 S is that the β-hairpin
261 harboring Asn358 in HCoV-NL63 S features a deletion of 10 residues significantly
262 shortening it in PDCoV S (Fig 3A-C). Moreover, the topologically equivalent glycan
263 linked to residue Asn-184 of PDCoV S is protruding away from domain B and does not
264 significantly cover it, in contrast to what was observed for HCoV-NL63 (Fig 3A-C).
265
266 OC43, HKU1 and bovine coronavirus (BCoV) are known to use 9-O-acetyl-sialylated
267 cellular receptors for attachment to host cells(43, 44). Structural and biochemical
268 studies showed that domain A mediates these interactions and mapped key residues
269 involved(25, 45) and nanoparticle-displayed multimeric OC43 S_1 subunit exhibited high
270 hemagglutination titer (Fig. 3 F). Comparison of the PDCoV, HKU1 and BCoV domain A
271 structures indicated PDCoV cannot interact with 9-O-acetyl-sialoglycans in a similar way
272 due to the absence of the strictly conserved residues involved in binding (BCoV Tyr162,
273 Glu182, Trp184 and His185) and of the loops delineating the binding cavity (Fig. 3 D-E).
274 In line with this observation, isolated or nanoparticle-displayed multimeric PDCoV S_1
275 subunit failed to interact with sialic acid using an erythrocyte hemagglutination assay

(Fig. 3 F), indicating that sialic acid (or at least the types of sialoglycans displayed on these erythrocytes) does not participate in PDCoV S attachment to host cells.

Domain B folds as a β-sandwich reminiscent of the equivalent domain of α-coronaviruses such as PRCV (r.m.s.d.= 2.1 Å over 108 aligned Cα positions), HCoV-NL63 (r.m.s.d.= 1.9 Å over 107 aligned Cα positions) and TGEV (r.m.s.d.= 3.0 Å over 109 aligned Cα positions) (Fig 4A-D)(22, 46, 47). The two PDCoV S glycosylation sites identified at Asn-311 and Asn-331 in domain B are topologically or strictly conserved with the HCoV-NL63 S glycans linked to Asn-486 and Asn-512, respectively. PRCV and TGEV B domains also feature topologically similar glycosylation sites on the solvent exposed surface of the β-sandwich and these glycans are likely to limit the immune response against this domain which is known to be the target of neutralizing antibodies for several coronaviruses(47-52). The glycan linked to Asn-506 in HCoV-NL63 S is absent in PDCoV S for which the equivalent residue is Ser-325 (Fig 4A and C). Since masking of receptor-binding residues has been suggested to assist HCoV-NL63 immune evasion(22), the reduced overall glycan coverage of PDCoV domain B could result from weaker immune pressure directed at the receptor-binding region in pigs compared to HCoV-NL63 S in humans.

Previous work showed that the loops located at the viral membrane distal end of the β-sandwich of domain B in α-coronavirus S glycoproteins are responsible for binding to diverse host receptors such as ACE2 (HCoV-NL63)(46) or pAPN (PRCV/TGEV)(47). Although the distal loops are significantly shorter for PDCoV compared to these two α-

299 coronaviruses, loop 1 and loop 3 contain several aromatic residues (Fig 4A-D). Since
300 aromatic residues in these loops have been shown to directly participate in receptor-
301 binding for HCoV-NL63, PRCV and TGEV, we speculate that they could also mediate
302 interactions of the PDCoV B domain with its receptor. As is the case for HCoV-NL63 S,
303 the PDCoV B domain has an opposite orientation, related by a ~180˚ rotation, to the
304 equivalent domain of β-coronavirus S glycoproteins(21). This results in burying the
305 distal loops of the β-sandwich through interactions with domain A belonging to the same
306 protomer and in turn restrain the availability of the putative receptor-binding motif to
307 interact with the receptor (Fig 3B-C). As a result, it is likely that PDCoV and HCoV-NL63
308 S glycoproteins can undergo similar conformational changes to those described for
309 domain B of SARS-CoV and MERS-CoV S to interact with their cognate receptors(23,
310 24, 26). A major difference, however, is that β-coronavirus S using domain B as
311 receptor-binding domain appear to spontaneously undergo these rearrangements
312 whereas α- and δ-coronavirus S do not and rely on a yet unidentified stimulus(22).
313
314 **Organization of the S_2 fusion machinery**
315 The C-terminal S_2 subunit trimer fuses the viral and cellular membranes at the onset of
316 infection and is the most conserved region among coronavirus S glycoproteins. The
317 PDCoV S_2 subunit is structurally similar to α- and β-coronavirus S_2 subunits such as
318 HCoV-NL63(22) (r.m.s.d.= 1.7 Å over 413 aligned Cα positions) and MHV(21) (r.m.s.d.=
319 2.2 Å over 291 aligned Cα positions), respectively. The conserved architecture of the S_2
320 fusion machinery across multiple genera highlights that coronaviruses rely on a
321 common fusion mechanism to enter host cells(53). Despite these striking similarities,

322 coronavirus fusion machineries exhibit differences with key functional implications for
323 their activation mechanism and potential for zoonotic spillover.
324
325 The S_2' activation loop, which connects the upstream helix to the fusion peptide and
326 regulates the spatial and temporal activation of fusion, is resolved in the PDCoV S
327 cryoEM map (Fig 5A-B), as was the case for the HCoV-NL63 S(22) and SARS-CoV S
328 structures(23, 24). However, the PDCoV S_2' loop is short (LTTRIGGR) and comprises 6
329 and 3 fewer residues than the HCoV-NL63 S (LPQRNIRSSRIAGR) and SARS-CoV S
330 (ILPDPLKPTKR) counterparts, respectively. In addition, the S_2' loop of these viruses
331 contains multiple positive charges, including two putative furin cleavage sites for HCoV-
332 NL63 S, whereas the PDCoV S_2' loop harbors a single positively charged residue (Arg-
333 669) in addition to the conserved Arg-673 residue (Fig 5A-B). These structural features
334 allow rationalizing the known protease requirements for fusion activation of the HCoV-
335 NL63 S glycoprotein, which is preferentially cleaved by furin in the endosomes, and of
336 the SARS-CoV S glycoprotein, which is preferentially processed by cathepsin L in the
337 endo-lysosomes, and explain the fact that trypsin-like TMPRSS proteases can also
338 trigger both proteins(10, 15). The paucity of positive charges in the PDCoV S_2' trigger
339 loop is in line with the requirement for trypsin or other pancreatic proteases to allow
340 virus passaging and the fact that PDCoV is exposed to high concentration of such
341 proteases in the enteric tract of infected pigs(28).
342
343 Studies on influenza hemagglutinin highlighted that glycans can modulate cleavage site
344 accessibility to proteases and in turn influence fusion activation(54, 55). Similar

observations were drawn from comparisons between the MERS-CoV and HKU4 S glycoproteins(14). Notably, the PDCoV S glycans linked to Asn-652 and Asn-661 decorate the trimer surface near the S_2' trigger loop and could limit accessibility to proteases and play a role in orchestrating fusion activation (Fig 5B). These glycans are conserved with an identical structural organization in HCoV-NL63 S and may have the same putative function(22).

Sequence alignment of representative S glycoproteins from viruses of the four coronavirus genera show that α- and δ-coronaviruses feature a 14-residue long insertion in the heptad-repeat 1 (HR1) and in the HR2 motifs, corresponding to two heptad-repeats, compared to β-coronaviruses (S2A Fig). γ-coronaviruses form a heterogeneous group comprising S glycoproteins without insertion but also S with one (BeCoV-SW1 and BdCoV-HKU22) or two (TurkeyCoV-MG10) additional heptad-repeats in HR1 and in HR2 compared to β-coronaviruses. The HR1 insertion is resolved in the PDCoV S cryoEM map (Fig 5C, residues 797-811) and corresponds to the addition of two helical turns (also visible in the HCoV-NL63 S structure) preceded by a loop (poorly resolved in the HCoV-NL63 S structure). This polypeptide segment is known to refold to form a central triple helical coiled-coil in the post-fusion S structure(53). The HR2 insertion cannot be visualized as this region is disordered in the PDCoV S reconstruction and in all other coronavirus pre-fusion S structures. Mapping the HR1 and HR2 insertions on the HCoV-NL63 S post-fusion core X-ray structure(56), however, reveals that these polypeptide segments are directly interacting within the 6-helix bundle (S2B Fig). This suggests that the strict correlation of their presence or absence in both

HR1 and HR2, along with the observation that insertions are always corresponding to an integer number of heptad repeats, is necessary to maintain the proper geometry of the fusion machinery and allow the conserved conformational changes driving membrane merger to take place with high efficiency.

Discussion

Structural and functional studies of coronavirus S glycoproteins are key to understanding host and tissue tropism as well as the mechanisms of receptor binding and fusion activation. The data reported in this manuscript establishes a strong connection between α- and δ-coronavirus S glycoproteins. HCoV-NL63, PRCV, TGEV and PDCoV B domains fold as similar β-sandwiches that are structurally distinct from the single β-sheet observed for the equivalent domain of β-coronaviruses(4, 5, 21, 25, 47). Moreover, the structures of HCoV-NL63 S and PDCoV S show that both glycoproteins share a common organization of their S_1 subunits in which the B domain directly interact with domain A from the same subunit to potentially limit accessibility of the receptor-binding loops to neutralizing antibodies. Sequence analysis indicates a strict correlation of the presence or absence of the HR1/HR2 insertions in the S glycoprotein sequence and an apparent evolutionary pressure restricting insertions/deletions to heptad-repeat units which we postulate to be necessary for efficient S refolding and fusion. Based on these criteria, we put forward that α- and δ-coronavirus S glycoproteins share closer evolutionary relationships with each other than they do with S of the other two coronavirus genera although insertions in HR1 and HR2 have also been detected in a subset of γ-coronavirus S proteins.

391

392 We previously recapitulated *in vitro* the proteolytic activation of MHV, SARS-CoV and
393 MERS-CoV pre-fusion S trimers, via trypsin incubation under limited proteolysis
394 conditions, which led to spontaneous refolding into post-fusion S_2 trimers (the ground
395 state of the fusion reaction)(53). In contrast, the PDCoV S glycoprotein remained largely
396 uncleaved even after extended incubation times with up to 5:1 molar ratio of S to trypsin
397 or chymotrypsin (0.1 mg/ml) (Figure 6). These results suggest that fusion activation of
398 PDCoV S, which is believed to be promoted by trypsin(28), involves an additional step
399 to expose the S_2' cleavage site, such as the receptor-binding induced conformational
400 changes described for MERS-CoV(13), SARS-CoV(57), MHV(58) and PEDV(16).
401 These distinct protease sensitivities are reminiscent of the differences reported between
402 clinical isolates (CV777) and cell-culture adapted (caDR13) PEDV strains for which
403 infectivity strictly requires or is hampered by trypsin, respectively(16). We put forward
404 that the S glycoprotein sequence and in turn structure of PDCoV and PEDV CV777
405 have evolved to be resistant to pancreatic proteases to which both viruses are exposed
406 in the enteric tract during infection. Fine-tuning of fusion activation is likely achieved by
407 restraining access to the S_2' cleavage site until receptor-binding occurs at the host cell
408 surface. This event could promote conformational changes exposing the S_2' site to allow
409 processing by trypsin or other proteases with exquisite spatial and temporal
410 coordination. In contrast, SARS-CoV, MERS-CoV or MHV are not expected to be
411 exposed to pancreatic proteases during the virus life-cycle and their S glycoproteins
412 presumably did not evolve with this selection pressure, explaining their sensitivity to
413 trypsin and chymotrypsin. In agreement with what has been postulated for SARS-

CoV(57) and PEDV caDR13(16), trypsin sensitivity could result in premature cleavage/triggering of the pre-fusion S trimer and attenuation of infectivity and viral fitness.

While completing this study, another group also determined a cryoEM reconstruction of the PDCoV S glycoprotein ectodomain and both structures can be superposed with excellent agreement (r.m.s.d. of 1.1 Å over 959 aligned Cα positions).

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences under Award Number 1R01GM120553 (DV) and T32GM008268 (A.C.W.) and a Pew Biomedical Scholars Award (D.V.). M.A.T. and F.A.R. acknowledge support from CNRS and Institut Pasteur recurrent funding. J.S. acknowledges support from the Netherlands Organization for Scientific Research (NWO, Rubicon 019.2015.2.310.006) and the European Molecular Biology Organisation (EMBO, ALTF 933-2015). Microscopy was performed at the Multi-scale Microscopy Core (MMC) at Oregon Health & Science University (OHSU) with technical support from the OHSU-FEI Living Lab and the OHSU Center for Spatial Systems Biomedicine (OCSSB). This work was also partly supported by the University of Washington's Proteomics Resource (UWPR95794).

Methods

Plasmids

A gene fragment encoding the PDCoV S ectodomain (residues 20-1098, Uniprot: W8Q9Y7) 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)(21, 59, 60), a thrombin cleavage site (LVPRGSLE), an 8-residue long Strep-Tag (WSHPQFEK) and a stop codon, as previously described(61). 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.

A human codon-optimized gene encoding for the ectodomain (residues 14-1180) of the SARS-CoV S protein (UniProt: P59594) was cloned into a modified pOPING vector(62) (Addgene) introducing a N-terminal Mu-phosphatase signal peptide and a C-terminal TEV protease cleavage site, a foldon and a hexa-histidine tag at the C-terminus of the construct.

Production of recombinant PDCoV S ectodomain in *Drosophila* S2 cells

To generate a stable *Drosophila* S2 cell line expressing the recombinant PDCoV S ectodomain, we used Effectene (Qiagen) and 2 µg of plasmid. Puromycin N-acetyl transferase was co-transfected as dominant selectable marker. Stable PDCoV S expressing cell lines were selected by addition of 7 µg/ml Puromycin (Invitrogen) 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 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

460 to pH 8.0, before affinity purification using StrepTactin Superflow column (IBA) followed
461 by gel filtration chromatography using Superose 6 10/300 GL column (GE Life
462 Sciences) equilibrated in 20 mM Tris-HCl pH 7.5 and 100 mM NaCl. The concentration
463 of the purified protein was estimated using absorption at 280 nm.
464
465 **Production of recombinant SARS S ectodomain in HEK293F cells**
466 Transient transfection of 250mL HEK293F cells at a density of 10^6 cells/mL was
467 performed using 293fectin (ThermoFisher) and Optimem (ThermoFisher). After 3 days
468 the cells were harvested before affinity purification with a Talon 5mL cobalt column
469 equilibrated in 25mM sodium phosphate pH 8.0, 300mM NaCl, 10mM Imidazole. The
470 purified protein was buffer exchanged into 20mM Tris pH 8.0, 150mM NaCl and
471 concentrated to 1.0mg/mL.
472
473 **CryoEM specimen preparation and data collection**
474 Two microliters of purified PDCoV S at ~ 0.5 mg/mL was triple-blotted(31) using 1.2/1.3
475 C-flat grids (Protochips), which had been glow discharged for 30 seconds at 20mA.
476 Grids were then plunge-frozen in liquid ethane using an FEI Mark I Vitrobot with 7.5
477 seconds blot time and an offset of -3mm at 100% humidity and 25°C. Data was
478 collected using SerialEM automatic data collection software(63) on a FEI Titan Krios
479 operated at 300kV and equipped with a Gatan Quantum GIF energy-filter operated in
480 zero-loss mode with a slit width of 20 eV and a Gatan K2 Summit direct electron
481 detector camera operated in super-resolution mode. The dose rate was adjusted to ~5
482 counts/pixel/s and each movie was acquired in counting mode fractionated in 75 frames

of 200 ms. 2,000 micrographs were collected in a single session using a defocus range comprised between 1.5 and 4.0 µm.

CryoEM data processing

Frame alignment was carried out using Motioncor2(64). The parameters of the microscope contrast transfer function were initially estimated using GCTF(65) and then using CTFFIND4(66). Particles were automatically picked using DoGPicker(67). Particle images were extracted and processed using Relion 2.0(68) with a box size of 640 pixels^2 and a pixel size of 0.665 Å. Following reference-free 2D classification, we ran 3D classification with C1 symmetry(69) using an initial model generated with e2initialmodel.py in EMAN2. 455,710 particles were selected to run a gold-standard 3D refinement imposing C3 symmetry using Relion 2.1 (70) that led to a map at 3.5 Å resolution. Post processing was done using Relion to apply an automatically generated B factor of -150 Å^2. Reported resolutions are based on the gold-standard FSC=0.143 criterion(70, 71) and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution(72). The soft mask used for FSC calculation had a 10 pixel cosine edge fall-off.

Model building and analysis

UCSF Chimera(73) was used to fit the HCoV-NL63 S structure(22) into the cryoEM map before manual rebuilding in Coot(32, 74). Glycan density coming after an NXS/T motif was initially hand built into the density where visible and glycan geometry was then refined using Rosetta, optimizing the fit-to-density as well as the energetics of

506 protein/glycan contacts. The final model was refined using the symmetric modeling
507 framework in Rosetta(33, 75). The quality of the final model was analyzed with
508 Molprobity(76) and Privateer(77). All figures were generated with UCSF Chimera(73).
509
510 **Mass Spectrometry**
511 250 pmol of PDCoV S was incubated in a freshly prepared solution containing 100mM
512 Tris pH 8.5, 2% sodium deoxycholate, 10mM tris(2-carboxyethyl)phosphine, and 40mM
513 iodoacetamide at 95 °C for five minutes followed by 25 °C for thirty minutes in the dark.
514 80 pmol of denatured, reduced, and alkylated PDCoV S was then diluted into freshly
515 made 50mM ammonium bicarbonate and incubated for 14 hours at 37 °C either with
516 1:75 (w:w) of trypsin (Sigma Aldrich), or chymotrypsin (Sigma Aldrich) or alpha lytic
517 protease (Sigma Aldrich). Formic acid was then added to a final concentration of 2% to
518 precipitate the sodium deoxycholate in the samples, followed by centrifugation at 14,000
519 rpm for 20 minutes. The supernatant containing the (glyco)-peptides was collected and
520 spun again at 14,000 rpm for 5 min immediately before sample analysis. For each
521 sample 8 µL was injected on a Thermo Scientific Orbitrap Fusion Tribrid mass
522 spectrometer. A 35-cm analytical column and a 3-cm trap column filled with ReproSil-
523 Pur C18AQ 5 µm (Dr. Maisch) beads were used. Nanospray LC-MS/MS was used to
524 separate peptides over a 110-min gradient from 5% to 30% acetonitrile with 0.1% formic
525 acid. A positive spray voltage of 2,100 was used with an ion-transfer-tube temperature
526 of 350 °C. An electron-transfer/higher-energy collision dissociation ion-fragmentation
527 scheme (34) was used with calibrated charge-dependent ETD parameters and
528 supplemental higher-energy collision dissociation energy of 0.15. A resolution setting of

120,000 with an AGC target of 2 × 105 was used for MS1, and a resolution setting of 30,000 with an AGC target of 1 × 105 was used for MS2. Data was searched with Protein Metrics Byonic software (78), using a small custom database of recombinant protein sequences including several coronavirus spike proteins, other viral glycoproteins and 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), F/W/Y/M/L (chymotrypsin) or T/A/S/V (alpha lytic protease) 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 cyteines was set as fixed modification, methionine oxidation as variable modification, and all four software-provided N-linked glycan databases were combined into a single non-redundant list used to identify glycopeptides. All glycopeptide hits were manually inspected and only those with quality peptide sequence information are reported here.

Proteolysis of PDCoV S and SARS S glycoproteins

Proteins at a concentration of 0.5 mg/mL (PDCoV S) or 1mg/mL (SARS-CoV S) were incubated with 0.1 mg/mL of either trypsin (SigmaAldrich) or chymotrypsin at 22 °C for two hours. This reaction was then used for analysis by SDS-PAGE.

Accession number(s)

The mass-spectrometry data have been deposited to PRIDE with accession code PXD007107 and includes the raw data, Byonic search results and the databases

used for protein sequences and N-linked glycan modifications. The EM map and PDB model have been deposited with accession codes EMD-7094 and 6BFU.

References

1. **Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY.** 2016. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov **15:**327-347.
2. **Vijay R, Perlman S.** 2016. Middle East respiratory syndrome and severe acute respiratory syndrome. Curr Opin Virol **16:**70-76.
3. **Schrauwen EJ, Fouchier RA.** 2014. Host adaptation and transmission of influenza A viruses in mammals. Emerg Microbes Infect **3:**e9.
4. **Li F, Li W, Farzan M, Harrison SC.** 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science **309:**1864-1868.
5. **Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, Zhang Y, Zhang W, Yuan Y, Bao J, Zhang B, Shi Y, Yan J, Gao GF.** 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature **500:**227-231.
6. **Menachery VD, Yount BL, Jr., Sims AC, Debbink K, Agnihothram SS, Gralinski LE, Graham RL, Scobey T, Plante JA, Royal SR, Swanstrom J, Sheahan TP, Pickles RJ, Corti D, Randell SH, Lanzavecchia A, Marasco WA, Baric RS.** 2016. SARS-like WIV1-CoV poised for human emergence. Proc Natl Acad Sci U S A **113:**3048-3053.
7. **Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Crameri G, Zhang SY, Wang LF, Daszak P, Shi ZL.** 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature **503:**535-538.
8. **Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, Baric RS, Jiang S, Li F.** 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc Natl Acad Sci U S A **111:**12516-12521.
9. **Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, Ji W, Li Y, Wu Y, Wang J, Iwamoto A, Woo PC, Yuen KY, Yan J, Lu G, Gao GF.** 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. Cell Host Microbe **16:**328-337.
10. **Millet JK, Whittaker GR.** 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res **202:**120-134.
11. **Belouzard S, Chu VC, Whittaker GR.** 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A **106:**5871-5876.
12. **Burkard C, Verheije MH, Wicht O, van Kasteren SI, van Kuppeveld FJ, Haagmans BL, Pelkmans L, Rottier PJ, Bosch BJ, de Haan CA.** 2014. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog **10:**e1004502.

13. **Millet JK, Whittaker GR.** 2014. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci U S A **111:**15214-15219.
14. **Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, Li F.** 2015. Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. J Virol **89:**9119-9123.
15. **Bosch BJ, Bartelink W, Rottier PJ.** 2008. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. J Virol **82:**8887-8890.
16. **Wicht O, Li W, Willems L, Meuleman TJ, Wubbolts RW, van Kuppeveld FJ, Rottier PJ, Bosch BJ.** 2014. Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture. J Virol **88:**7952-7961.
17. **Sabir JS, Lam TT, Ahmed MM, Li L, Shen Y, Abo-Aba SE, Qureshi MI, Abu-Zeid M, Zhang Y, Khiyami MA, Alharbi NS, Hajrah NH, Sabir MJ, Mutwakil MH, Kabli SA, Alsulaimany FA, Obaid AY, Zhou B, Smith DK, Holmes EC, Zhu H, Guan Y.** 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science **351:**81-84.
18. **Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276-278.
19. **Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, Peiris JS, Guan Y, Rambaut A.** 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature **459:**1122-1125.
20. **Ksiazek TG, Rota PA, Rollin PE.** 2011. A review of Nipah and Hendra viruses with an historical aside. Virus Res **162:**173-183.
21. **Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey FA, Veesler D.** 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature **531:**114-117.
22. **Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, DiMaio F, Bosch BJ, Veesler D.** 2016. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol **23:**899-905.
23. **Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, Lu G, Wu Y, Yan J, Shi Y, Zhang X, Gao GF.** 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat Commun **8:**15092.
24. **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 Res **27:**119-129.
25. **Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, Corbett KS, Graham BS, McLellan JS, Ward AB.** 2016. Pre-fusion structure of a human coronavirus spike protein. Nature **531:**118-121.
26. **Pallesen J, Wang N, Corbett KS, Wrapp D, Kirchdoerfer RN, Turner HL, Cottrell CA, Becker MM, Wang L, Shi W, Kong WP, Andres EL, Kettenbach AN, Denison MR, Chappell JD, Graham BS, Ward AB, McLellan JS.** 2017. Immunogenicity and structures

635 of a rationally designed prefusion MERS-CoV spike antigen. Proc Natl Acad Sci U S A
636 doi:10.1073/pnas.1707304114.
637 27. **Liu C, Tang J, Ma Y, Liang X, Yang Y, Peng G, Qi Q, Jiang S, Li J, Du L, Li F.** 2015. Receptor
638 usage and cell entry of porcine epidemic diarrhea coronavirus. J Virol **89:**6121-6125.
639 28. **Hu H, Jung K, Vlasova AN, Chepngeno J, Lu Z, Wang Q, Saif LJ.** 2015. Isolation and
640 characterization of porcine deltacoronavirus from pigs with diarrhea in the United
641 States. J Clin Microbiol **53:**1537-1548.
642 29. **Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M,**
643 **Zheng BJ, Chan KH, Yuen KY.** 2012. Discovery of seven novel Mammalian and avian
644 coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene
645 source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene
646 source of gammacoronavirus and deltacoronavirus. J Virol **86:**3995-4008.
647 30. **Walls A, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey F, Veesler D.** 2016.
648 Crucial steps in the structure determination of a coronavirus spike glycoprotein using
649 cryo-electron microscopy. Protein Sci doi:10.1002/pro.3048.
650 31. **Snijder J, Borst AJ, Dosey A, Walls AC, Burrell A, Reddy VS, Kollman JM, Veesler D.**
651 2017. Vitrification after multiple rounds of sample application and blotting improves
652 particle density on cryo-electron microscopy grids. J Struct Biol **198:**38-42.
653 32. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot.
654 Acta Crystallogr D Biol Crystallogr **66:**486-501.
655 33. **DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits TC, Cheng**
656 **Y, Baker D.** 2015. Atomic-accuracy models from 4.5-A cryo-electron microscopy data
657 with density-guided iterative local refinement. Nat Methods **12:**361-365.
658 34. **Frese CK, Zhou H, Taus T, Altelaar AF, Mechtler K, Heck AJ, Mohammed S.** 2013.
659 Unambiguous phosphosite localization using electron-transfer/higher-energy collision
660 dissociation (EThcD). J Proteome Res **12:**1520-1525.
661 35. **Stewart-Jones GB, Soto C, Lemmin T, Chuang GY, Druz A, Kong R, Thomas PV, Wagh K,**
662 **Zhou T, Behrens AJ, Bylund T, Choi CW, Davison JR, Georgiev IS, Joyce MG, Kwon YD,**
663 **Pancera M, Taft J, Yang Y, Zhang B, Shivatare SS, Shivatare VS, Lee CD, Wu CY, Bewley**
664 **CA, Burton DR, Koff WC, Connors M, Crispin M, Baxa U, Korber BT, Wong CH, Mascola**
665 **JR, Kwong PD.** 2016. Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A,
666 B, and G. Cell doi:10.1016/j.cell.2016.04.010.
667 36. **Gristick HB, von Boehmer L, West AP, Jr., Schamber M, Gazumyan A, Golijanin J,**
668 **Seaman MS, Fatkenheuer G, Klein F, Nussenzweig MC, Bjorkman PJ.** 2016. Natively
669 glycosylated HIV-1 Env structure reveals new mode for antibody recognition of the CD4-
670 binding site. Nat Struct Mol Biol **23:**906-915.
671 37. **Lin YP, Xiong X, Wharton SA, Martin SR, Coombs PJ, Vachieri SG, Christodoulou E,**
672 **Walker PA, Liu J, Skehel JJ, Gamblin SJ, Hay AJ, Daniels RS, McCauley JW.** 2012.
673 Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin.
674 Proc Natl Acad Sci U S A **109:**21474-21479.
675 38. **Zhou Y, Lu K, Pfefferle S, Bertram S, Glowacka I, Drosten C, Pohlmann S, Simmons G.**
676 2010. A single asparagine-linked glycosylation site of the severe acute respiratory
677 syndrome coronavirus spike glycoprotein facilitates inhibition by mannose-binding lectin
678 through multiple mechanisms. J Virol **84:**8753-8764.

39. **Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD, Jr., Thackray LB, Young MD, Mason RJ, Ambrosino DM, Wentworth DE, Demartini JC, Holmes KV.** 2004. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc Natl Acad Sci U S A **101:**15748-15753.

40. **Jeffers SA, Hemmila EM, Holmes KV.** 2006. Human coronavirus 229E can use CD209L (L-SIGN) to enter cells. Adv Exp Med Biol **581:**265-269.

41. **Krempl C, Ballesteros ML, Zimmer G, Enjuanes L, Klenk HD, Herrler G.** 2000. Characterization of the sialic acid binding activity of transmissible gastroenteritis coronavirus by analysis of haemagglutination-deficient mutants. J Gen Virol **81:**489-496.

42. **Sanchez CM, Gebauer F, Sune C, Mendez A, Dopazo J, Enjuanes L.** 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. Virology **190:**92-105.

43. **Huang X, Dong W, Milewska A, Golda A, Qi Y, Zhu QK, Marasco WA, Baric RS, Sims AC, Pyrc K, Li W, Sui J.** 2015. Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. J Virol **89:**7202-7213.

44. **Vlasak R, Luytjes W, Spaan W, Palese P.** 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. Proc Natl Acad Sci U S A **85:**4526-4529.

45. **Peng G, Xu L, Lin YL, Chen L, Pasquarella JR, Holmes KV, Li F.** 2012. Crystal structure of bovine coronavirus spike protein lectin domain. J Biol Chem **287:**41931-41938.

46. **Wu K, Li W, Peng G, Li F.** 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc Natl Acad Sci U S A **106:**19970-19974.

47. **Reguera J, Santiago C, Mudgal G, Ordono D, Enjuanes L, Casasnovas JM.** 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLoS Pathog **8:**e1002859.

48. **Ying T, Prabakaran P, Du L, Shi W, Feng Y, Wang Y, Wang L, Li W, Jiang S, Dimitrov DS, Zhou T.** 2015. Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. Nat Commun **6:**8223.

49. **Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, Ji X, Dimitrov DS.** 2006. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J Biol Chem **281:**15829-15836.

50. **Hwang WC, Lin Y, Santelli E, Sui J, Jaroszewski L, Stec B, Farzan M, Marasco WA, Liddington RC.** 2006. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. J Biol Chem **281:**34610-34616.

51. **Sui J, Li W, Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallarico AS, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA.** 2004. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc Natl Acad Sci U S A **101:**2536-2541.

52. **Zhu Z, Chakraborti S, He Y, Roberts A, Sheahan T, Xiao X, Hensley LE, Prabakaran P, Rockx B, Sidorov IA, Corti D, Vogel L, Feng Y, Kim JO, Wang LF, Baric R, Lanzavecchia A, Curtis KM, Nabel GJ, Subbarao K, Jiang S, Dimitrov DS.** 2007. Potent cross-reactive

neutralization of SARS coronavirus isolates by human monoclonal antibodies. Proc Natl Acad Sci U S A **104:**12123-12128.

53. **Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch B-J, Rey FA, Veesler D.** 2017. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proceedings of the National Academy of Sciences **114:**11157-11162.
54. **Tse LV, Hamilton AM, Friling T, Whittaker GR.** 2014. A novel activation mechanism of avian influenza virus H9N2 by furin. J Virol **88:**1673-1683.
55. **Kawaoka Y, Naeve CW, Webster RG.** 1984. Is virulence of H5N2 influenza viruses in chickens associated with loss of carbohydrate from the hemagglutinin? Virology **139:**303-316.
56. **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:**15205-15215.
57. **Matsuyama S, Ujike M, Morikawa S, Tashiro M, Taguchi F.** 2005. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. Proc Natl Acad Sci U S A **102:**12543-12547.
58. **Matsuyama S, Taguchi F.** 2009. Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. J Virol **83:**11133-11141.
59. **Eckert DM, Malashkevich VN, Kim PS.** 1998. Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. J Mol Biol **284:**859-865.
60. **Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS.** 2006. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. Nature **439:**38-44.
61. **Walls A, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey FA, Veesler D.** 2017. Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. Protein Sci **26:**113-121.
62. **Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ.** 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res **35:**e45.
63. **Mastronarde DN.** 2005. Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol **152:**36-51.
64. **Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA.** 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods **14:**331-332.
65. **Zhang K.** 2016. Gctf: Real-time CTF determination and correction. J Struct Biol **193:**1-12.
66. **Rohou A, Grigorieff N.** 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol **192:**216-221.
67. **Voss NR, Yoshioka CK, Radermacher M, Potter CS, Carragher B.** 2009. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J Struct Biol **166:**205-213.
68. **Kimanius D, Forsberg BO, Scheres SH, Lindahl E.** 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife **5**.
69. **Scheres SH.** 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol **180:**519-530.

70. **Scheres SH, Chen S.** 2012. Prevention of overfitting in cryo-EM structure determination. Nat Methods **9:**853-854.
71. **Rosenthal PB, Henderson R.** 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol **333:**721-745.
72. **Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SH, 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.
73. **Goddard TD, Huang CC, Ferrin TE.** 2007. Visualizing density maps with UCSF Chimera. J Struct Biol **157:**281-287.
74. **Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G.** 2015. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr D Biol Crystallogr **71:**136-153.
75. **DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I.** 2011. Modeling symmetric macromolecular structures in Rosetta3. PLoS One **6:**e20450.
76. **Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC.** 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr **66:**12-21.
77. **Agirre J, Iglesias-Fernandez J, Rovira C, Davies GJ, Wilson KS, Cowtan KD.** 2015. Privateer: software for the conformational validation of carbohydrate structures. Nat Struct Mol Biol **22:**833-834.
78. **Bern M, Kil YJ, Becker C.** 2012. Byonic: advanced peptide and protein identification software. Curr Protoc Bioinformatics **Chapter 13:**Unit13 20.

Figure Captions

Fig 1. CryoEM structure of the PDCoV S protein. A, A representative micrograph of vitreous ice-embedded PDCoV S protein at 3.4 µm defocus. Scale bar: 510 Å. **B,** Selected 2D class averages of the PDCoV S protein. Scale bar: 85 Å. **C-D,** Side (C) and top (D) views of the PDCoV S cryoEM map filtered at 3.5 Å resolution and sharpened with a B-factor of -150 $Å^2$. The density is colored per protomer. **E-F,** Ribbon representation of the PDCoV S trimer structure rendered with the same orientations as in panels C-D. One protomer is colored according to the indicated structural domains whereas the other two protomers are colored gray.

802

803 **Fig 2. Glycosylation profile of the PDCoV S protein. A-B,** Two orthogonal views of
804 the PDCoV S trimer rendered as ribbons. Glycan density extracted from the
805 unsharpened reconstruction is colored green for one protomer and grey for the other
806 two protomers. Labels indicate the position of N-linked glycosylated asparagine
807 residues. **C,** Schematic summary of all detected N-linked glycans. Each site shows the
808 most extensive glycan structure detected, either by mass-spectrometry or cryoEM. A full
809 overview of all detected N-linked glycans is provided in Supplementary Table 1. Glycan
810 moities are represented as symbols according to the key and the structural domains are
811 individually colored and indicated in a linear representation of the PDCoV S sequence.
812 **D-E**, Ribbon representation of PDCoV (D) and HCoV-NL63 (E) S protomers with
813 glycans visualized by cryoEM shown as green spheres.

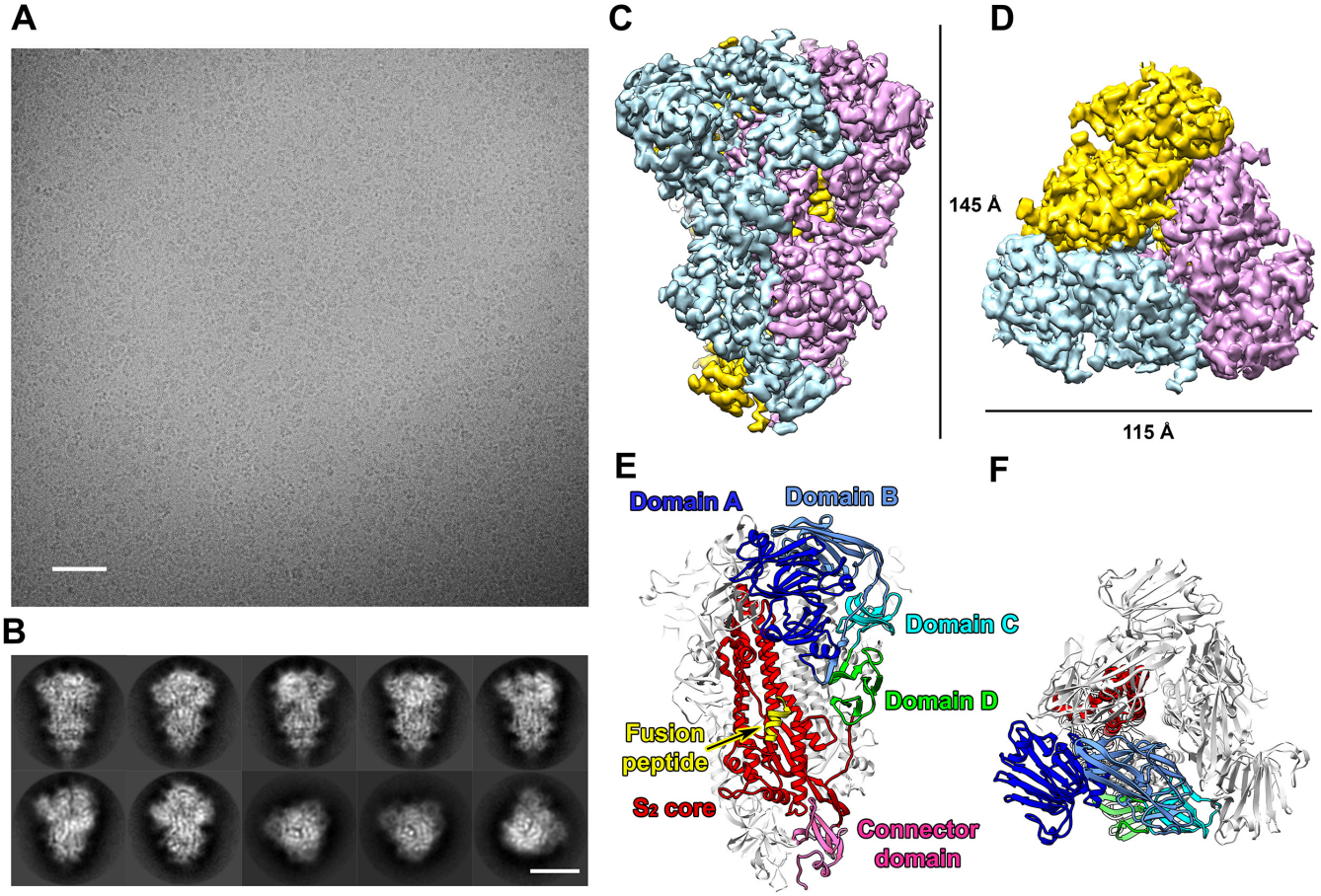
814

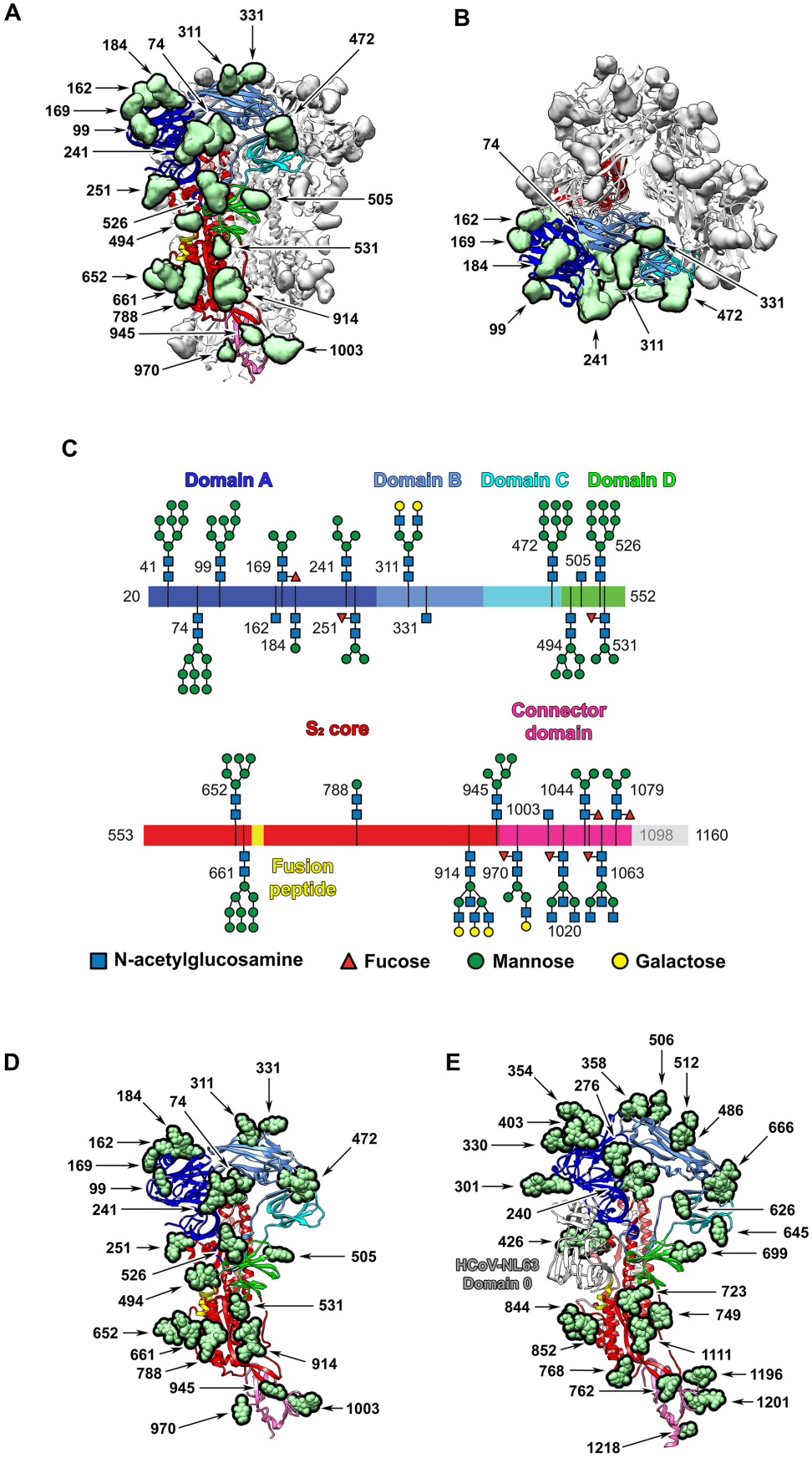
815

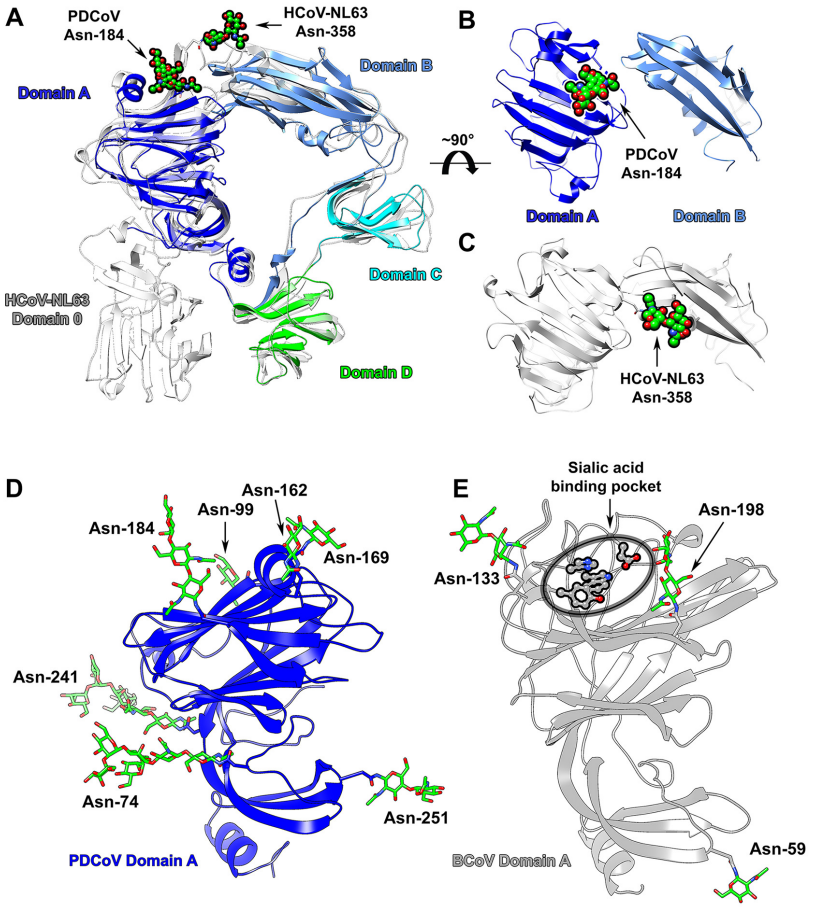
816 **Fig 3. Structural features of the PDCoV S_1 subunit and the galectin-like domain A.**
817 **A,** Superposition of the PDCoV and HCoV-NL63 S_1 subunits highlights the absence of
818 domain 0 in PDCoV S. **B,** View of the interface between PDCoV S A and B domains
819 showing the Asn-184 glycan points away from domain B. **C,** View of the interface
820 between HCoV-NL63 S A and B domains showing the Asn-358 glycan contributes to
821 masking the receptor-binding loops. **D,** Ribbon representation of PDCoV domain A. **E,**
822 Ribbon representation of BCoV domain A oriented identically to panel (D). Highly
823 conserved residues involved in sialic acid recognition are shown in ball and stick
824 representation. Glycans are rendered as spheres in panels A-C or sticks in panels D-E

825 and colored per atom type (carbon: green, nitrogen: blue and oxygen: red). **F,** The
826 PDCoV S_1 subunit C-terminally tagged with the Fc portion of human IgG (S_1-Fc) was
827 tested for its hemagglutination potential of an erythrocyte suspension of human or rat
828 origin, either alone or premixed with protein A-coupled nanoparticles to increase the
829 avidity of S_1-Fc proteins for sialic acids. The sialic-acid binding S_1 subunit of HCoV-
830 OC43 (GenBank: AAR01015.1) C-terminally fused to human Fc portion was used as a
831 positive control. 'Mock' indicates the condition where no S_1 subunit was used (negative
832 control). Wells positive for hemagglutination are encircled.
833
834
835 **Fig 4. Structural comparison of α- and δ-coronavirus receptor-binding domains.**
836 **A-D,** Ribbon rendering of the receptor-binding domain (domain B) of the δ-genus
837 PDCoV S (A) and α-genus PRCV S (B), HCoV-NL63 S (C) and TGEV S (D). Loops that
838 have been implicated in receptor-binding for α-coronaviruses are indicated. Key
839 aromatic residues that have been shown to take part in α-coronavirus receptor-binding
840 and putatively involved in δ-coronavirus receptor-binding are highlighted. Disulphide
841 bonds that stabilise receptor binding loops are indicated and glycans within the domain
842 are shown as sticks (carbon: green, nitrogen: blue and oxygen: red).
843
844
845 **Fig 5. Structural features of the PDCoV S_2 subunit. A,** Ribbon representation of the
846 PDCoV S trimer with the S_2 subunit core of one protomer colored from blue to red (from
847 N-terminus to C-terminus). **B,** Zoomed-in view of the S_2' activation loop region. Two

848 glycans, linked to Asn-669 and Asn-673, that are strictly conserved in HCoV-NL63 S are
849 shown as sticks (carbon: green, nitrogen: blue and oxygen: red). For comparison, the
850 equivalent residues in the HCoV-NL63 S protein are indicated in gray. **C,** The PDCoV S
851 glycoprotein features an insertion of 14 amino acid residues in HR1, compared to the β-
852 coronavirus MHV S protein, folding as an extended loop and an helical extension of two
853 turns. The residues accounting for this HR1 insertion interact with the complementary
854 insertion in HR2 in the post-fusion conformation (Fig S2 B).
855
856
857 **Fig 6. The PDCoV S glycoprotein is resistant to digestive enzymes.** Purified SARS
858 S (1 mg/ml) and PDCoV S (0.5 mg/ml) glycoproteins were incubated with 0.1 mg/ml
859 trypsin or chymotrypsin for 2 hours at 22 °C. The digestion reactions were analyzed on
860 a 12% SDS-PAGE gel. After incubation, the SARS S protein was extensively
861 proteolyzed whereas a large fraction of the PDCoV S protein remains intact.
862
863
864
865

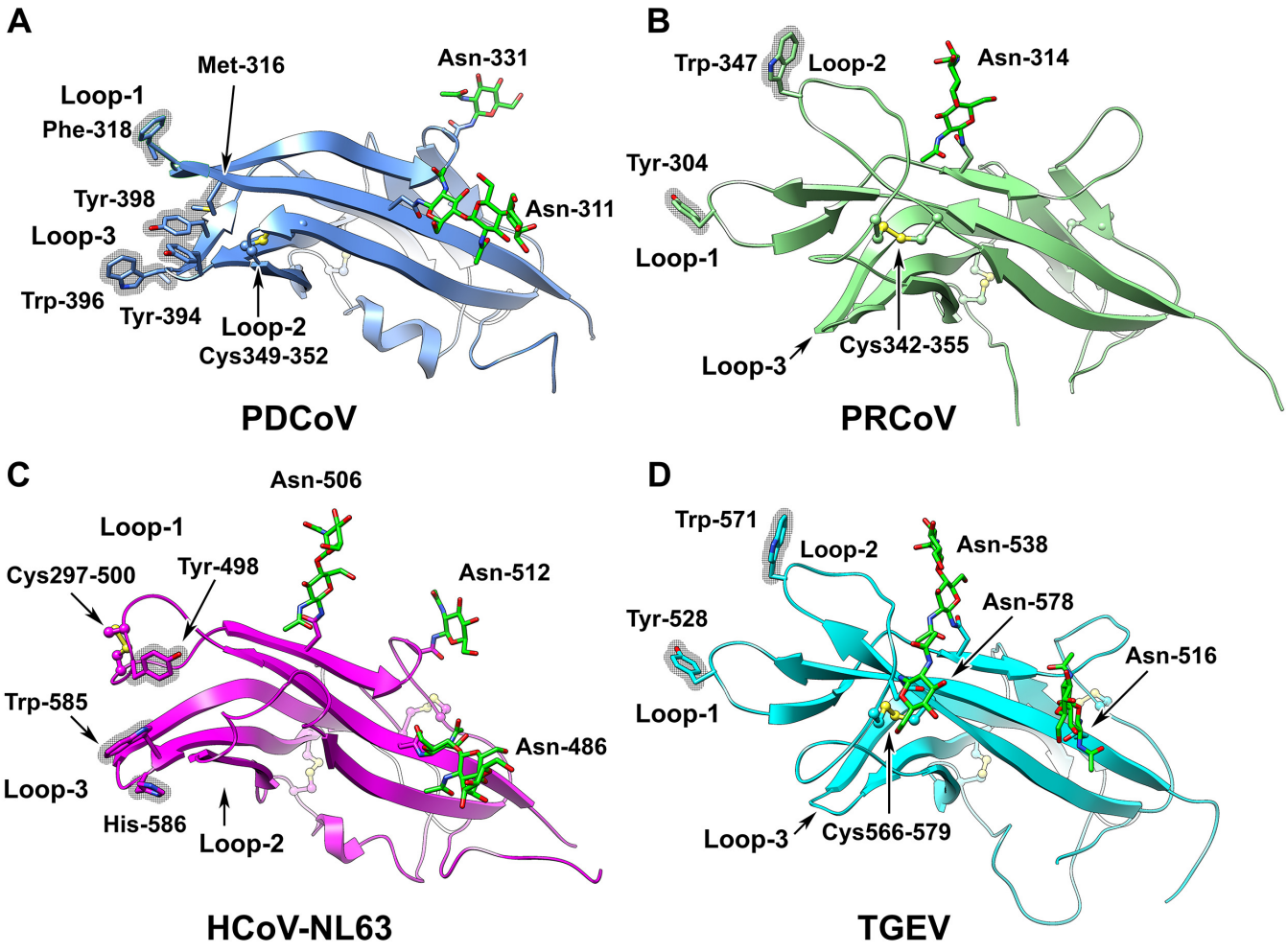






F

		1:2	1:4	1:8	1:16	1:32	1:64	1:128	1:256	Nano particle	HA titer
Human erythrocytes	PDCoV S_1-Fc									-	0
										+	0
	HCoV-OC43 S_1-Fc									-	0
										+	64
	Mock									-	0
										+	0
Rat erythrocytes	PDCoV S_1-Fc									-	0
										+	0
	HCoV-OC43 S_1-Fc									-	0
										+	256
	Mock									-	0
										+	0



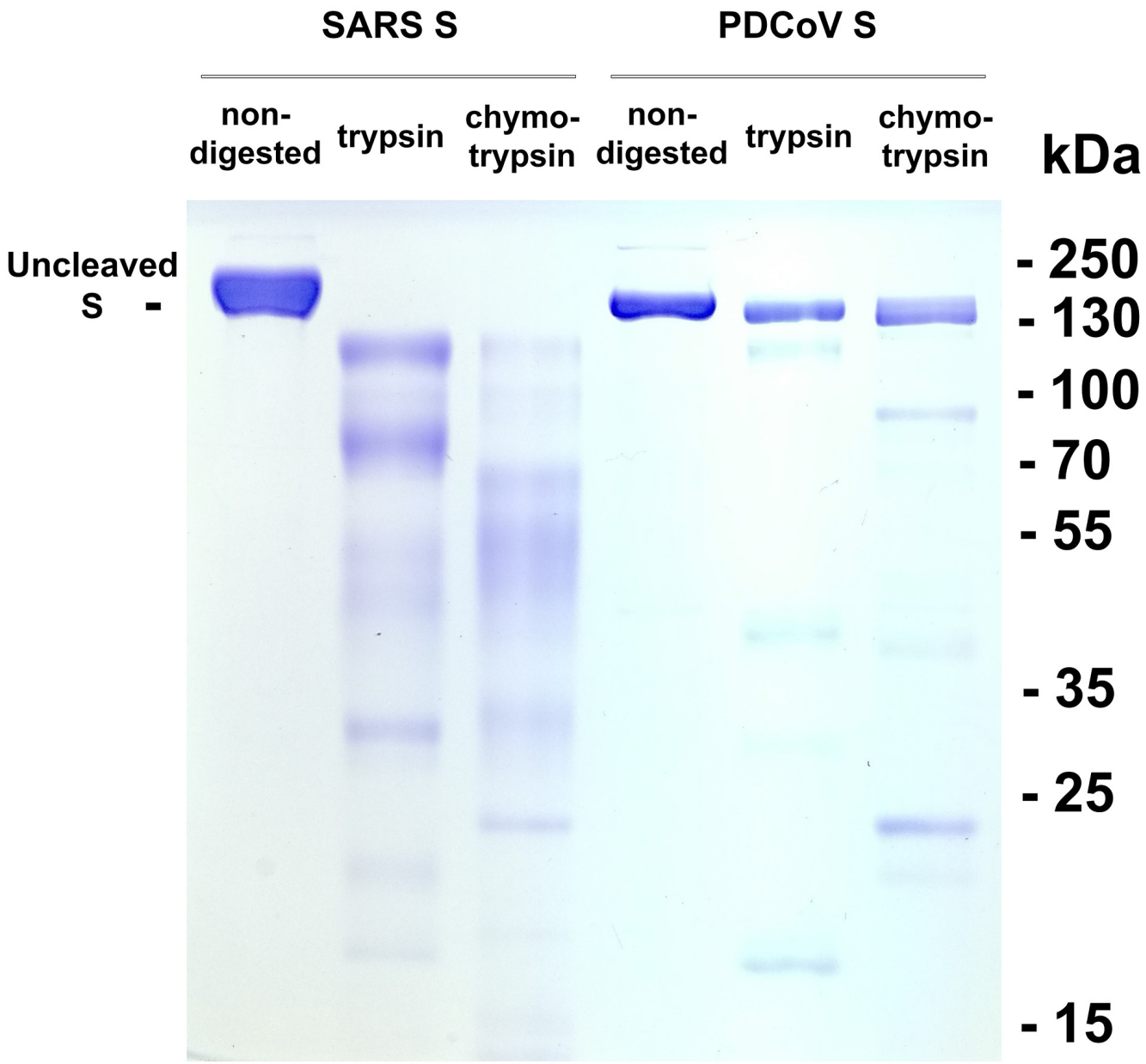


Table 1 Data collection and refinement statistics

Data collection	
Number of particles	455710
Pixel size (Å)	1.33
Voltage (kV)	300
Electron dose (e–/Å^2)	23.5
Refinement	
Resolution (Å)	3.5
Map-sharpening B factor (Å^2)	-150
Model validation	
Favored rotamers (%)	98
Poor rotamers (%)	0.35
Ramachandran allowed (%)	99.69
Ramachandran outliers (%)	0.31
Clash score	2.2
MolProbity score	1.27