Cryo-EM structure of porcine delta coronavirus spike protein in the pre-fusion state

Jian Shang 1,*, Yuan Zheng 1,*, Yang Yang 1,*, Chang Liu 1, Qibin Geng 1,
Wanbo Tai 2,3, Lanying Du 2, Yusen Zhou 3, Wei Zhang 4,5,#, Fang Li 1,#

1 Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA
2 Lindsley F. Kimball Research Institute, New York Blood Center, New York, NY 10065, USA
3 State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China
4 Department of Diagnostic and Biological Sciences, School of Dentistry, University of Minnesota, Minneapolis, MN 55455, USA
5 Characterization Facility, College of Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA

* These authors contributed equally to this work.
# Correspondence:
Fang Li (lifang@umn.edu); Wei Zhang (zhangwei@umn.edu)

Key words: viral spike glycoprotein, pre-fusion, receptor binding, membrane fusion, immune evasion, cryo-electron microscopy, single-particle reconstruction

Running title: Structure, function, and evolution of PdCoV spike
Abstract

Coronavirus spike proteins from different genera are divergent, although they all mediate coronavirus entry into cells by binding to host receptors and fusing viral and cell membranes. Here we determined the cryo-EM structure of porcine delta coronavirus (PdCoV) spike protein at 3.3-angstrom resolution. The trimeric protein contains three receptor-binding S1 subunits that tightly pack into a crown-like structure and three membrane-fusion S2 subunits that form a stalk. Each S1 subunit contains two domains, N-terminal domain (S1-NTD) and C-terminal domain (S1-CTD). PdCoV S1-NTD has the same structural fold as alpha- and beta-coronavirus S1-NTDs as well as host galectins, and it recognizes sugar as its potential receptor. PdCoV S1-CTD has the same structural fold as alpha-coronavirus S1-CTDs, but its structure differs from that of beta-coronavirus S1-CTDs. PdCoV S1-CTD binds to an unidentified receptor on host cell surfaces. PdCoV S2 is locked in the pre-fusion conformation by structural restraint of S1 from a different monomeric subunit. PdCoV spike possesses several structural features that may facilitate immune evasion by the virus, such as its compact structure, concealed receptor-binding sites, and shielded critical epitopes. Overall, this study reveals that delta-coronavirus spikes are structurally and evolutionally more closely related to alpha-coronavirus spikes than to beta-coronavirus spikes; it also has implications for the receptor recognition, membrane fusion, and immune evasion by delta-coronaviruses as well as coronaviruses in general.
Significance

In this study we determined the cryo-EM structure of porcine delta coronavirus (PdCoV) spike protein at 3.3 angstrom. This is the first atomic structure of a spike protein from the delta coronavirus genus, which is divergent in amino acid sequences from the well-studied alpha- and beta-coronavirus spike proteins. In the current study, we described the overall structure of the PdCoV spike and the detailed structure of each of its structural elements. Moreover, we analyzed the functions of each of the structural elements. Based on the structures and functions of these structural elements, we discussed the evolution of PdCoV spike protein in relation to the spike proteins from other coronavirus genera. This study combines the structure, function, and evolution of coronavirus spike proteins, and provides many insights into the receptor recognition, membrane fusion, immune evasion, and evolution of PdCoV spike protein.
Introduction

Coronaviruses are large enveloped RNA viruses that can be classified into four genera: α, β, γ, and δ (1). Both α- and β-coronaviruses infect mammals, γ-coronaviruses infect birds, and δ-coronaviruses infect mammals and birds (1). Representative coronaviruses include: human NL63 coronavirus (HCoV-NL63) and porcine transmissible gastroenteritis coronavirus (TGEV) from α genus; mouse hepatitis coronavirus (MHV), bovine coronavirus (BCoV), SARS coronavirus (SARS-CoV) and MERS coronavirus (MERS-CoV) from β genus; avian infectious bronchitis virus (IBV) from γ genus; porcine delta coronavirus (PdCoV) from δ genus (2). Coronaviruses from different genera demonstrate distinct serotypes, mainly due to the divergence of their envelope-anchored spike proteins (3). The spike proteins mediate viral entry into host cells by first binding to host receptors through their S1 subunit and then fusing host and viral membranes through their S2 subunit (4). Hence they are critical determinants of viral host range and tissue tropism, and also induce most of the host immune responses (5). Knowing the structure and function of the spike proteins from different genera is critical for understanding cell entry, pathogenesis, evolution, and immunogenicity of coronaviruses (6).

The receptor recognition pattern by coronaviruses is complicated (7). The S1 subunits from α- and β-coronavirus spikes contain two domains, the N-terminal domain (S1-NTD) and C-terminal domain (S1-CTD). Depending on the virus, either one or both of the S1 domains can function as the receptor-binding domain (RBD) by binding to host receptors. On the one hand, S1-CTDs from α- and β-coronaviruses have different tertiary
structures, but they share a common structural topology, indicating a common evolutionary origin and subsequent divergent evolution of S1-CTDs (7). α-coronavirus S1-CTDs recognize either angiotensin-converting enzyme 2 (ACE2) or aminopeptidase-N (APN) as their protein receptor, whereas β-coronavirus S1-CTDs recognize either ACE2 or dipeptidyl peptidase 4 (DPP4) (8-16). Hence S1-CTDs likely have undergone further divergent evolution to recognize different receptors. On the other hand, S1-NTDs from α- and β-coronaviruses both have the same structural fold as human galectins, and they recognize either sugar receptors or a protein receptor CEACAM1 (17-23). Hence it has been suggested that coronavirus S1-NTDs originated from host galectins and have undergone divergent evolution to recognize different receptors (7). These studies on receptor recognition by coronaviruses have revealed complex evolutionary relationships among the spikes from different genera.

The membrane fusion mechanism for coronavirus spikes is believed to be similar to those used by “class 1” viral membrane-fusion proteins (24, 25). The best studied such protein is hemagglutinin (HA) from influenza virus (26, 27). Influenza HA exists in two structurally distinct conformations. Its “pre-fusion” conformation on mature virions is a trimer, already cleaved by host proteases into receptor-binding subunit HA1 and membrane fusion subunit HA2 that remain associated. During the membrane fusion process, HA1 dissociates and HA2 undergoes a dramatic conformational change to reach its “post-fusion” conformation: two heptad repeat (HR) regions from each HA2 subunit, HR-N and HR-C, refold into a six-helix bundle, and a previously buried hydrophobic fusion peptide (FP) becomes exposed and inserts into host membrane. The cryo-EM structures of α- and β-coronavirus spikes in the pre-fusion conformation have recently
been determined (28-31). The overall architecture of α- and β-coronavirus spikes is
similar to, albeit more complex than, that of influenza HA. Biochemical studies have
identified parts of S2 that form six-helix bundle structures and hence likely correspond to
HR-N and HR-C respectively (32-34), and another part of S2 that associates with
membranes and hence likely corresponds to FP (35, 36). It was demonstrated that α-
coronavirus spikes are heavily glycosylated, with S2 more heavily glycosylated than S1,
as a viral strategy for immune evasion (29). These studies on membrane fusion by α- and
β-coronavirus spikes have suggested a common molecular mechanism for membrane
fusion shared by coronavirus spikes and other class 1 viral membrane fusion proteins (37,
38).

PdCoV from the δ genus is a highly lethal viral pathogen in piglets (39-41).
Compared to the extensive studies on α- and β-coronavirus spikes, much less is known
about the structure and function of δ-coronavirus spikes. It is not clear which of their S1
domains functions as the RBD, where the structural elements of S2 are located, how δ-
coronavirus spikes are structurally and evolutionarily related to the spikes from other
genera, or what strategies δ-coronavirus spikes use to evade host immune surveillance.
This study fills in these critical gaps by determining the cryo-EM structure of PdCoV
spike and revealing its functions in receptor binding, viral entry and immune evasion.

Results and Discussion

Overall structure of PdCoV spike
To capture PdCoV spike in the pre-fusion conformation, we constructed and prepared PdCoV spike ectodomain (S-e) without the transmembrane anchor or intracellular tail (Fig. 1A). We also excluded a short pre-transmembrane region (PTR) because this region is hydrophobic and can adversely affect protein solubility (42). Instead, we replaced these regions with a GCN4 trimerization tag followed by His$_6$ tag. We expressed PdCoV S-e in insect cells, and purified it to homogeneity. We collected cryo-EM data on PdCoV S-e, and determined its structure at 3.3Å resolution (Table 1; Fig. 1B, Fig. 2).

The atomic structure of pre-fusion PdCoV S-e contains residues from 52 to 1017, covering all of the key structural elements except HR-C (Fig. 1A). The overall trimeric structure of PdCoV spike is similar to, but more compact than, those of $\alpha$- and $\beta$-coronavirus spikes: PdCoV spike has a length of 130Å from S1 to S2 and a width of 50Å at S2 (Fig. 1C). S2 itself spans 100Å in length (Fig. 1D). Three S1 subunits form a crown-like structure and sit on top of the trimeric S2 stalk (Fig. 1C, 1D). Three S1-CTDs are located at the top and center of the spike trimer, whereas three S1-NTDs are located on the lower and outer side of S1-CTDs (Fig. 3A, 3B, 3C, 3D). The S1-CTD mainly stacks with the S1-NTD from the same monomeric subunit, although there also exist inter-subunit interactions between S1-CTDs from different subunits and between S1-CTD and S1-NTD from different subunits. In contrast, the S1 trimer of $\beta$-genus MHV spike has an intertwined quaternary structure, with S1-CTD from one subunit mainly stacking with S1-NTD from another subunit (Fig. 4A) (30). Like PdCoV spike, the S1-CTD in $\alpha$-genus HCoV-NL63 spike also mainly stacks with the S1-NTD from the same subunit (Fig. 4B) (29). Moreover, whereas each subunit of PdCoV S1 contains only one...
S1-NTD, each subunit of HCoV-NL63 S1 contains two, possibly resulting from gene
duplication (Fig. 4B) (29). Connecting S1 and S2 are two subdomains, SD1 and SD2, and
a long loop (Fig. 3A, 3B). The structure of PdCoV S2 is in the pre-fusion conformation
and can be aligned well with those of α- and β-coronavirus S2 fragments (Fig. 4A, 4B).
HR-C is missing in both the current PdCoV S2 structure and previously published α- and
β-coronavirus S2 structures, suggesting that this region is poorly ordered. Our structural
model also includes glycans N-linked to 39 residues on the trimer (13 on each monomeric
subunit). In this article, we will illustrate the structures and functions of each of the
structural elements in PdCoV spike.

Structure, function, and evolution of PdCoV S1-NTD

PdCoV S1-NTD adopts a β-sandwich fold identical to human galectins (Fig. 5A). Its core structure consists of two anti-parallel β-sheet layers: one is seven-stranded and
the other is six-stranded. On top of the core structure is a short α-helix. Underneath the
core structure is another three-stranded β-sheet and another α-helix. The S1-NTDs from
α- and β-coronaviruses have the same galectin fold (Fig. 5B, 5C). Like PdCoV S1-NTD,
α-coronavirus S1-NTDs contain a short α-helix on top of the core structure, but β-
coronavirus S1-NTDs contain a ceiling-like structure in the same location. The galectin
fold of PdCoV S1-NTD suggests that like some of the α- and β-coronavirus S1-NTDs,
PdCoV S1-NTD may recognize sugar as host receptors to facilitate initial viral
attachment to cells, and hence it may function as a viral lectin.
We investigated the sugar-binding capability of PdCoV S1-NTD. To this end, we expressed and purified recombinant PdCoV S1-NTD containing a C-terminal His$_6$ tag, and carried out an ELISA assay to examine whether it binds sugar (Fig. 5D). More specifically, PdCoV S1-NTD was incubated with mucin, which contains a variety of sugar chains on its surface; subsequently, the mucin-bound PdCoV S1-NTD was detected using antibodies recognizing its His$_6$ tag. The result showed that PdCoV S1-NTD bound to mucin. Thus, PdCoV S1-NTD bound to the sugar moiety of mucin and can potentially recognize sugar as its receptor. The sugar-binding site in PdCoV S1-NTD is currently unknown. Because the sugar-binding site in β-genus BCoV S1-NTD and the galactose-binding site in human galectins are both located on top of the core structure (18, 43), the sugar-binding site in PdCoV S1-NTD may also be located in the same region (Fig. 5A, 5C).

The above structural and functional analyses of PdCoV S1-NTD provide insight into the evolution of coronavirus S1-NTDs from different genera. Previously, based on the structures and functions of β-coronavirus S1-NTDs, we hypothesized that ancestral coronaviruses acquired a galectin gene from the host and incorporated it into their spike gene, which began to encode S1-NTD; we further predicted that the S1-NTDs from other genera also contain the galectin fold. Both the structure of PdCoV S1-NTD presented here and the structures of α-coronavirus S1-NTDs determined by recent studies confirmed our earlier prediction and lent further support to our previous hypothesis. Hence, coronavirus S1-NTDs from different genera likely all have the same evolutionary origin, which might be the host galectin, and have conserved the galectin fold through evolution.
Structure, function, and evolution of PdCoV S1-CTD

PdCoV S1-CTD adopts a β-sandwich fold also containing two β-sheet layers: one is a three-stranded anti-parallel β-sheet and the other is a three-stranded mixed β-sheet (Fig. 6A). Its structure is similar to the β-sandwich core structure of α-coronavirus S1-CTDs, but different from the core structure of β-coronavirus S1-CTDs that contains a single β-sheet layer (Fig. 6B, 6C). We previously showed that despite their different structural folds, α- and β-coronavirus S1-CTDs share the same structural topology (i.e., connectivity of secondary structural elements) (7). Similarly, PdCoV S1-CTD also shares the same structural topology with β-coronavirus S1-CTDs. Because α- and β-coronaviruses widely use their S1-CTD as the main RBD by recognizing protein receptors, PdCoV S1-CTD may also recognize a protein receptor and function as the main RBD.

We examined the possibility of PdCoV S1-CTD recognizing a receptor on the surface of mammalian cells. To this end, we expressed and purified recombinant PdCoV S1-CTD containing a C-terminal Fc tag, and performed a flow cytometry assay to detect the binding of PdCoV S1-CTD-Fc to mammalian cells (Fig. 6D). Here the cell-bound PdCoV S1-CTD was detected using antibodies recognizing its Fc tag. The result showed that PdCoV S1-CTD-Fc bound to both human and pig cells with significantly higher affinity than Fc alone, suggesting that PdCoV S1-CTD binds to a receptor on the surface of both human and pig cells. Although PdCoV S1-CTD demonstrates higher affinity for human cells than for pig cells, it is unknown whether PdCoV infects human cells since receptor recognition is only one of several factors that can impact coronavirus infections.
We further investigated whether PdCoV S1-CTD recognizes ACE2 or APN, two known protein receptors for α-coronavirus S1-CTDs. To this end, we prepared and purified recombinant PdCoV S1-CTD containing a C-terminal His<sub>6</sub> tag, and carried out a dot-blot assay to examine whether it binds ACE2 or APN (Fig. 6E). The result showed that PdCoV S1-CTD does not bind ACE2 or APN. As positive controls, TGEV S1-CTD binds APN, whereas SARS-CoV S1-CTD binds ACE2. Taken together, these results demonstrate that PdCoV S1-CTD likely functions as the main RBD and binds a yet-to-be-identified receptor on the surface of human and pig cells.

The receptor-binding site in PdCoV S1-CTD is currently unknown. In α-coronavirus S1-CTDs, the three loops on the top of the β-sandwich core function as receptor-binding motifs (RBMs) by binding to their respective protein receptor, ACE2 for HCoV-NL63 and APN for TGEV. In PdCoV S1-CTD, the same three loops are structurally similar to their counterparts in α-coronavirus S1-CTDs. Hence, these three loops in PdCoV S1-CTD may bind to a protein receptor and function as RBMs. In the current structure, the S1-CTD is in a closed conformation, with its putative RBMs pointing towards the S1-NTD and unavailable for receptor binding. To bind its receptor, the S1-CTD would need to switch to an open conformation by “standing up” on the spike trimer and rendering the putative RBMs available for receptor binding.

Based on the above structural and functional analyses, we discuss the evolution of coronavirus S1-CTDs. Because S1-CTD is located on the tip of the pre-fusion spike trimer, it is the most exposed region on the surface of virions and thereby is under heavy immune pressure to evolve. Possibly as a consequence of immune pressure, S1-CTD is
structurally divergent among different coronavirus genera: α- and δ-coronavirus S1-CTDs have a β-sandwich core, whereas β-coronavirus S1-CTDs have a β-sheet core. The RBMs are located on the very tip of S1-CTDs, and are even more structurally divergent than the core structure of S1-CTDs. The RBMs in α- and δ-coronavirus S1-CTDs are three short discontinuous loops; depending on the virus, their RBM loops can bind APN (as in TGEV), ACE2 (as in HCoV-NL63), or a yet-to-be-identified receptor (as in PdCoV). The RBM in β-coronavirus S1-CTDs is a long continuous subdomain; depending on the virus, their RBM can bind ACE2 (as in SARS-CoV) or DPP4 (as in MERS-CoV). Despite their structural divergence, the S1-CTDs from different genera share the same structural topology in their cores (7). These results suggest that these S1-CTDs have a common evolutionary origin and have undergone divergent evolution.

Moreover, our study demonstrates that PdCoV S1-CTD is structurally and evolutionarily more closely related to α-coronavirus S1-CTDs than to β-coronavirus S1-CTDs.

Structures, functions, and evolution of S1 subdomains

The structures of SD1 and SD2 are similar to their counterparts in α- and β-coronavirus spikes (Fig. 3B). SD1 adopts a small β-sandwich fold containing two antiparallel β-sheets: one is two-stranded and the other is five-stranded. SD2 also adopts a small β-sandwich fold containing two three-stranded β-sheets: one is antiparallel and the other is mixed. Interestingly, both SD1 and SD2 consist of discontinuous regions: majority of their sequences are to the C-terminus of S1-CTD, but they also each contain a region to the N-terminus of S1-CTD. Based on these structural data, SD1 and SD2 might have evolved later than S1-NTD and S1-CTD. The main function of the two S1
subdomains is to connect S1 and S2, but SD1 also plays a role in membrane fusion as discussed below.

Structure, function, and evolution of S2

The overall structure of the pre-fusion trimeric PdCoV S2 is similar to those of α- and β-coronaviruses. Two central helices, CH-N and CH-C, from each subunit form a six-helix inter-subunit interface. Based on previous biochemical and structural studies using isolated regions in S2, HR-N corresponds to a region consisting of four helices and connecting loops, and HR-C corresponds to a disordered region (Fig. 7A, 7B) (30). The exact location of FP is uncertain, but it may correspond to a region consisting of two helices and a connecting loop (30). Examination of the pre-fusion and post-fusion structures of influenza HA2 suggests that during the conformational changes of PdCoV S2, HR-N from each subunit in the pre-fusion conformation would need to fold into one long central helix as part of the six-helix bundle of the post-fusion structure (Fig. 7C).

Hence, like influenza HA2, part of the CH-C in PdCoV S2 should also be part of the HR-N, such that the other parts of HR-N can anchor upon CH-C and extend towards the membrane-distal direction (Fig. 7A). Like the FP in influenza HA2, the FP in PdCoV S2 would also need to change its conformation, spring out towards the membrane-distal direction, and insert into the target membrane. The reason why HR-N and FP are locked in their pre-fusion conformation is likely because S1-CTD and SD1 from another subunit sit on top of them respectively, and prevent them respectively from extending towards the membrane-distal direction. The stacking between S1 and S2 from two different subunits contributes to the compact structure of PdCoV spike trimer. Two protease cleavages, one...
at the S1/S2 boundary and the other on the N-terminus of FP, can potentially remove the structural restraint of S1 on S2, allowing the conformational changes of S2 to occur (30, 37, 44). Both the structural and mechanistic similarities between coronavirus S2 and influenza HA2 suggest that the two viral membrane-fusion proteins are evolutionarily related (4). The above analysis will need to be confirmed by the atomic structure of post-fusion PdCoV S2.

**Immune evasion strategies by PdCoV spike**

The structure of PdCoV spike suggests immune evasion strategies by PdCoV spike. First, the PdCoV spike has a compact structure. The six domains and six subdomains of trimeric S1 are tightly packed (Fig. 3B, 3C), which reduces the surface area of the spike protein. Despite its compact structure, S1 maintains the two-RBD system, giving the virus more options in receptor selections than a single-RBD system would do. Second, in the current structure, PdCoV S1-CTD is in a closed conformation with its putative RBM loops facing S1-NTD and inaccessible to the host receptor (Fig. 3D). Upon infecting host cells, S1-CTD would need to switch to an open conformation to render the putative RBM loops accessible to the host receptor. The closed-to-open conformational change of S1-CTD has been observed for β-genus MERS-CoV and SARS-CoV spikes (28). This mechanism can minimize the exposure of the putative RBM loops to the immune system. Third, our structural model of PdCoV spike contains glycans N-linked to 39 residues (13 on each subunit); there are also another 24 predicted, but not observed, N-linked glycosylation sites (8 on each subunit) (Fig. 8A, 8B). Most of these sites are located on the surface of S1, which is in contrast to α-genus HCoV-NL63
spike where S2 is more heavily glycosylated than S1. Thus, while it was previously suggested that HCoV-NL63 spike evades host immune surveillance mainly by glycan shielding its S2 epitopes (29), PdCoV spike appears to evade host immune surveillance mainly by glycan shielding its S1 epitopes. For example, the putative sugar-binding site in PdCoV S1-NTD is surrounded by glycans, which reduces the accessibility of this site to the immune system (Fig. 8C). As a comparison, the sugar-binding site in β-genus BCoV S1-NTD is also shielded, not by glycans, but by the ceiling-like structure on top of the core structure (18). Taken together, PdCoV spike has several structural features that may facilitate viral immune evasion, such as reducing surface areas, concealing receptor-binding sites, and shielding critical S1 epitopes.

Conclusions

In this study we determined the cryo-EM structure of PdCoV spike at 3.3 Å. To our knowledge, this is the first atomic structure of a spike protein from the δ coronavirus genus, which is divergent in amino acid sequences from the well-studied α- and β-coronavirus spikes. Our study reveals a compact PdCoV spike trimer locked in the pre-fusion conformation. The trimeric S1 contains six domains (three copies of S1-NTD and S1-CTD each) and six subdomains (three copies of SD1 and SD2 each) that tightly pack into a crown-like structure. PdCoV S1-NTD has the same galectin fold as α- and β-coronavirus S1-NTDs; it binds sugar and can potentially recognize sugar as its receptors. These results expand our knowledge on the structures and functions of S1-NTDs from different coronavirus genera, and provide further evidence on the common host origin of coronavirus S1-NTDs. PdCoV S1-CTD has the same β-sandwich fold as α-coronavirus
S1-CTDs, and this structural fold differs from the \( \beta \)-sheet fold of \( \beta \)-coronavirus S1-CTDs. However, S1-CTDs from all coronavirus genera share the same structural topology, suggesting a common evolutionary origin of coronavirus S1-CTDs. PdCoV S1-CTD binds to an unidentified receptor on mammalian cell surfaces, and may function as the main RBD. Moreover, PdCoV S1-CTD is in a closed conformation with its putative receptor-binding sites buried; it would need to switch to an open conformation for receptor binding. The structures of both S1-NTD and S1-CTD of PdCoV are more similar to those of \( \alpha \)-coronaviruses than to those of \( \beta \)-coronaviruses, and hence PdCoV spike is evolutionarily more closely related to \( \alpha \)-coronavirus spikes than to \( \beta \)-coronavirus spikes.

The trimeric PdCoV S2 forms the stalk of the spike protein. Each of the S2 subunits is locked in the pre-fusion conformation by structural constraint of S1 from a different monomeric subunit. More specifically, HR-N and FP are prevented from re-folding into their post-fusion conformation by the steric restrictions from S1-CTD and SD1, respectively, of another subunit. PdCoV spike possesses several structural features that appear to facilitate its evasion from host immune surveillance, such as its compact structure, the closed conformation of its S1-CTD, and heavy glycosylation near critical epitopes in S1. Overall, our study combines the structure and function of PdCoV spike, and provides many insights into the receptor recognition, membrane fusion, immune evasion, and evolution of PdCoV spike as well as coronavirus spikes in general.
Acknowledgements

This work was supported by NIH grant R01AI089728 (to F. Li) and fund from University Minnesota (to W. Zhang). Initial cryo-EM images were collected using FEI Tecnai TEMs maintained at the Characterization Facility of the University of Minnesota. Final cryo-EM data were collected at the EM facility of Purdue University. We thank Valorie Bowman, Thomas Klose, Yue Liu and Steve Wilson for help in data collection and image processing, and Hinh Ly for comments on the manuscript. Initial image analysis and computation work was performed using the workstations at the Basic Sciences Computing Laboratory of the University of Minnesota Supercomputing Institute. The cryoEM map has been deposited in the Electron Microscopy Data Bank (EMD) under accession codes 7063. The atomic model has been deposited in the Protein Data Bank (PDB) under accession codes 6B7N.
References


Materials and Methods

Expression, purification, and treatment of PdCoV spike ectodomain

PdCoV spike ectodomain (S-e) (residues 18-1077) was cloned into pFastBac vector (Life Technologies Inc.) with a N-terminal honeybee melittin signal peptide and C-terminal GCN4 and His$_6$ tags. It was expressed in sf9 insect cells using the Bac-to-Bac system (Life Technologies Inc.) and purified as previously described (15). Briefly, the protein was harvested from cell culture medium, and purified sequentially on Ni-NTA column and Superdex200 gel filtration column (GE Healthcare). Because we showed earlier that low pH could facilitate trimer formation (45), we incubated PdCoV S-e in buffer containing 0.1 M sodium citrate (pH 5.6) at room temperature for 1 hour, and then re-purified it on Superdex200 gel filtration column in buffer containing 20 mM Tris pH7.2 and 200 mM NaCl.

Cryo-electron microscopy

For sample preparation, aliquots of PdCoV S-e (3 μl, 0.35 mg/ml, in buffer containing 2 mM Tris pH7.2 and 20 mM NaCl) were applied to glow-discharged CF-2/1-4C C-flat grids (Protochips). The grids were then plunge-frozen in liquid ethane using a FEI MarkIII Vitrobot system (FEI Company).

For data collection, images were recorded using a Gatan K2 Summit direct electron detector in the direct electron counting mode (Gatan), attached to a Titan-Krios TEM (FEI Company), at Purdue University. The automated software Leginon (46) was used to collect ~2,100 movies at 22,500x magnification and at a defocus range of between 0.5 and 3 μm. Each movie had a total accumulated exposure of 52 e/Å$^2$. 

23
fractionated in 55 frames of 200 ms exposure. Data collection statistics are summarized in Table 1.

For data processing, the recorded movies were corrected for beam-induced motion using MotionCor2 (47). The final image was bin-averaged to give the pixel size to be 1.3 Å. The parameters of the microscope contrast transfer function were estimated for each micrograph using GCTF (48). Particles were automatically picked and extracted using RELION 2.0 on a GPU workstation with a box size of 256 pixels. Initially, particles were subjected to 2D alignment and clustering using RELION 2.0, and the best classes were selected for an additional 2D alignment. Some of the particles on 2D class averages appear to have a tail (Fig. S1A), which may correspond to HR-C. Nevertheless, the weak density of the tail region suggests that this region is poorly ordered, and hence this region was not included in subsequent map calculation and model building. All of the particles, with or without the tail, were subjected to 3D auto-refine with a mask covering the overall shape of the particles (excluding the tail region) to yield the map. The orientations of the particles used in the final reconstruction map sufficiently covered the whole sphere in the Fourier space to allow calculation of a 3D map with isotropic resolution. The map was sharpened with modulation transfer function of K2 operated at 300 kV using RELION 2.0 post processing. Reported resolution was based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion, and Fourier shell correction curves were corrected for the effects of soft masking by high-resolution noise substitution (49). Data processing statistics are summarized in Table 1.

Model building and refinement
For atomic model building, the cryo-EM structure of HCoV-NL63 spike (PDB: 5865SZS) were divided into 7 parts (S1-NTD, SD2', SD1', S1-CTD, SD1'', SD2'' and S2), and fitted into the cryo-EM map of PdCoV S-e individually using UCSF Chimera (50) and Coot (51). Model rebuilding was performed manually in Coot based on the well-defined continuous density of the main chain, and sequence register assignment was guided mainly by the density of N-linked glycans and of bulky amino acid residues. The structural model was refined using Phenix (52) with geometry restraints and three-fold noncrystallographic symmetry constraints. Refinement and manual model correction in Coot were carried out iteratively until there was no more improvement in geometry parameters. The quality of the final model was analyzed with MolProbity (53) and EMRinger (54). The validation statistics of the structural model are summarized in Table 1.

ELISA sugar-binding assay

PdCoV S1-NTD containing a C-terminal His$_6$ tag was expressed and purified in the same way as PdCoV S-e, and assayed for its sugar-binding capability using an ELISA assay as previously described (18). Briefly, ELISA plates were pre-coated with bovine mucin (1 mg/ml) at 37 °C for 1 hour. After blocking with 1% BSA at 37 °C for 1 hour, PdCoV S1-NTD (1 μg/ml) was added to the plates and incubated with mucin at 37 °C for 1 hour. After washes with PBS buffer, the plates were incubated with anti-His$_6$ antibody (Santa Cruz) at 37 °C for 1 hour. Then the plates were washed with PBS and incubated with HRP-conjugated goat anti-mouse IgG antibody (1:5,000) at 37 °C for 1 hour. After more washes with PBS, enzymatic reaction was carried out using ELISA substrate (Life Technologies Inc.) and stopped with 1 M H$_2$SO$_4$. Absorbance at 450 nm (A450) was...
measured using Tecan Infinite M1000 PRO Microplate Reader (Tecan Group Ltd.). Five replicates were done for each sample. Porcine epidemic diarrhea virus (PEDV) S1 and SARS-CoV S1-CTD were prepared as previously described (15, 55), and PdCoV S1-CTD was prepared as described below; these three proteins were used in the assay as controls.

*Dot-blot receptor-binding assay*

PdCoV S1-CTD containing a C-terminal His<sub>6</sub> tag was expressed and purified in the same way as PdCoV S-e, and assayed for its receptor-binding capability using a dot-blot receptor-binding assay as previously described (55). Briefly, 5 µM receptor (human ACE2 or porcine APN) was dotted onto nitrocellulose membranes. The membranes were dried and blocked with 1% BSA, and then incubated with 1 µM PdCoV S1-CTD at 4 °C for 2 hours. After washes with PBS buffer, the membranes were incubated with anti-His<sub>6</sub> antibody (Life Technologies Inc.) at 4 °C for 2 hours, washed with PBS, incubated with HRP-conjugated goat anti-mouse IgG antibody (1:5,000) at 4 °C for 2 hours, and washed with PBS. Finally, the receptor-bound proteins were detected using a chemiluminescence reagent (ECL plus, GE Healthcare). Recombinant human ACE2 and porcine APN were prepared as previously described (13, 15).

*Flow cytometry cell-binding assay*

PdCoV S1-CTD containing a C-terminal Fc tag was expressed, purified, and assayed for its cell-binding capability by flow cytometry as previously described (56). Briefly, human (HeLa and A549) and pig (ST and PK15) cells were incubated with PdCoV S1-CTD-Fc (40 µg/ml), or human IgG-Fc control, at room temperature for 30 min, followed by incubation with fluorescein isothiocyanate (FITC)-labeled anti-human
IgG-Fc antibody for 30 min. The cells were then analyzed for the binding using flow cytometry.
Table 1 Data and model statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Microscope</td>
<td>Titan Krios</td>
</tr>
<tr>
<td>Voltage (keV)</td>
<td>300</td>
</tr>
<tr>
<td>Defocus range (µm)</td>
<td>1.0 to 4.0</td>
</tr>
<tr>
<td>Movies</td>
<td>2168</td>
</tr>
<tr>
<td>Frames per movie</td>
<td>55</td>
</tr>
<tr>
<td>Dose rate (e⁻/Å²/s)</td>
<td>4.7</td>
</tr>
<tr>
<td>Total dose per movie (e⁻/Å²)</td>
<td>51.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data processing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Particles</td>
<td>87,002</td>
</tr>
<tr>
<td>Symmetry</td>
<td>C3</td>
</tr>
<tr>
<td>Provided B-factor (Å²)</td>
<td>-150</td>
</tr>
<tr>
<td>Map resolution (Å)</td>
<td>3.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Validation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSF Chimera CC(57)</td>
<td>0.865</td>
</tr>
<tr>
<td>EMRinger Score(54)</td>
<td>2.77</td>
</tr>
<tr>
<td>MolProbity Score(53)</td>
<td>1.91</td>
</tr>
<tr>
<td>All-atom clashscore(53)</td>
<td>5.48</td>
</tr>
<tr>
<td>Rotamers outliers (%)</td>
<td>0.78</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>99.59</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R.m.s deviations</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond length (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.437</td>
</tr>
</tbody>
</table>
Figure Legends:

Figure 1. Overall structure of PdCoV S-e in the pre-fusion conformation. (A) Schematic drawing of PdCoV S-e (spike ectodomain). S1: receptor-binding subunit. S2: membrane-fusion subunit. GCN4-His<sub>6</sub>: GCN4 trimerization tag followed by His<sub>6</sub> tag. S1-NTD: N-terminal domain of S1. S1-CTD: C-terminal domain of S1. CH-N and CH-C: central helices N and C. FP: fusion peptide. HR-N and HR-C: heptad repeats N and C. Residues in shaded regions (N-terminus, GCN4 tag, and His<sub>6</sub> tag) were not traced in the structure. (B) Cryo-EM maps of PdCoV S-e with atomic model fitted in. The maps have a contour of 6.6 σ. (C) Cryo-EM structure of pre-fusion PdCoV S-e. Each of the monomeric subunits is colored differently. (D) Structure of a monomeric subunit in the pre-fusion conformation. The structural elements are colored in the same way as in panel (A).

Figure 2. Cryo-EM data analysis of PdCoV S-e. (A) Representative micrographs of frozen-hydrated PdCoV S-e particles and representative 2D class averages in different orientations. Arrow indicates a poorly ordered tail region in some of the particles. (B) Gold-standard Fourier shell correlation (FSC) curves. The resolution was determined to be 3.3 Å. The 0.143 and 0.5 cut-off values are indicated by horizontal grey bars. (C) Final cryo-EM map of PdCoV S-e colored according to the local resolution.

Figure 3. Structure of PdCoV S1. (A) Schematic drawing of PdCoV S1. SD1: subdomain 1. SD2: subdomain 2. SD1 consists of two discontinuous regions SD1’ and SD1”. SD2 consists of two discontinuous regions SD2’ and SD”. (B) Structure of
monomeric S1. Domains and subdomains are colored in the same way as in panel (A).

Residue ranges for each of the domains and subdomains are labeled. (C) Structure of 
trimeric S1, viewed from the side. Each of the monomeric subunits is colored differently.
The empty space under S1 is occupied by S2, which is not shown here. (D) Structure of 
trimeric S1, viewed from the top. Each of the monomeric subunits is colored differently.

Figure 4. Structural alignments of PdCoV spike with the spikes from other 
coronavirus genera. (A) Alignment of PdCoV and β-genus MHV spikes. PdCoV spike 
is colored in magenta. MHV spike (PDB ID: 3JCL) is colored in cyan. (B) Alignment of 
PdCoV and α-genus HCoV-NL63 spikes. PdCoV spike is colored in magenta. HCoV-
NL63 spike (PDB ID: 5SZS) is colored in green. Each subunit of PdCoV S1 contains 
only one S1-NTD, whereas each subunit of HCoV-NL63 S1 contains two.

Figure 5. Structure and function of PdCoV S1-NTD. (A) Structure of PdCoV S1-
NTD. The putative sugar-binding site is indicated by the question mark. (B) Structure of 
α-genus HCoV-NL63 S1-NTD (PDB ID: 5SZS). (C) Structure of β-genus BCoV S1-
NTD (PDB ID: 4H14). (D) ELISA sugar-binding assay for PdCoV S1-NTD. Here the 
ELISA plates were pre-coated with sugar-rich mucin, and then PdCoV S1-NTD was 
added and incubated with mucin. Mucin-bound S1-NTD was detected using antibodies 
recognizing its C-terminal His6 tag. Porcine epidemic diarrhea virus (PEDV) S1 was used 
as the positive control; PdCoV S1-CTD, SARS-CoV S1-CTD, and BSA were used as 
negative controls. Plate without mucin was used as an additional negative control.
Statistic analyses were performed using two-tailed t-test. Error bars indicate S.E.M. (n=5). *** P<0.001.

**Figure 6. Structure and function of PdCoV S1-CTD.** (A) Structure of PdCoV S1-CTD. The putative RBM loops are indicated by the question mark. (B) Structure of α-genus HCoV-NL63 S1-CTD (PDB ID: 3KBH). (C) Structure of β-genus SARS-CoV S1-CTD (PDB ID: 2AJF). (D) Flow cytometry assay for the binding of PdCoV S1-CTD to the surface of mammalian cells. Cell-bound PdCoV S1-CTD was detected using antibodies recognizing its C-terminal Fc tag. Fc or cells only were used as negative controls. Statistic analyses were performed using two-tailed t-test. Error bars indicate S.E.M. (n=4). *** P<0.001. (E) Dot-blot receptor-binding assay for PdCoV S1-CTD. Here the receptor (either APN or ACE2) was first dotted onto a membrane. Subsequently, PdCoV S1-CTD was dotted and incubated with the receptor. Receptor-bound S1-CTD was detected using antibodies recognizing its C-terminal His<sub>6</sub> tag. TGEV and SARS-CoV S1-CTDs were used as positive controls. PBS buffer was used as a negative control.

**Figure 7. Structure and function of PdCoV S2.** (A) Structure of the pre-fusion monomeric PdCoV S2 only including CH-C, HR-N and FP. Arrow indicates the direction in which HR-N would need to extend to reach the post-fusion conformation. Question mark indicates part of CH-C that likely is part of HR-N. Residue ranges for each of the structural elements are labeled. (B) S1-CTD and SD1 from a different subunit stack with HR-N and FP, respectively, preventing them from switching to their post-fusion conformation. Scissor indicates the proteolysis sites to the N-terminus of FP. (C)
Structures of influenza HA2 in the pre-fusion and post-fusion conformations (PDB IDs: 709 2YPG and 1QU1). Arrow indicates the direction in which HR-N would need to extend to reach the post-fusion conformation. Scissor indicates the proteolysis sites to the N-terminus of FP.

**Figure 8. Glycosylation sites on the surface of PdCoV spike.** (A) Distribution of N-linked glycosylation sites on the one-dimensional structure of PdCoV spike. Ψ indicates N-linked glycosylate site. Those on the top indicate glycans observed in the structure. Those at the bottom indicate predicted, but not observed, glycosylate sites. Predicted glycosylation sites in the N-terminal region and HR-C were not included because these two regions were not traced in the structure. (B) Distribution of N-linked glycosylation sites on the three-dimensional structure of PdCoV spike. Observed glycans are in dark blue. Predicted, but not observed, glycosylation sites are in light blue. (C) Distribution of N-linked glycosylation sites in monomeric S1. Question marks indicate the putative sugar-binding site in S1-NTD and putative RBMs in S1-CTD, respectively.
A block diagram with amino acid positions labeled.

B: Structural models with a contour of 6.6σ.

C: A pre-fusion trimer model with dimensions 130 Å and 100 Å.

D: Monomeric subunit structure with labeled domains CH-N, FP, CH-C, HR-N, S1-CTD, and S1-NTD.
A

\[ SD1 = SD1' + SD1'' \]
\[ SD2 = SD2' + SD2'' \]

B

S1 monomer (side view)

C

S1 trimer (side view)

D

S1 trimer (top view)