

Structure and binding properties of Pangolin-CoV Spike glycoprotein inform the evolution of SARS-CoV-2.

Antoni G. Wrobel (✉ antoni.wrobel@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0002-6680-5587>

Donald J. Benton (✉ donald.benton@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0001-6748-9339>

Pengqi Xu

Precision Medicine Center, The 10 Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen

Annabel Borg

The Francis Crick Institute

Chloë Rouston

The Francis Crick Institute

Stephen R. Martin

The Francis Crick Institute

Peter B. Rosenthal

The Francis Crick Institute <https://orcid.org/0000-0002-0387-2862>

John J. Skehel

The Francis Crick Institute <https://orcid.org/0000-0002-8519-9404>

Steven J. Gamblin (✉ steve.gamblin@crick.ac.uk)

The Francis Crick Institute <https://orcid.org/0000-0001-5331-639X>

Research Article

Keywords: SARS-CoV-2, electron cryomicroscopy, pangolin

Posted Date: September 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-83072/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on February 5th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-21006-9>.

1 **Structure and binding properties of Pangolin-CoV Spike** 2 **glycoprotein inform the evolution of SARS-CoV-2**

3
4 Antoni G. Wrobel^{#*1}, Donald J. Benton^{#*1}, Pengqi Xu^{4,1}, Annabel Borg², Chloë
5 Roustan², Stephen R. Martin¹, Peter B. Rosenthal³, John J. Skehel¹, and Steven J.
6 Gamblin^{*1}

7 ¹Structural Biology of Disease Processes Laboratory; ²Structural Biology Science
8 Technology Platform; ³Structural Biology of Cells and Viruses Laboratory, Francis
9 Crick Institute, NW1 1AT London, United Kingdom. ⁴Precision Medicine Center, The
10 Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen, 518107, Guangdong,
11 China

12 # equal contribution

13 * to whom correspondence should be addressed: antoni.wrobel@crick.ac.uk,
14 donald.benton@crick.ac.uk, steve.gamblin@crick.ac.uk

15
16 **Coronaviruses of bats and pangolins are implicated in the origin and evolution of**
17 **the pandemic SARS-CoV-2. We show that Spikes from Pangolin-CoVs, closely**
18 **related to SARS-CoV-2, bind strongly to human and pangolin ACE2 receptors. We**
19 **also report Cryo-EM structure of Pangolin-CoV S and show it adopts a fully-closed**
20 **conformation and that, aside from the Receptor-Binding Domain, it resembles the**
21 **spike of a bat coronavirus RaTG13 more than that of SARS-CoV-2.**

22
23 Despite intensive research into the origins of the COVID-19 pandemic, the
24 evolutionary history of its causative agent SARS-CoV-2 remains unclear^{1,2}. SARS-
25 CoV-2 belongs to the subgenus of sarbecoviruses, for which horseshoe bats
26 (*Rhinolophus sp.*) are the reservoir species^{1,3,4}. However, others have suggested⁵, and
27 we recently demonstrated⁶, that the bat coronavirus RaTG13, the closest known relative
28 of SARS-CoV-2, is unlikely to be able to infect human cells because of the very low
29 affinity of its Spike protein for the human receptor ACE2. For this reason, it has been
30 speculated that SARS-CoV-2 could have reached the human population via an
31 intermediate host⁵. A number of recent studies reported the existence of sarbecoviruses
32 highly similar to SARS-CoV-2 in diseased Malayan pangolins (*Manis javanica*) and

33 thus pangolins were proposed to have played a role in the emergence of the current
34 pandemic⁷⁻¹⁰.

35

36 **The affinity of Pangolin-CoV S proteins for ACE2 receptors**

37 To characterise the pangolin virus Spike and compare it with that of SARS-CoV-
38 2, we expressed and purified two different Pangolin-CoV Spike ectodomains. These are
39 based on the sequences of viruses isolated from pangolins seized in China's Guangdong
40 province in 2019^{8,9}. We also produced recombinant ectodomains of ACE2 proteins from
41 human, bat (*Rhinolophus ferreus*) and pangolin in order to perform comparative
42 biolayer interferometry assays. Both pangolin proteins (referred to as Pangolin-CoV S
43 and Pangolin-CoV S') showed strong (<100 nM) binding towards the human ACE2,
44 approximately ten-fold weaker binding to pangolin ACE2, and very weak binding to bat
45 ACE2 (**Fig. 1A** and **Fig. S1**). A similar pattern of binding was observed for SARS-CoV-
46 2 S (**Fig. 1B**); preferred and strong binding to human ACE2, weaker binding to pangolin
47 ACE2 and very weak binding to bat ACE2.

48 These binding data demonstrate functional similarity between the RBDs of the
49 S proteins from Pangolin-CoV and SARS-CoV-2, that correlates with the high degree
50 of sequence identity, and three-dimensional structure similarity, between them (**Fig 1D**).
51 None of the three species of ACE2 were bound strongly by the bat virus RaTG13 S (data
52 not shown). This observation correlates with the substantial sequence differences
53 between the RBD of RaTG13 and the RBDs of the other two species (**Fig. 1D**).

54

55 **Cryo-EM structure of Pangolin-CoV S**

56 We have determined the structure of the Pangolin-CoV S protein at 2.9 Å by
57 Cryo-EM (**Fig. 2 & Fig. S2**). The structure is of similar resolution to our recent structure
58 of SARS-CoV-2 S⁶, enabling a detailed comparison between the two. Overall, the
59 structure of the Pangolin-CoV S (**Fig. 2A**) is similar to the closed form of the SARS-
60 CoV-2 and the RaTG13 S; the most striking feature is that all of the resolvable particles
61 on the grid are in the closed conformation (compared with 83% in the uncleaved SARS-
62 CoV-2 S sample and 34% in the furin-cleaved in our previous study⁶). Comparison of
63 the structures of S of Pangolin-CoV and SARS-CoV-2 identifies two specific, major
64 differences that likely account for this feature.

65 Firstly, an amino-acid substitution in the otherwise highly conserved sequences
66 in the interface between RBD neighbours in the S trimer, likely contributes to a more

67 stable packing arrangement that favours the closed conformation (**Fig. 2B**). In detail,
68 there is a salt bridge between in the closed form of SARS-CoV-2 formed by Lys417 and
69 Glu406 in the RBD. In the Pangolin-CoV, an arginine is substituted at position 417 and,
70 while it also makes a salt bridge with Glu406, the unique side-chain properties of the
71 arginine residue induce different conformers at Arg403 and Tyr505 that enable
72 additional stacking interactions and the formation of a hydrogen bond to the mainchain
73 of Tyr369 in the neighbouring RBD. These interactions would be expected to contribute
74 additional stabilisation to the RBD/RBD packing, hence favouring the closed form.
75 Furthermore, in Pangolin-CoV S there are also two additional glycans close to the RBD
76 interface (**Fig. S4**).

77 Secondly, the presence of a leucine residue at position 50 in the NTD-associated
78 intermediate subdomain of Pangolin-CoV, compared with a serine residue in SARS-
79 CoV-2, promotes a conformational arrangement that is further indicative of the closed
80 form of S. Occupancy of a bulky, hydrophobic leucine (instead of the smaller, polar Ser)
81 leads the helix (residues 294 to 304) to shift 1.5 Å (to the right as viewed in **Fig. 2C**)
82 compared with SARS-CoV-2, stabilising the formation of a helix-turn-helix structure
83 between the two intermediate domains, which is not present in the SARS-CoV-2 but is
84 present in RaTG13 S (**Fig. 2C** and **Fig S3A**). Folding of this motif has the effect of
85 shifting the neighbouring RBD-associated subdomain as a rigid-body (to the left as
86 viewed in **Fig. 2D**). A similar arrangement, of the helix-turn-helix, and rigid-body
87 position of the domain are seen in the closed conformation of RaTG13 S (**Fig. S3B**).
88 Moreover, analysis of the open conformations of S found in SARS-CoV-2 S shows that
89 the RBD-associated intermediate domain shifts in the opposite direction upon S opening
90 (**Fig. 2D**).

91 Taken together, these observations suggest several sequence-based differences,
92 compared with SARS-CoV-2, that likely account for the Pangolin-CoV spike adopting
93 an all-closed conformation. In an earlier work, we described the closed conformation
94 adopted by the bat CoV RaTG13 S protein⁶. In that case, chemical crosslinking was
95 required to stabilise the protein for Cryo-EM analysis, and so the possibility existed that
96 the cross-linking had influenced its structure. The fact that the current closed
97 conformation of pangolin-CoV S is remarkably similar, outside of the RBD, to the
98 RaTG13 S suggests that the structure of the latter was probably not materially affected
99 by the crosslinking.

100 The likely role of the closed conformation for shielding the fusiogenic apparatus
101 of S2 has been detailed in several places¹¹⁻¹³, as with the need for the open conformation
102 to facilitate receptor binding^{14,15}. The similarity in affinity of intact SARS-CoV and
103 SARS-CoV-2 Spike trimers and their isolated RBDs for ACE2 receptor^{16,17} implies that
104 there is not a large energetic cost to opening of the S1 structure. Thus, the shielding
105 effect of the closed conformation contributes to the stability of the trimer. When the
106 virus encounters an ACE2-receptor-decorated cell surface, then the binding of ACE2
107 stabilises the open form of the Spike by inhibiting its re-closure. Thus, the presence of
108 ACE2 receptors likely stimulates the opening of S and its priming for subsequent
109 membrane fusion.

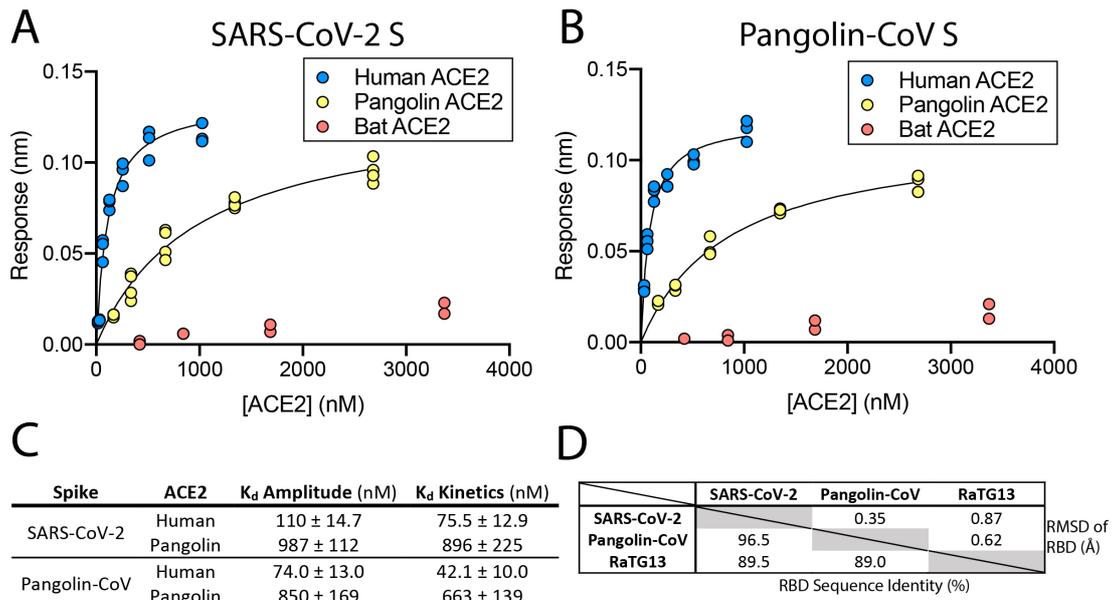
110

111 **Conclusion**

112 While the non-RBD component of the S protein is very similar between SARS-
113 CoV-2 and the bat virus RaTG13 (96% identity within S1), the sequence (97% identity)
114 and structure (RMSD 0.35 Å, **Fig. 1D**) of the RBD of SARS-CoV-2, in contrast, is
115 remarkably similar to that of Pangolin-CoV, particularly at the ACE2 binding site
116 (sequence identity between SARS-CoV-2 and RaTG13 is just 76% in this region). This
117 close similarity of RBDs between Pangolin-CoV and SARS-CoV-2 correlates with the
118 near identical binding properties of their two S proteins (**Fig. 1**). This suggests that, even
119 though Pangolin-CoV and SARS-CoV-2 have significant sequence differences beyond
120 their RBDs, especially in the NTD which for the Pangolin-CoV Spikes resembles more
121 that of bat viruses ZXC21 and ZC45 than RaTG13, pangolin viruses might well be
122 capable of infecting humans. In contrast, given the immeasurably low affinity of bat
123 RaTG13 S for human ACE2, it seems unlikely that at least this class of presumed
124 precursor bat viruses would infect humans.

125 There are conflicting reports on whether the RBD of Pangolin-CoV S, while very
126 similar in sequence to the RBD of the current pandemic virus, is the immediate precursor
127 to the SARS-CoV-2 RBD^{18,19}. Our results suggest that the effective zoonotic range for
128 this class of coronaviruses, beyond bats, may include species that, like pangolins, have
129 ACE2 receptors similar to the human ACE2. Consequently, there are likely to be other,
130 as yet unidentified, viruses that harbour RBDs of similar sequence and binding
131 properties to SARS-CoV-2 and Pangolin-CoV. The existence of such RBDs in the
132 relevant zoonotic background might account for the emergence of SARS-CoV-2
133 possibly via a recombination of bat viruses similar to RaTG13 with viruses perhaps not

134 dissimilar to Pangolin-CoV. It is also important to note that various species of bat, even
135 within *Rhinolophus* genus, show considerable differences in their ACE2 sequences and
136 that it has not been possible to demonstrate direct binding of spike proteins from the
137 viruses most closely related to SARS-CoV-2 to bat ACE2. Thus, perhaps S of bat viruses
138 bind a different, as yet unidentified, cellular receptor(s).
139



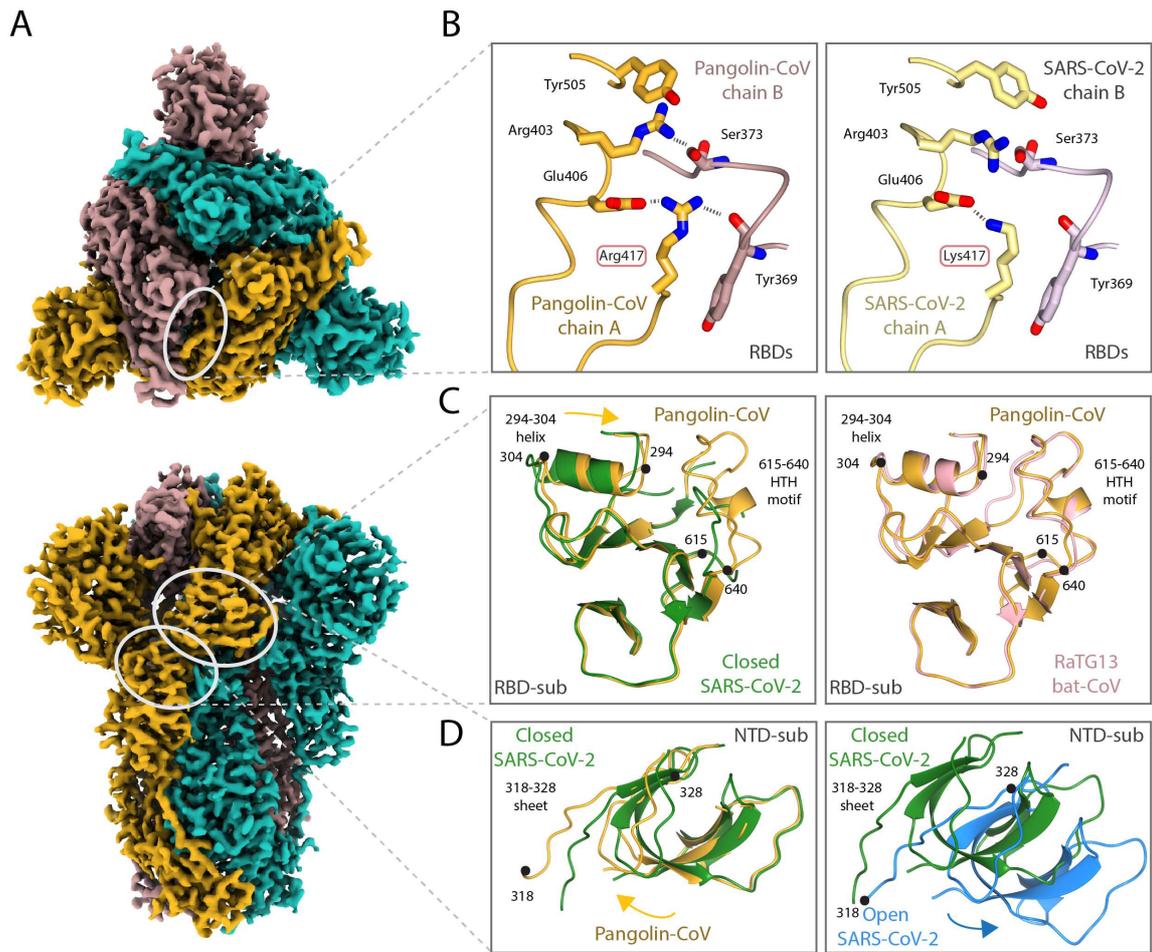
140

141

142 **Figure 1: Binding of Pangolin-CoV and SARS-CoV-2 S to ACE2s from different**
 143 **species. (A,B)** Plots of biolayer interferometry amplitudes for human (blue), pangolin
 144 (yellow) and bat (red) ACE2s binding to SARS-CoV-2 S (A) and pangolin-CoV S (B).
 145 (C) Equilibrium dissociation constants determined from the analysis of the data in A
 146 and B (K_d Amplitude) compared with values determined from analysis of the
 147 corresponding kinetic data (K_d Kinetics) (see Fig. S1). (D) Right hand side: RMSD of
 148 atom positions in the structures of RaTG13 S (6ZGF)⁶, closed conformation of SARS-
 149 CoV-2 S (6ZGE)⁶, and Pangolin-CoV S RBD determined in this study. Left hand-side:
 150 sequence identity of the RBDs from the same viruses.

151

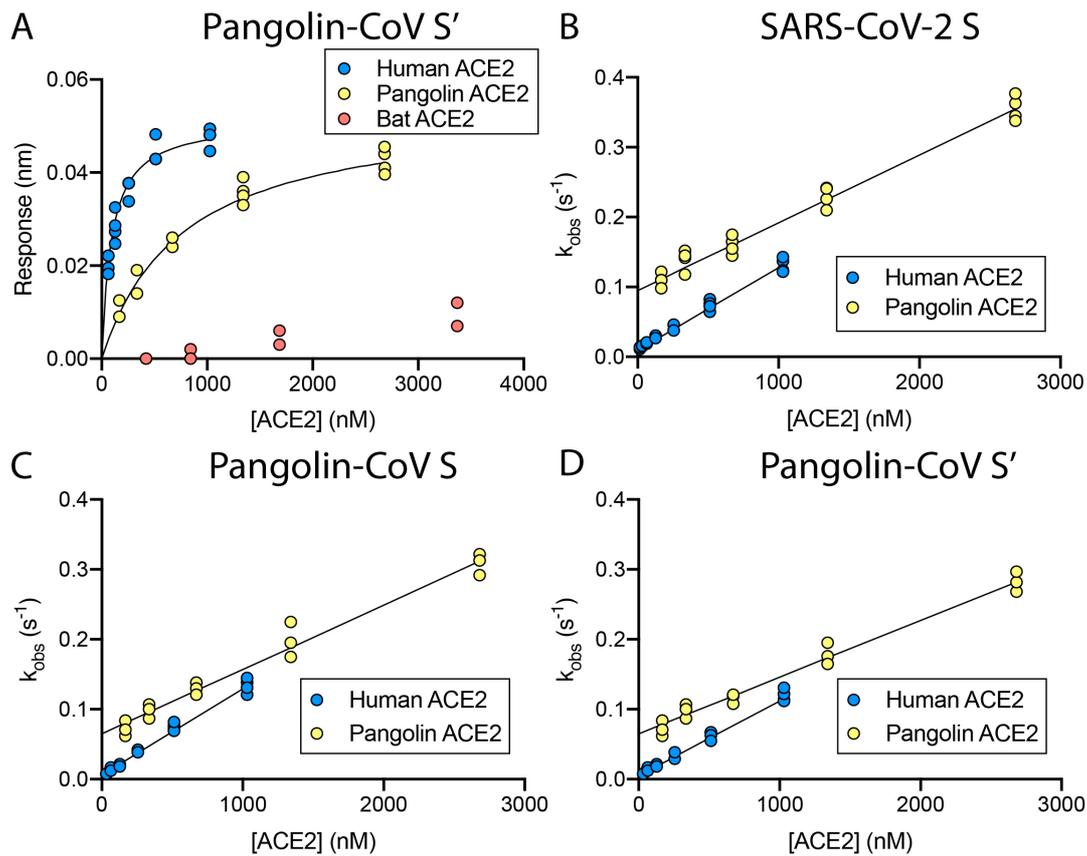
152



153

154

155 **Figure 2: Structure of Pangolin-CoV Spike protein.** (A) EM density representation
 156 from the 2.9 Å map of Pangolin-CoV S viewed from down the three-fold axis (top panel)
 157 and in the orthogonal view (lower panel). The subunits are coloured in blue, golden and
 158 brown. The white ovals identify the areas shown in molecular representation on the
 159 right. (B) Comparison of the RBD/RBD interface from the pangolin-CoV (left) and
 160 SARS-CoV-2 S highlighting the Arg417Lys substitution. (C) Comparison of the RBD-
 161 associated subdomains of the pangolin-CoV (golden) and closed form of SARS-CoV-2
 162 (green) in the left hand panel, showing the different positioning of the 294-304 helix
 163 and the presence of the 615-640 helix-turn-helix in the pangolin structure and, in the
 164 right-hand panel, the overlap of the same Pangolin-CoV S structure (golden) with the
 165 corresponding region from the RaTG13 (pink). (D) Comparison of (left) the NTD-
 166 associated subdomain of Pangolin-CoV (golden) with that of the closed form of SARS-
 167 CoV-2 (green) showing the different domain orientations between them; (right) the
 168 closed (green) and open (blue) conformations of the NTD-associated subdomain of
 169 SARS-CoV-2 showing that the shift in orientation of the NTD-associated subdomain on
 170 Spike opening is in the opposite direction to the shift seen between the Pangolin-CoV
 171 and closed SARS-CoV-2 conformations shown in the left panel.



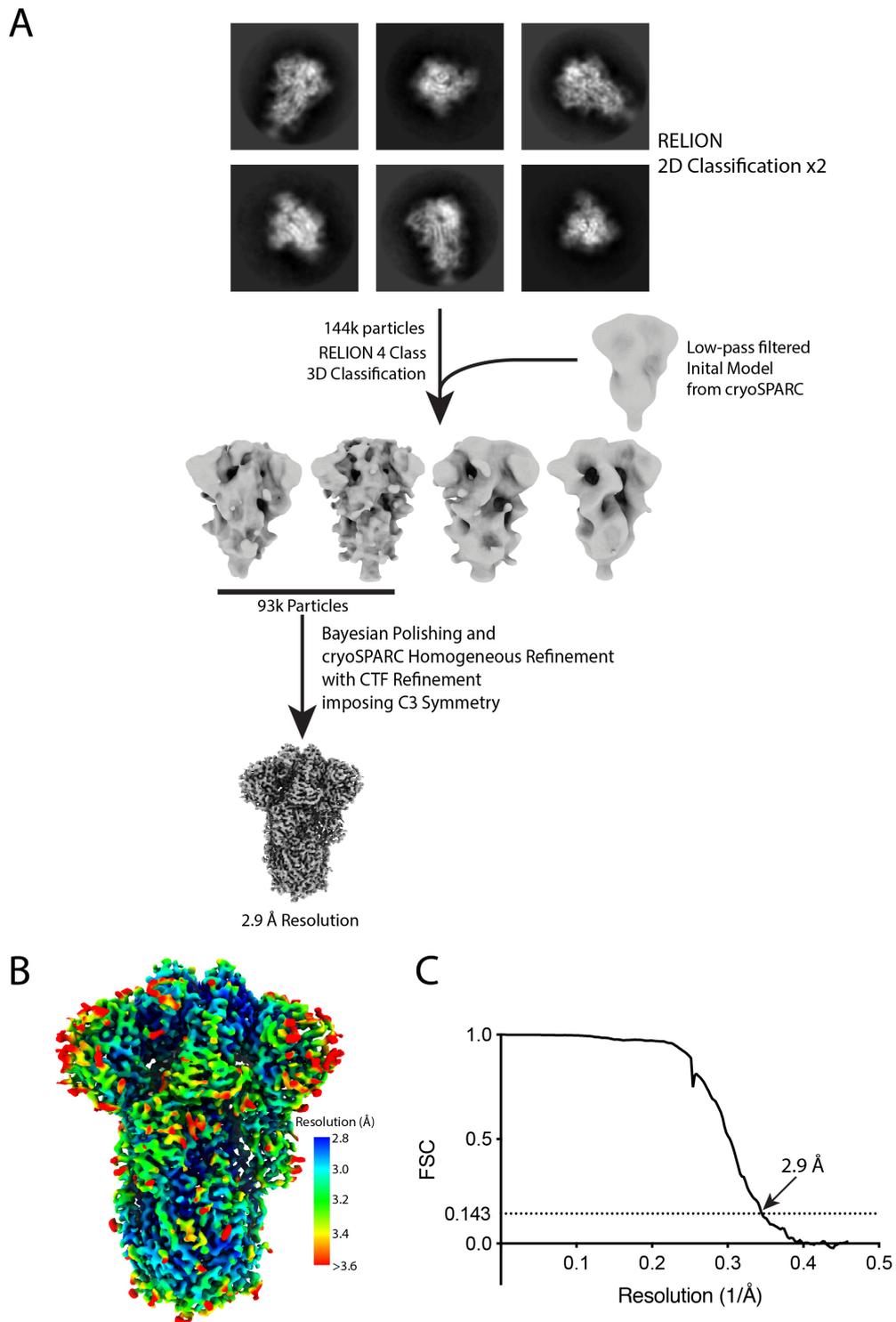
172

173

174 **Figure S1: Biolayer interferometry on Spikes binding to ACE2s from different**
 175 **species. (A)** Plots of biolayer interferometry amplitudes for human (blue), pangolin
 176 (yellow) and bat (red) ACE2s binding to pangolin-CoV S'. **(B, C, D)** Plots of observed
 177 rate constants for human (blue) and pangolin (yellow) ACE2s binding to SARS-CoV-2
 178 S **(B)**, Pangolin-CoV S **(C)** and Pangolin-CoV S' **(D)**. **(E)** Association rate constants,
 179 dissociation rate constants and equilibrium dissociation constants.

180

181

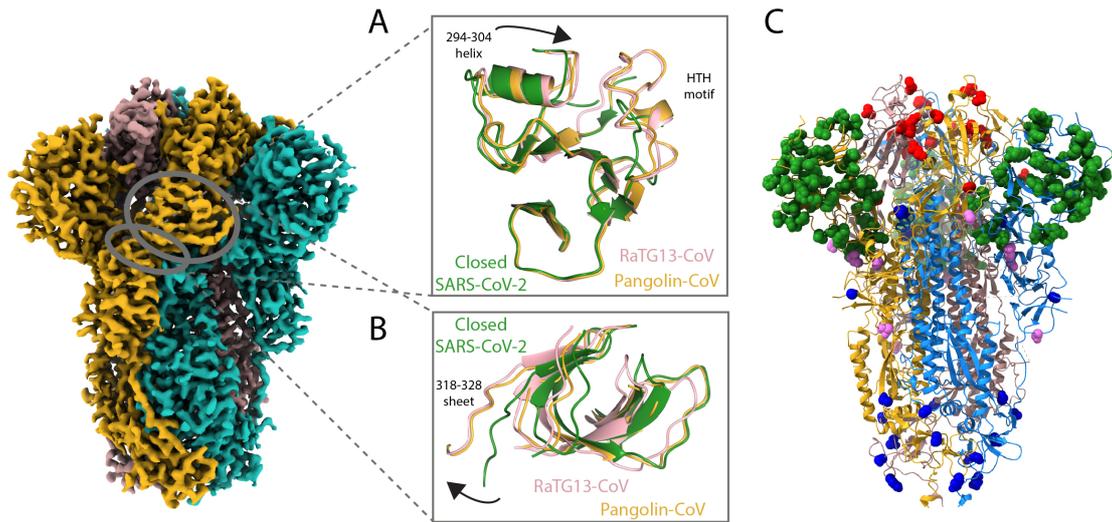


182

183

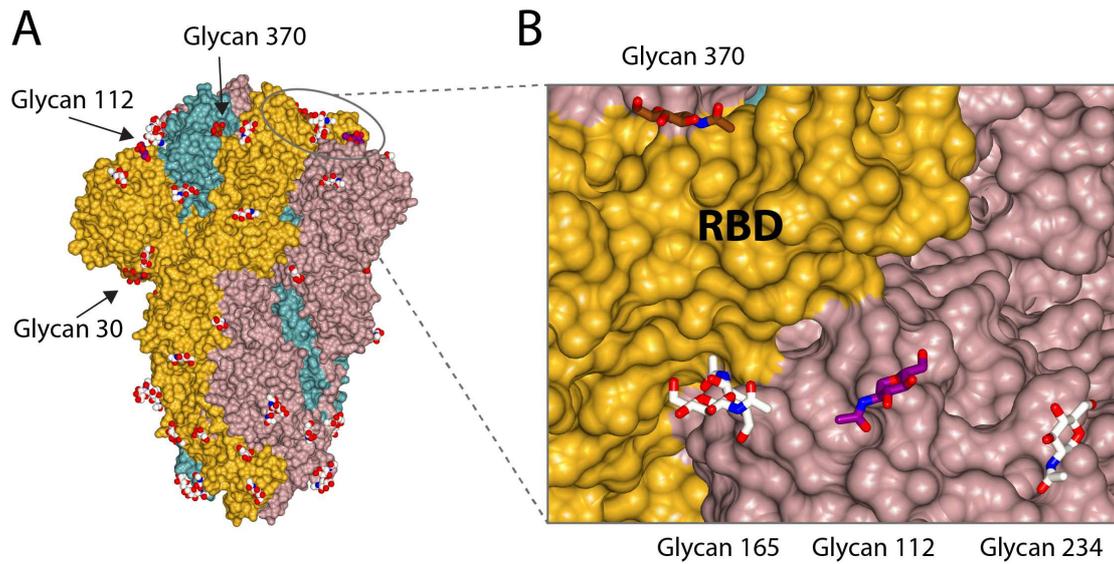
184 **Figure S2: EM Data processing.** (A) Schematic to show steps in cryoEM data
 185 processing. (B) Final EM density coloured by local resolution. (C) Fourier Shell
 186 Correlation (FSC) of final EM map.

187



188

189 **Figure S3: Comparison of Pangolin-CoV S with those from RaTG13 and SARS-**
 190 **CoV-2.** (A, B) Panels corresponding to Fig. 2A, 2C (A) and 2D (B). (C) Molecular
 191 model of the Pangolin-CoV S protein coloured with monomers in blue, golden, and
 192 brown in ribbon representation and substitutions between the Pangolin-CoV S and
 193 SARS-CoV-2 S shown as spheres in: green (for NTD), light plum (for NTD- and RBD-
 194 associated subdomains), red (for RBDs), and navy (S2).
 195



196

197

198 **Figure S4: N-linked glycosylation of Pangolin-CoV S protein.** (A) All glycans
 199 present in the Cryo-EM structure of the Pangolin-CoV S shown in balls representation,
 200 with those present in RaTG13 S and Pangolin-CoV S, but not SARS-CoV-2, shown in
 201 brown and that on Asn112, present uniquely in Pangolin-CoV, in purple. (B) Glycans
 202 on residues 165 and 234 suggested to be important for either inhibiting or enhancing the
 203 RBD erection²⁰ shown in pale cream together with the two other glycans 370 and 112,
 204 not present in SARS-CoV-2 spike, in proximity of the RBD.
 205

206 **Author Contributions**

207 A.G.W, D.J.B, P.X, A.B., C.R, S.R.M performed research, collected and analysed data;
208 A.G.W, D.J.B, P.B.R, J.J.S, S.J.G conceived and designed research and wrote the paper.

209 **Conflict Statement**

210 We have no conflicts of interest to declare.

211 **Data Availability**

212 The map and model have been deposited in the Electron Microscopy Data Bank,
213 <http://www.ebi.ac.uk/pdbe/emdb/> (Accession Nos. XXX,XXX,...). The model has been
214 deposited in the Protein Data Bank, <https://www.ebi.ac.uk/pdbe/> (PDB ID codes XXX,
215 XXX, ...). [Accession numbers will be available before publication].

216 **Acknowledgements**

217 We would like to acknowledge Andrea Nans of the Structural Biology Science
218 Technology Platform for assistance with data collection, Phil Walker and Andrew
219 Purkiss of the Structural Biology Science Technology Platform and the Scientific
220 Computing Science Technology Platform for computational support. We thank Lesley
221 Calder, Peter Cherepanov, George Kassiotis and Svend Kjaer for discussions. This work
222 was funded by the Francis Crick Institute which receives its core funding from Cancer
223 Research UK (FC001078 and FC001143), the UK Medical Research Council
224 (FC001078 and FC001143), and the Wellcome Trust (FC001078 and FC001143). P.X.
225 is also supported by the 100 Top Talents Program of Sun Yat-sen University, the
226 Sanming Project of Medicine in Shenzhen (SZSM201911003) and the Shenzhen
227 Science and Technology Innovation Committee (Grant No. JCYJ20190809151611269).

228

229 References

- 230 1. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of
231 probable bat origin. *Nature* **579**, 270–273 (2020).
- 232 2. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in
233 China. *Nature* **579**, 265–269 (2020).
- 234 3. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science*
235 (80-.). **310**, 676–679 (2005).
- 236 4. Yang, L. *et al.* Novel SARS-like betacoronaviruses in bats, China, 2011.
237 *Emerg. Infect. Dis.* **19**, 989–991 (2013).
- 238 5. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The
239 proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450–452 (2020).
- 240 6. Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein
241 structures inform on virus evolution and furin-cleavage effects. *Nat. Struct.*
242 *Mol. Biol.* 1–5 (2020). doi:10.1038/s41594-020-0468-7
- 243 7. Lam, T. T. Y. *et al.* Identifying SARS-CoV-2 related coronaviruses in Malayan
244 pangolins. *Nature* 1–6 (2020). doi:10.1038/s41586-020-2169-0
- 245 8. Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan
246 pangolins. *Nature* **583**, 286–289 (2020).
- 247 9. Liu, P. *et al.* Are pangolins the intermediate host of the 2019 novel coronavirus
248 (SARS-CoV-2)? *PLOS Pathog.* **16**, e1008421 (2020).
- 249 10. Zhang, T., Wu, Q. & Zhang Correspondence, Z. Probable Pangolin Origin of
250 SARS-CoV-2 Associated with the COVID-19 Outbreak In Brief. *Curr. Biol.* **30**,
251 1346-1351.e2 (2020).
- 252 11. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein.
253 *Science* (80-.). eabd4251 (2020). doi:10.1126/science.abd4251
- 254 12. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu.*
255 *Rev. Virol.* (2016). doi:10.1146/annurev-virology-110615-042301
- 256 13. Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike
257 glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* (2017).
258 doi:10.1073/pnas.1708727114
- 259 14. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion
260 conformation. *Science* (80-.). (2020). doi:10.1126/science.aax0902
- 261 15. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2
262 Spike Glycoprotein. (2020). doi:10.1016/j.cell.2020.02.058
- 263 16. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature*
264 1–4 (2020). doi:10.1038/s41586-020-2179-y
- 265 17. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain
266 bound to the ACE2 receptor. *Nature* (2020). doi:10.1038/s41586-020-2180-5
- 267 18. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage
268 responsible for the COVID-19 pandemic. *Nat. Microbiol.* 1–10 (2020).
269 doi:10.1038/s41564-020-0771-4
- 270 19. Li, X. *et al.* Emergence of SARS-CoV-2 through recombination and strong
271 purifying selection. *Sci. Adv.* eabb9153 (2020). doi:10.1126/sciadv.abb9153
- 272 20. Henderson, R. *et al.* Glycans on the SARS-CoV-2 Spike Control the Receptor
273 Binding Domain Conformation. *bioRxiv Prepr. Serv. Biol.* 2020.06.26.173765
274 (2020). doi:10.1101/2020.06.26.173765

275

Figures

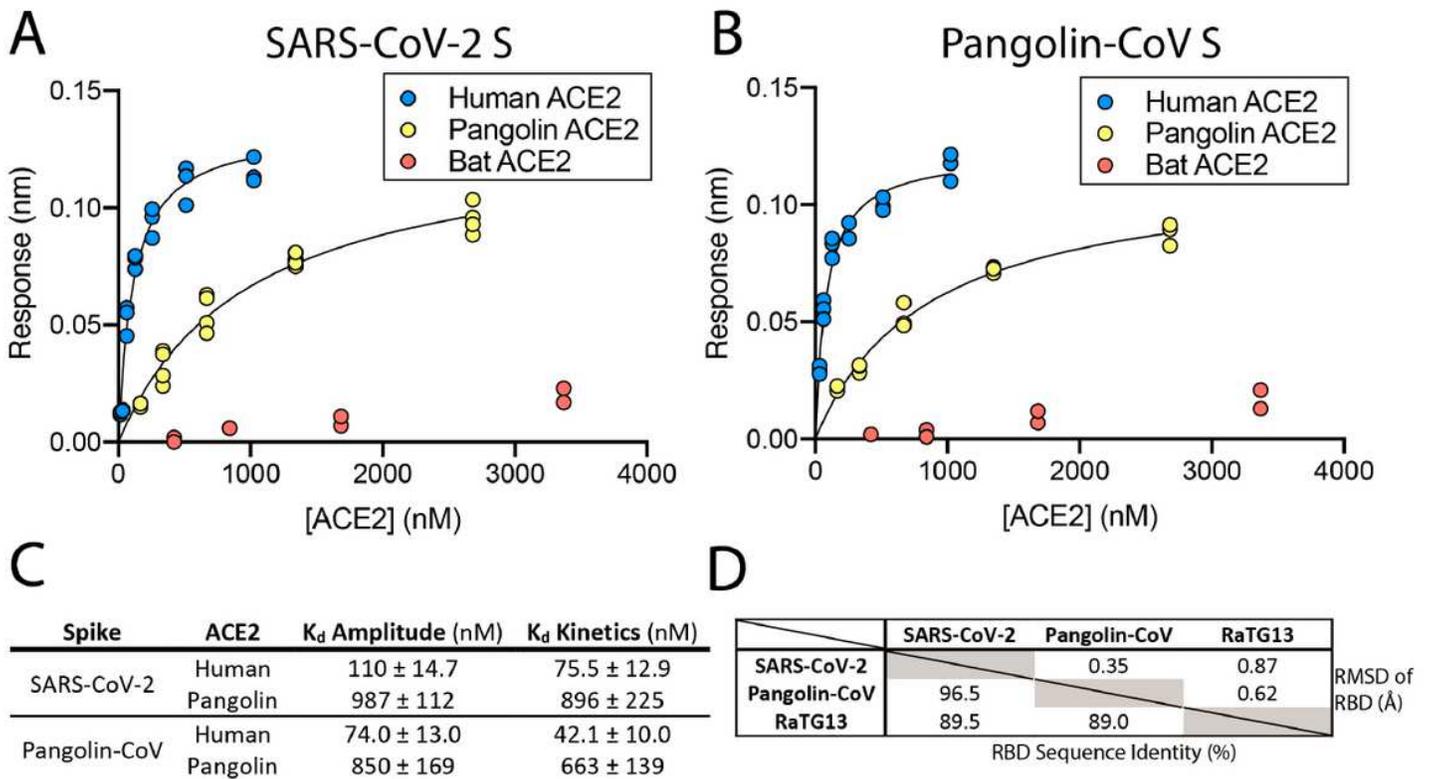


Figure 1

Binding of Pangolin-CoV and SARS-CoV-2 S to ACE2s from different species. (A,B) Plots of biolayer interferometry amplitudes for human (blue), pangolin (yellow) and bat (red) ACE2s binding to SARS-CoV-2 S (A) and pangolin-CoV S (B). (C) Equilibrium dissociation constants determined from the analysis of the data in A and B (K_d Amplitude) compared with values determined from analysis of the corresponding kinetic data (K_d Kinetics) (see Fig. S1). (D) Right hand side: RMSD of atom positions in the structures of RaTG13 S (6ZGF)6, closed conformation of SARS CoV-2 S (6ZGE)6, and Pangolin-CoV S RBD determined in this study. Left hand-side: sequence identity of the RBDs from the same viruses.

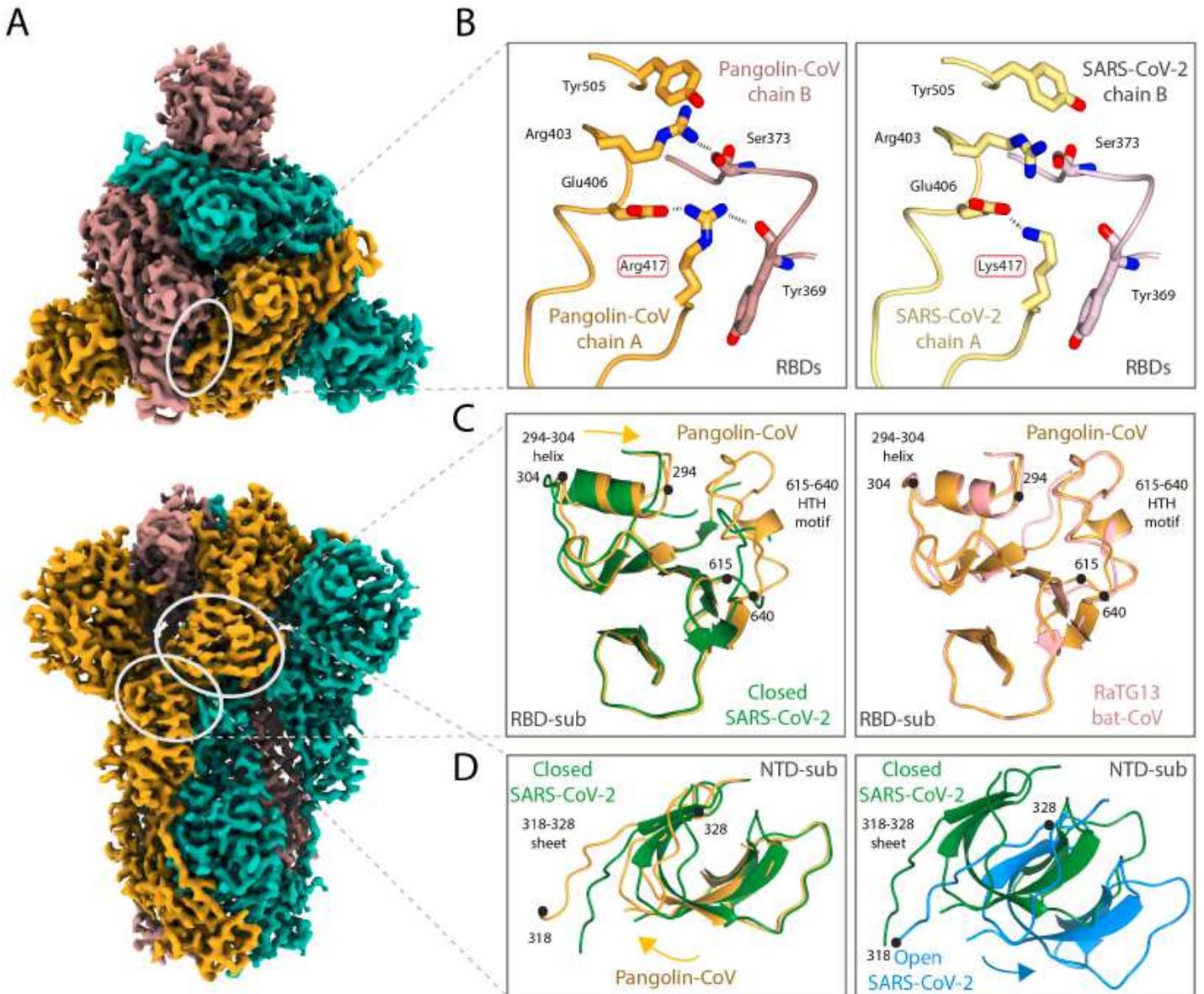


Figure 2

Structure of Pangolin-CoV Spike protein. (A) EM density representation from the 2.9 Å map of Pangolin-CoV S viewed from down the three-fold axis (top panel) and in the orthogonal view (lower panel). The subunits are coloured in blue, golden and brown. The white ovals identify the areas shown in molecular representation on the right. (B) Comparison of the RBD/RBD interface from the pangolin-CoV (left) and SARS-CoV-2 S highlighting the Arg417Lys substitution. (C) Comparison of the RBD associated subdomains of the pangolin-CoV (golden) and closed form of SARS-CoV-2 (green) in the left hand panel, showing the different positioning of the 294-304 helix and the presence of the 615-640 helix-turn-helix in the pangolin structure and, in the right-hand panel, the overlap of the same Pangolin-CoV S structure (golden) with the corresponding region from the RaTG13 (pink). (D) Comparison of (left) the NTD associated subdomain of Pangolin-CoV (golden) with that of the closed form of SARS CoV-2 (green) showing the different domain orientations between them; (right) the closed (green) and open (blue)

conformations of the NTD-associated subdomain of SARS-CoV-2 showing that the shift in orientation of the NTD-associated subdomain on Spike opening is in the opposite direction to the shift seen between the Pangolin-CoV and closed SARS-CoV-2 conformations shown in the left panel.