1

Characterization of the SARS-CoV-2 Spike in an Early Prefusion Conformation

3

4 Tingting Li[1,2,#], Qingbing Zheng[1,2,#], Hai Yu[1,2,#], Dinghui Wu[3,#], Wenhui Xue[1,2,#],

5 Yuyun Zhang[1,2], Xiaofen Huang[1,2], Lizhi Zhou[1,2], Zhigang Zhang[1,2], Zhenghui Zha[1,2],

6 Tingting Chen[1,2], Zhiping Wang[1,2], Jie Chen[1,2], Hui Sun[1,2], Tingting Deng[1,2], Yingbin

7 Wang[1,2], Yixin Chen[1,2], Qinjian Zhao[1,2], Jun Zhang[1,2], Ying Gu[1,2,*], Shaowei Li[1,2,*],

8 Ningshao Xia[1,2,*]

9

10 [1] State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, School

11 of Life Sciences, School of Public Health, Xiamen University, Xiamen, China 361102

12 [2] National Institute of Diagnostics and Vaccine Development in Infectious Disease,

13 Xiamen University, Xiamen, China 361102

14 [3] Department of Pulmonary Medicine, The First Affiliated Hospital of Xiamen

15 University, Xiamen, China 361003

16

17

18 [#] T.L., Q.Z., D.W., H.Y. and W.X. contributed equally to this work.

19 [*] To whom correspondence may be addressed. E-mail: (Y.G.) guying@xmu.edu.cn or

20 (S.L.) shaowei@xmu.edu.cn or (N.X.) nsxia@xmu.edu.cn

21

22 **Running title:** SARS-CoV-2 spike in an early prefusion conformation

23 **Abstract**

24 Pandemic coronavirus disease 2019 (COVID-19) is caused by the emerging severe

25 acute respiratory syndrome coronavirus 2 (SARS-CoV-2), for which there are no

26 efficacious vaccines or therapeutics that are urgently needed. We expressed three

27 versions of spike (S) proteins—receptor binding domain (RBD), S1 subunit and S

28 ectodomain—in insect cells. RBD appears monomer in solutions, whereas S1 and S

29 associate into homotrimer with substantial glycosylation. The three proteins confer

30 excellent antigenicity with six convalescent COVID-19 patient sera. Cryo-electron

31 microscopy (cryo-EM) analyses indicate that the SARS-CoV-2 S trimer dominate in a

32 unique conformation distinguished from the classic prefusion conformation of

33 coronaviruses by the upper S1 region at lower position ~15 Å proximal to viral

34 membrane. Such conformation is proposed as an early prefusion state for the SARS-

35 CoV-2 spike that may broaden the knowledge of coronavirus and facilitate vaccine

36 development.

37 **Key words:** COVID-19, SARS-CoV-2, spike, cryo-electron microscopy, antigenicity,

38 early prefusion conformation

39

Introduction

The novel coronavirus grouped in betacoronavirus genus has become the third serious virus intruder to human in the coronaviridae, after sever acute respiratory syndrome coronaviruses (SARS-CoV) and middle east respiratory syndrome coronavirus (MERS-CoV), recently named SARS-CoV-2. In the phylogenic tree of the coronaviruses, SARS-CoV-2 is genetically close to some bat coronavirus and SARS-CoV, however, with its origin undefined[1]. SARS-CoV-2 causative disease "Coronavirus disease 2019" (abbreviated "COVID-19") is characterized by high fever, dry cough, difficulty breathing and sever atypical pneumonia, which usually be confirmed by virus RNA positive or pulmonary computed tomography (CT) in clinical practice[2, 3]. In terms of higher human-to-human transmissibility, SARS-CoV-2 has spread over 118 countries and areas, and led to over 125,288 confirmed cases worldwide and at least 4,614 deaths, as of March 12th 2020. The World Health Organization (WHO) has declared the SARS-CoV-2 epidemic as a pandemic of international concern and updates the COVID-19 situation every day.

SARS-CoV-2 is an enveloped, single and positive-stranded RNA virus encapsulated with a genome of ~30 kb. At least three membrane proteins including the surface spike protein (S), an integral membrane protein (M), a membrane protein (E). Like other coronaviruses, S is responsible for initiating the engagement to a specific cellular receptor angiotensin-converting enzyme 2 (ACE2) and mediating the cell-virus membrane fusion by the class I fusion mechanism[4, 5]. Thus, S is the main target for neutralizing antibodies against viral infection and the core immunogen constituent of

62 vaccine design. S is consisted of S1 and S2 subunits and the cleavage on S1/S2

63 boundary by protease during biosynthesis is prerequisite for coronaviruses cellular

64 membrane fusion and subsequent infection[6]. SARS-CoV-2 evolves a 4-residue

65 insertion (RRAR) as potential furin cleavage site rather than SARS-CoV and other bat

66 coronaviruses, which may contribute to the higher transmissibility of this novel

67 coronavirus[6, 7]. Previous studies suggested the infection process of MERS-CoV[8] and

68 SARS[9] viruses, where S trimer undergoes conformational transition from a prefusion

69 conformation ready for ACE2 binding to a postfusion conformation for eventual virus-

70 cell membrane fusion. Structure determination of SARS-CoV and MERS-CoV spike

71 trimers captured a variety of scenarios in the prefusion conformation showing partial

72 (one or two) receptor-binding domain (RBD) in the "up" conformation and the rest in

73 the "down", and all in either "up" or "down". The conformation transition from "down"

74 to "up" could expose the receptor binding site, and the subsequent receptor engagement

75 would lead to a substantial conformation rearrangement of S trimer from prefusion

76 conformation to postfusion. Two recent studies[7, 10] reported cryo-electron microscope

77 (cryo-EM) structures of SARS-CoV-2 spike trimers in the prefusion conformation with

78 2 RBDs down and 1 RBD up. In the case of SARS-CoV, this conformational change

79 during RBDs "down" to "up" was associated with the binding of receptor ACE2 as well

80 as the recognition of neutralizing monoclonal antibodies [11].

81   A safe and efficacious vaccine is urgently needed to control and eliminate the

82 SARS-CoV-2 infection. Various forms of vaccine candidates, mostly aiming to elicit

83 neutralizing antibodies against S proteins, are under preclinical research or even

84   subjected to clinical trials[12]. Here, we cloned S ectodomain and its fragments RBD and

85   S1 into recombinant baculovirus and expressed the proteins in insect cells. We found

86   that S and S1 formed homotrimer in solutions and the three proteins reacted well with

87   COVID-19 convalescent human sera. Cryo-EM analysis demonstrated the S trimer

88   unexpectedly retains a unique conformation distinguished from the classic prefusion

89   conformation of coronavirus spikes, that may represent an early state rather than the

90   known prefusion conformation of S spike. These results might broaden the knowledge

91   on coronavirus virology and provide another protective conformation of S trimer for

92   structure-based vaccine design against SARS-CoV-2 infection and its causative

93   COVID-19.

94

95   **Results**

96   **Construct design, expression and purification of SARS-CoV-2 S proteins**

97   To screen a potent immunogen for COVID-19 vaccine development, we designed three

98   constructs—S ectodomain, S1 and RBD—for the SARS-CoV-2 Spike (S) protein

99   expression by aligning the SARS-CoV-2 S gene (Genbank accession no. NC_045512.2)

100   to a SARS-CoV strain (Genbank accession no. NC_004718) S gene sequence in terms

101   of structure-defined domain profile of the SARS-CoV S protein (Fig. 1A). The gene of

102   SARS-CoV-2 S ectodomain encoding amino acids (aa) 15-1,213 with removal of its

103   original signal sequence was cloned to the downstream of the gp67 signal sequence in

104   pAcgp67B plasmid vector (Fig.1B). and with its C-terminal addition of a thrombin

105   cleavage site, a T4 trimerization foldon motif and his tag. The segments S1 (aa 15-

106 680) and RBD (aa 319-541) were cloned similar to S ectodomain, keeping gp67 signal

107 peptide and his tag to facilitate secretory outside cell and affinity purification,

108 respectively, but without the thrombin site and T4 foldon (Fig. 1A). The three

109 constructed plasmids were respectively co-transfected into Sf9 insect cells with v-

110 cath/chiA gene deficient baculovirus DNA for the generation and amplification of

111 recombinant baculovirus, which were then harnessed to infect Hive Five insect cells to

112 eventually produce recombinant proteins, named S, S1 and RBD respectively.

113   The recombinant proteins were mostly soluble expressed and secreted into the

114 culture medium. The centrifugation supernatants of cell culture went through metal

115 affinity chromatography using Ni-NTA resin. S, S1 and RBD proteins were mainly

116 eluted in a separation fractions under 250 mM imidazole elution, and resolved as

117 molecular weight (m.w.) of ~180 kDa, 110 kDa and 35 kDa, respectively, in SDS-

118 PAGE as indicated by a corresponding western blotting (WB) using anti-His antibody

119 as detection antibody (Fig. 1C). Interestingly, about one half S proteins were cleaved

120 into S1 (identical migration site to S1 lane in Fig. 1C) and S2 (about 80kDa developed

121 in anti-His WB) possibly by innate furin of insect cell that was also found in other cases

122 of enzymatic cleave while protein expression in insect cell, such as Flu HA [13]. The

123 eluted S fraction was further polished by Superdex 200 to remove contaminative

124 proteins (Fig. 1D). These peaks fractionated at retention volume 28mL, 36mL, 48mL,

125 and 65mL, were further harvested and subjected to SDS-PAGE analysis. The results

126 indicated that S proteins together with cleaved S1/S2 were resolved at peak 1 in size-

127 exclusion chromatography (Fig. 1D) and showed a high purity of over 95% total

128   S/S1/S2 in gel (Fig. 1E). Overall, one-step Ni-NTA affinity chromatography produced

129   RBD with 95% purity and a yield of 30 mg per L cell culture, S1 with about 90% purity

130   and 10 mg per L yield, while further purification through a size-exclusion

131   chromatography (SEC), the resultant S sample had over 95% purity regarding intact S

132   and cleaved S1/S2, and was harvested in a yield of 1 mg per L cell culture. These data

133   set up a start point for further optimization on expression and purification process of

134   SARS-CoV-2 S immunogen candidates through insect baculovirus expression vector

135   system (BEVS).

136

137   **Physiochemical properties of SARS-CoV-2 S-RBD, S1 and S proteins**

138   We next investigated the physiochemical properties of the recombinant S protein and

139   its fragments purified from insect cells, including association potential, thermal

140   stability and glycosylation situation. Firstly, high pressure size-exclusion

141   chromatography (HPSEC) and sedimentation velocity analytical ultracentrifugation

142   (SV-AUC) analyses were carried out to measure the oligomerization potential of the

143   three proteins in solution. RBD, S1 and S all showed single major peak in HPSEC

144   profiles at elution volume of 9.0 mL, 5.5 and 5.3 mL, respectively (Fig. 2A and Fig.

145   2B). RBD, S1 and S were further verified by SV-AUC, where RBD sedimented as

146   single species of 3.1S in c(s) profile, corresponding to apparent molecular weight 22

147   kDa (Fig. 2D); S1 existed as a dominant species of 11.3 S (estimated as 277 kDa

148   corresponding to S1 trimer) and a minor aggregate form of 20 S (Fig. 2E); S and cleaved

149   S1/S2 resolved as 15.2 S, equivalent to 577 kDa, approximately as the theoretical

150   molecular weight of intact S trimer. The three proteins were further analyzed by

151   differential scanning calorimetry (DSC) that was usually used to investigate the inner

152   thermostability of macromolecules or their complexes[14]. RBD and S1 showed one

153   major peak at comparable thermal denaturation midpoints (Tm) of 46.0 °C and 45.5 °C,

154   respectively (Fig. 3G and 3H), whereas S sample showed two major peaks at Tm of

155   45.5°C (identical to Tm of S1) and 64.5°C (Fig. 3I), which might reflect the coexistence

156   of intact S and cleaved S1/S2.

157   On the other hand, we investigated the glycosylation extent of the three protein by

158   enzymatic deglycosylation analysis. Endo H could unleash the chithobiose core of high

159   mannose and some hybrid oligosaccharides from N-liked glycoproteins, therefore

160   remove the extended branches of glycans and leave the one N acetylglucosamine

161   (GlcNAc) on N-linked glycoproteins. While PNGase F would release N-linked glycan

162   moieties between GlcNAc and ASN residues within a glycoprotein. It should be noted

163   that glycosylation in insect cells is featured as terminal mannose glycans, unlike

164   complex sialylated glycans in mammalian cells, and glycosylation is known to correlate

165   the immunogenicity and broad-coverage protection of a glycoprotein immunogen[15, 16].

166   After the treatment of either Endo H or PNGase F, RBD showed no discernible decrease

167   of molecular weight in SDS-PAGE/anti-His WB, S1 and S2 both demonstrated nearly

168   ~10 kDa decrease, and the intact S exhibited substantial shrinkage in molecular weight

169   of about ~20 kDa decrease (Fig. 2J). The analyses conclude that the glycosylation

170   extent within S glycoprotein is RBD < S1 ~ S2, consistent to the predicted glycosylation

171   profile of S polypeptide (Fig. 1A).

172

Reactivity of SARS-CoV-2 RBD, S1 and S proteins against convalescent COVID-19 human sera

We next evaluated the antigenicity of the three versions of S proteins by WB and ELISA using a panel of six COVID-19 convalescent human sera, which was collected from COVID-19 patients after they recovered from the disease in the First Affiliated Hospital of Xiamen University. Eight reducing SDS gel duplicates of the one depicted in Fig. 1C were prepared for WB analysis using these six convalescent sera and two control sera from health human (Fig. 3A-3H, left panel). As expected, intact S protein bands reacted well with all the six convalescent sera (Fig. 3A-3H, left panel). Unexpectedly, five of six sera showed no or very weak reactivities against RBD, only Serum #6 possessed RBD's activity. Among the five sera with lower RBD-reactivity, Serum #2, #3 and #5 well recognized S1 and the cleaved S1 in lane S, suggesting these sera may specifically react with NTD of S1. S2 demonstrated reaction activity against all the six sera, like the intact S. No detectable reaction was observed in the control sera (Fig. 3G and 3H). Inconsistent to the WB results, RBD, S1 and S shared comparable reactivities against the convalescent sera in ELISA, although the sera per se presented varied reaction titers (represented as ET50) following the reaction sequence: Serum #5 > #2 > #3> #1 > # 6 > #4 (Fig. 3A-3H, right panel).

Taken together, RBD, S1 and S proteins from insect cells maintain the native-like SARS-CoV-2 epitopes. These epitopes in native virion should be immunogenic in COVID-19 patients and capable of eliciting high antibody titer in the convalescent

194 phase of SARS-CoV-2 infection. Among these epitopes, most RBD epitopes are strictly

195 tertiary conformation-dependent sites that are damaged upon the mild denatured

196 condition of reductant and SDS treatment, NTD within S1 bears some linear epitopes,

197 whereas S2 part essentially has linear epitopes that are immunogenic in all COVID-19

198 patients (n=6).

199

200 **Cryo-EM structures of SARS-CoV-2 S proteins**

201 To examine the structure of the trimeric S ectodomain with native sequence, we

202 prepared cryo-EM grids using the Ni-NTA purified S proteins and collected 1,513

203 electron micrograph movies. Most of motion-corrected micrographs demonstrated

204 plenty of well-dispersed particles with an approximate size as the canonical coronavirus

205 S trimer (Fig. 4A). A total of 162,645 particles were picked out for multiple rounds of

206 2D classification, consequently, 37,147 particles grouped into top 10 classes, rendering

207 typical feature of S trimer in prefusion conformation as recently reported[7, 10], were

208 selected for further analysis (Fig. 4B). 3D reconstruction (applying 3-fold symmetry)

209 yielded the density map of prefusion spike (S-pre) at resolution of 5.43 Å (Fig. 3D and

210 Supplementary Fig. 1A).

211   Structurally, three S monomers intertwine around each other and associate to

212 homotrimers with 145 Å height seen from side-view and 160 Å diameter in top-view

213 (Fig. 4C and 4D). We then recruited the recently reported cryo-EM map of S prefusion

214 trimer (EMD-21374, at resolution of 3.17 Å, low pass to 5.43 Å prior to structural

215 comparison) and compared our cryo-EM map at same resolution (Fig. 4D). It was

216   worthy noted that the compared prefusion SARS-CoV-2 S trimer was engineered with

217   site-directed mutations to stabilize prefusion conformation and expressed in 239F cells.

218   The mutant included two stabilizing proline mutations at residues 986, 987 and a

219   "GSAS" substitution at the furin cleavage site[7]. Surprisingly, the alignment

220   demonstrated that the two cryo-EM structures share similar mushroom-shaped

221   architecture in particular nearly identical at stalk moiety (S2 region), but our S-pre

222   shows the cap part (S1 region) at ~15Å lower position than the reported S trimer in

223   RBD-down prefusion conformation (Fig. 4D). Regarding to substantial mismatch at the

224   density of 3 S1 subunits, we respectively fitted 5 individual domains (NTD, RBD, SD1,

225   SD2 and S2) of the SARS-CoV-2 S structure (PDB code 6VSB) to our S-pre map. In

226   the fitting map, NTD, RBD, SD2 and S2 could be well placed in the S-pre map,

227   especially for the latter two, which reflects the aforementioned good match at the stalk

228   of the mushroom-shape (Supplementary Fig. 2). However, there is no observable

229   density between RBD and SD2 to accommodate an SD1 model (Supplementary Fig. 2),

230   which suggests SD1 region is dramatically flexible in our S-pre structure (Fig. 4E and

231   4F). When the combined model of fitted NTD-RBD-SD2-S2 was superimposed to the

232   original S protomer structure (PDB code 6VSB, Chain A, RBD in down conformation),

233   both NTD and RBD in the original S obviously move and rotate up against our

234   combined model (Fig. 4G). The structural comparison demonstrated that the S-pre

235   trimer retains a unique conformation different from the prefusion conformation of the

236   two reported SARS-CoV-2 spike structures (PDB codes 6VSB and 6VXX).

237   We then compared the conformation of our S-pre structure with that of 21

238   deposited coronavirus S models. Six representative S structures[7, 17-20] from four known

239   genus (α-, β-, γ- and δ-genus) in coronaviridae are respectively fitted to the S-pre map

240   (Supplementary Fig. 3). Five S trimer structures of other coronaviruses share similar

241   prefusion conformation with the reported SARS-CoV-2 S structure but substantially

242   distinct with the unique conformation of our S-pre.

243   Apart from most particles classified as S-pre in our sample, 2D classifications also

244   showed five classes of few particles (2,951) assuming an elongated rosette-shape

245   assembly. These particles were further reconstructed and yielded a structure at lower

246   resolution of 8.40 Å (Supplementary Fig. 1B) that could be considered as post-fusion

247   spike (S-post), as the structure has similar shape but shorter length (~170 Å) as

248   compared to the postfusion spike of SARS-CoV[21] and the presumed one observed in

249   native SARS-CoV-2 virion (BioRxiv, https://doi.org/10.1101/2020.03.02.972927)

250   (Supplementary Fig. 4) . Fitting the S-post map with the core region structure of SAR-

251   CoV-2 S2 subunit in post-fusion conformation (PDB code 6LXT) indicated that our S-

252   post exhibits roughly rod shape similar with the post-fusion structure (Supplementary

253   Fig. 4).

254

255   **Conformational transition of SARS-CoV-2 spike from early prefusion to**

256   **postfusion**

257   Briefly, we've obtained two conformations of SARS-CoV-2 spike from insect cells.

258   The dominant one maintains the similar mushroom-shaped trimer as the previous

259   models, while the S1 region substantially diverges. The other conformation essentially

260   resembles the postfusion state. However, we could not find the classic prefusion

261   conformation in our sample. We next tried to figure out at which stage the unique

262   conformation occurs during the spike conformation change. The space relationship of

263   NTD or RBD to S2 domain in the unique, RBD-down and RBD-up prefusion

264   conformations (Fig. 5A) was measured by a reference plant approximately parallel to

265   the viral membrane. The plane is defined by the positions of three equivalent Cα atoms

266   (residue 694 be used) from three S2 subunits of the trimer structures. Numerical data

267   shows that (1) the NTD and RBD in the unique conformation retain the lowest position

268   in the three prefusion conformations; (2) the NTD and RBD of RBD-down prefusion

269   stretch upward 16.6° rotation/16 Å elevation, and 13.1°/18 Å, respectively, with respect

270   to the unique conformation; (3) from RBD-down to RBD-up prefusion state, the RBD

271   elevates 9 Å with an additional rotation whereas the NTD remaining nearly stationary

272   (Supplementary Movie S1). The resultant "up" RBD is ready for ACE2 binding and

273   the spike eventually is rearranged to postfusion state upon RBD-ACE2 interaction[11, 22]

274   (Fig. 5B). The motion trend of NTD and RBD from prefusion to postfusion state in

275   conformational transition is away from the viral membrane, suggesting the unique

276   conformation may occur earlier than RBD-down prefusion conformation, named as

277   "early prefusion conformation" (Fig. 5). This early prefusion conformation might exist

278   in other coronaviruses as well.

279

280

Discussion

SARS-CoV-2 has crossed the species barrier and sweep over the planet by person-to-person transmission in an R0 ~2.56 rate[23], first wave in China and the second wave booming outside China. WHO has declared the event as another pandemic infectious disease in human history, and the epidemiology of SARS-CoV-2 infection is still in data accumulation. Although the biology and virology of SARS-CoV-2 remain elusive, in terms of knowledge on other coronaviruses, the spikes decorating the SARS-CoV-2 virion play a critical role in viral attachment and entry to host cells. Cryo-EM structures of spikes in the prefusion conformation, and RBD-bound receptor ACE2 have indicated the engagement of SARS-CoV-2 to cellular membrane requires a serial of conformational change of RBDs. The change is presumed from the start point of 3 RBDs down in the prefusion conformation, then RBD(s) up for ACE2 binding, and eventually spike is rearranged to postfusion. In this study, we suggest that the SARS-CoV-2 spike may retain at more precedent state than the classic prefusion conformation that has been determined for other coronaviruses. This early prefusion conformation features that the cap of the mushroom-shaped spike constituted by three S1 subunits is more proximal to viral membrane by 15 Å than in the classic prefusion conformation.

The SARS-CoV-2 spike expressed in insect cells predominantly retains a unique early prefusion conformation, which was repeatable in at least three batches of samples and is ascribed to two possible reasons – native aa sequence used in the S ectodomain construct and over-expression in insect cells. There is about a half of S proteins undergoing cleavage on the S1/S2 boundary site in the purified samples both after the

303   first Ni-NAT and the second SEC purification (Fig. 1C and 1E). Further analyses

304   suggest the split between S1 and S2 likely takes no effect on the trimerization of S

305   trimer. It is known that the insect cells can confer post-translation glycosylation for

306   protein over-expression as mammalian cells despite the latter can produce more

307   complex sialyation[24], and thus provide an alternative way to generate glycoprotein in

308   native conformation. Our results indicate that RBD, NTD and S2 domains of SARS-

309   CoV-2 demonstrated different glycosylation extent in insect cells, however, RBD, S1

310   and S proteins comparably react well with six convalescent COVID-19 human sera

311   albeit they differ in domain composition, polypeptide length and oligomerization.

312   There are numbers of SARS-CoV-2 vaccine candidates, including inactivated,

313   vectored, recombinant and nucleotide vaccine forms, under preclinical research.

314   Various versions of S proteins are the major targets for vaccine immunogen candidate.

315   In addition to potent neutralizing antibody elicitation upon immunization, potential

316   antibody-dependent disease enhancement (ADE) is the major concern for an efficacious

317   SARS-CoV-2 vaccine. ADE has been found in the development of numbers of virus

318   vaccine candidates, including respiratory syndrome virus (RSV), dengue fever [25, 26],

319   human immunodeficient virus (HIV), SARS-CoV, MERS-CoV [25-27] and so on. It is

320   believed that ADE is associated with non-neutralization epitope attribute and / or

321   specific antibody isotype[28, 29], in which virus-bound antibody would promote the viral

322   infection to immune cells through Fc fragment targeting γFc receptors on the cellular

323   surface and enhance the disease severity. Therefore, the strategy of vaccine design

324   against SARS-CoV-2 should include the consideration of antigen region selection,

325 glycosylation number/extent and exactly presented prefusion conformation. The

326 prefusion conformation needed to be maintained is exemplified by the case of RSV

327 vaccine candidate in which F trimer in prefusion is much potent than postfusion[30-32].

328 Hence the early prefusion conformation proposed for SARS-CoV-2 spike should be

329 drawn an attention for immunogen design as well as the prefusion one.

330   In conclusion, we obtain three kinds of S proteins showing excellent antigenicity

331 and find an early prefusion conformation for SARS-CoV-2 spike. Nevertheless, the

332 molecular level detail for such conformation and the underlying immunogenicity

333 should be further investigated, and whether this conformation recapitulates the exact

334 state of spike in native SARS-CoV-2 virion remains to be determined.

335

336 **Materials and Methods**

337 **Cloning, protein expression and purification**

338 The SARS-CoV-2 S gene (Genbank accession no. NC_045512.2) was synthesized and

339 cloned into a baculovirus shuttle vector pAcgp67B (BD Biosciences, CA, USA) using

340 Gibson assembly. The S construct encoding aa 15-1,213 (numbered as original

341 sequence), contains a thrombin site, a T4 foldon domain to assist in trimerization and a

342 C-terminal 10-His tag for purification. For S1 construct contains gene encoding aa 15-

343 680 followed by a 10-His tag. The RBD construct (aa 319-541) also contains 10-his tag

344 to facilitate purification. In all three constructs, the natural signal peptide (aa 1-14

345 analyzed by SignalP tool) was replaced with a gp67 secretion signal peptide at N-

346 terminus.

347   The expression and purification of proteins were performed as described

348   previously[33]. All plasmids were co-transfected with linearized 2.0 DNA (deficient in *v-*

349   *cath/chiA*genes) (Expression Systems, CA, USA) into *Sf9* insect cells (Thermo Fisher

350   Scientific, MA, USA), according to the protocol provided by the manufacturer

351   (Expression Systems). The transfection supernatant was harvested and amplified 2

352   times to obtain a high titer of the recombinant viruses. Hive Five cells (BTI-TN-5B1-

353   4) (Thermo Fisher Scientific) were cultured in ESF921 medium (Expression Systems)

354   and infected with recombinant virus at an multiplicity of infection (MOI) of 5 in the

355   exponential growth phase (2×10^6 cells/ml, 95% viability) at 28°C for 72 h. The culture

356   media was centrifuged at 8,000 rpm for 20 min. Then the supernatant was dialyzed

357   against phosphate-buffered saline (PBS), pH 7.4, and purified with Ni-sepharose fast

358   flow 6 resin (GE Healthcare, Boston, USA) by the elution with 250 mM imidazole. The

359   protein concentrations of the final purified samples were measured with Pierce™BCA

360   Protein Assay Kit (Thermo Fisher Scientific).

361

362   **SDS-PAGE and western blot**

363   Protein samples were mixed with loading buffer and boiled for 10 min, and subjected

364   to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Equal

365   amounts of proteins for each sample were loaded onto two SDS-PAGE gels, one for

366   western blotting and one for Coomassie staining. The proteins were electrophoresed for

367   70 min at 80 V in a BioRad MINI-PROTEAN Tetra system (BioRad Laboratories, CA,

368   USA), and the gel was stained with Coomassie Brilliant Blue R-250 (Bio-Rad) for 30

369   min at room temperature. For western blotting, separated proteins were transferred onto

370   a nitrocellulose membrane (Whatman, Dassel, Germany) using a Trans-Blot Turbo

371   transfer system (Bio-Rad). The membrane was blocked and then incubated for 1 h with

372   an His-tag-specific mouse mAb antibody (Proteintech, Rosemont, USA) or human sera

373   (1:500 dilution). Unbound antibody was removed by five 5-min washes and the

374   membrane was incubated with alkaline phosphatase-conjugated goat anti-mouse

375   secondary antibody or goat anti-human IgG secondary antibody (Abcam, Cambridge,

376   UK). Membranes were washed again and then developed using SuperSignal ELISA

377   Pico Chemiluminescent Substrate Kit (Thermo Fisher Scientific).

378

379   **Enzyme-Linked Immunosorbent Assay (ELISA)**

380   Purified proteins were coated onto 96-well microtiter plates at 100 ng/well in PBS at

381   37°C for 4 h. The background was blocked with 1 × Enzyme dilution buffer (PBS +

382   0.25% casein + 1% gelatin + 0.05% proclin-300) at 37°C for 2 h. Sera were diluted

383   started at 1:100 followed with three-fold serially dilution, and added to the wells (100 µl)

384   and incubated at 37°C for 1 h. Horseradish peroxidase (HRP)-labeled mouse anti-

385   human antibody (Abcam) was used as secondary antibody at 1:5,000 for 30 min. Wells

386   were washed again and the reaction catalyzed using o-phenylenediamine (OPD)

387   substrate at 37°C for 10 min. The OD450nm (reference, OD620nm) was measured on

388   a microplate reader (TECAN, Männedorf, Switzerland), with a cut-off value 0.1. The

389   Half effective titers (ET50) was calculated by sigmoid trend fitting using GraphPad

390   Prism software.

Size-Exclusive Chromatography (SEC)

Ni-NTA purified S proteins were further loaded into Superdex200 (GE Healthcare), the fractions were harvested and analyzed by SDS-PAGE. All high-purity RBD, S1 and S proteins were subjected to HPLC (Waters; Milford, MA) analysis using a TSK Gel G5000PWXL7.8 × 300 mm column (TOSOH, Tokyo, Japan) equilibrated in PBS, pH 7.4. The system flow rate was maintained at 0.5 mL/min and eluted proteins were detected at 280 nm.

Analytical Ultracentrifuge (AUC)

The AUC assay was performed using a Beckman XL-Analytical ultracentrifuge (Beckman Coulter, Fullerton, CA), as described elsewhere[34]. The sedimentation velocity (SV) was carried out at 20°C with diluted proteins in PBS. The AN-60 Ti rotor speed was set to 20,000-30,000 rpm according to the molecular weight of the control proteins. Data was collected using SEDFIT computer software, kindly provided by Dr. P. C. Shuck (NIH, Bethesda, MA, USA). Multiple curves were fit to calculate the sedimentation coefficient (S) using continuous sedimentation coefficient distribution model [c(s)], and then the c(s) used to estimate protein molar mass.

Differential scanning calorimetry (DSC)

Differential scanning calorimetry (DSC) was carried out on the S proteins using a MicroCal VP-DSC instrument (GE Healthcare, MicroCal Products Group, Northampton, MA) as described previously[14]. In brief, all samples with a concentration

413   of 0.2 mg/mL were measured at a heating rate of 1.5°C /min with the scan temperature

414   ranging from 10°C to 90°C. The melting temperatures (Tm) were calculated using

415   MicroCal Origin 7.0 (Origin-Lab Corp., Northampton, MA) software assuming a non-

416   two-state unfolding model.

417

418   **Endo-H and PNGase-F digestion**

419   The Endo-H (NEB) and PNGase-F (NEB) digestions were performed according to the

420   protocol offered by instruction. In brief, the deglycosylation reactions were carried out

421   using 10ug S proteins with 5uL of Endo H or PNGase F and incubated at 37°C

422   overnights. The reactions were loaded in to SDS-PAGE and analyzed by Western

423   blotting using anti-His as detecting anybody.

424

425   **Cryo-EM sample preparation and data collection.**

426   Aliquots (3 μL) of 0.5 mg/mL purified SARS-CoV-2 S protein were loaded onto glow-

427   discharged (60 s at 20 mA) holey carbon Quantifoil grids (R1.2/1.3, 200 mesh,

428   Quantifoil Micro Tools) using a Vitrobot Mark IV (ThermoFisher Scientific) at 100%

429   humidity and 4°C. Data were acquired using the EPU software to control a FEI Tecnai

430   F30 transmission electron microscope (ThermoFisher Scientific) operated at 300 kV.

431   and equipped with a ThermoFisher Falcon-3 direct detector. Images were recorded in

432   the 58-frame movie mode at a nominal magnification of 93,000X with a pixel size of

433   1.12 Å. The total electron dose was set to 46 e$^-$ Å^{-2} and the exposure time was 1.5 s.

434 537 micrographs were collected with a defocus range comprised between 1.5 and 2.8

435 μm.

436

Cryo-EM data processing

438 Movie frame alignment and contrast transfer function estimation of each aligned

439 micrograph were carried out with the programs Motioncor[35] and Gctf[36]. Particles were

440 picked by the 'Templete picker' session of cryoSPARC v2[37]. Two rounds of reference-

441 free 2D classification were performed and well-defined particle images were selected

442 and non-uniform 3D refinement, 3D reconstruction with C3 symmetry were performed

443 using cryoSPARC v2. The resolutions of the final maps were estimated on the basis of

444 the gold-standard FSC curve with a cutoff at 0.143[38]. Density-map-based visualization

445 and segmentation were performed with Chimera[39].

446

Acknowledgments

452

Financial Disclosure

454 The funders had no role in study design, data collection and analysis, decision to publish,

455 or preparation of the manuscript.

456

Competing Interest

The authors have declared that no competing interests exist.

459

Author Contributions

Y.G, S.L. and N.X. designed the study. T.L., Q.Zheng., H.Y., D.W., W.X., Y.Z.,

X.H., L.Z., Z.Zhang., Z.Zhai., T.C., Z.W., J.C., H.S. and T.D. performed experiments.

T.L., Q.Z., H.Y., Y.W., Y.C., Q.Zhao., J.Z., Y.G., S.L. and N.X. analyzed data. T.L.,

Q.Z., H.Y., Y.G., and S.L. wrote the manuscript. T.L., Q.Zheng., H.Y., D.W., W.X.,

Q.Zhao., J.Z. Y.G., S.L., and N.X. participated in discussion and interpretation of the

results. All authors contributed to experimental design.

467

References

1 Chan JF, Kok KH, Zhu Z *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* 2020; **9**:221-236.

2 Zhu N, Zhang D, Wang W *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382**:727-733.

3 Wang D, Hu B, Hu C *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 2020.

4 Zhou P, Yang XL, Wang XG *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**:270-273.

5 Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol* 2020.

6 Millet JK, Whittaker GR. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Research*; **202**:120-134.

7 Wrapp D, Wang N, Corbett KS *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, NY)* 2020.

8 Walls AC, Xiong X, Park Y-JJ *et al.* Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* 2019; **176**:1026.

9 Yuan Y, Cao D, Zhang Y *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike

487 glycoproteins reveal the dynamic receptor binding domains. *Nature communications* 2017;
488 **8**:15092.

489 10 Walls AC, Park Y-JJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and
490 Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020.

491 11 Gui M, Song W, Zhou H *et al.* Cryo-electron microscopy structures of the SARS-CoV spike
492 glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell research* 2017;
493 **27**:119-129.

494 12 Lu S. Timely development of vaccines against SARS-CoV-2. *Emerg Microbes Infect* 2020; **9**:542-
495 544.

496 13 Sui J, Hwang WC, Perez S *et al.* Structural and functional bases for broad-spectrum
497 neutralization of avian and human influenza A viruses. *Nature structural & molecular biology* 2009;
498 **16**:265-273.

499 14 Zhang X, Wei M, Pan H *et al.* Robust manufacturing and comprehensive characterization of
500 recombinant hepatitis E virus-like particles in Hecolin(®). *Vaccine* 2014; **32**:4039-4050.

501 15 Medina RA, Stertz S, Manicassamy B *et al.* Glycosylations in the globular head of the
502 hemagglutinin protein modulate the virulence and antigenic properties of the H1N1 influenza
503 viruses. *Science translational medicine* 2013; **5**.

504 16 Liu W-CC, Jan J-TT, Huang Y-JJ, Chen T-HH, Wu S-CC. Unmasking stem-specific neutralizing
505 epitopes by abolishing N-linked glycosylation sites of influenza hemagglutinin proteins for vaccine
506 design. *Journal of virology* 2016.

507 17 Walls AC, Tortorici MA, Frenz B *et al.* Glycan shield and epitope masking of a coronavirus spike
508 protein observed by cryo-electron microscopy. *Nat Struct Mol Biol* 2016; **23**:899-905.

509 18 Yuan Y, Cao D, Zhang Y *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike
510 glycoproteins reveal the dynamic receptor binding domains. *Nature communications* 2017;
511 **8**:15092.

512 19 Shang J, Zheng Y, Yang Y *et al.* Cryo-EM structure of infectious bronchitis coronavirus spike
513 protein reveals structural and functional evolution of coronavirus spike proteins. *PLoS Pathog* 2018;
514 **14**:e1007009.

515 20 Shang J, Zheng Y, Yang Y *et al.* Cryo-Electron Microscopy Structure of Porcine Deltacoronavirus
516 Spike Protein in the Prefusion State. *J Virol* 2018; **92**.

517 21 Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein
518 in complex with its host cell receptor ACE2. *PLoS Pathog* 2018; **14**:e1007236.

519 22 Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-
520 CoV-2 by full-length human ACE2. *Science (New York, NY)* 2020.

521 23 Zhao S, Musa SS, Lin Q *et al.* Estimating the Unreported Number of Novel Coronavirus (2019-
522 nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the
523 Early Outbreak. *Journal of clinical medicine* 2020; **9**.

524 24 Palomares LA, Srivastava IK, Ramírez OT, Cox MMJMJ. Glycobiotechnology of the Insect Cell-
525 Baculovirus Expression System Technology. *Advances in biochemical engineering/biotechnology*
526 2018.

527 25 Wang SF, Tseng SP, Yen CH *et al.* Antibody-dependent SARS coronavirus infection is mediated
528 by antibodies against spike proteins. *Biochem Biophys Res Commun* 2014; **451**:208-214.

529 26 Wan Y, Shang J, Sun S *et al.* Molecular Mechanism for Antibody-Dependent Enhancement of
530 Coronavirus Entry. *J Virol* 2020; **94**.

531   27 Kam YW, Kien F, Roberts A *et al.* Antibodies against trimeric S glycoprotein protect hamsters
532   against SARS-CoV challenge despite their capacity to mediate FcgammaRII-dependent entry into
533   B cells in vitro. *Vaccine* 2007; **25**:729-740.

534   28 Dejnirattisai W, Supasa P, Wongwiwat W *et al.* Dengue virus sero-cross-reactivity drives
535   antibody-dependent enhancement of infection with zika virus. *Nature immunology* 2016;
536   **17**:1102-1108.

537   29 Rey FAA, Stiasny K, Vaney M-CC, Dellarole M, Heinz FX. The bright and the dark side of human
538   antibody responses to flaviviruses: lessons for vaccine design. *EMBO reports* 2018; **19**:206-224.

539   30 Ngwuta JO, Chen M, Modjarrad K *et al.* Prefusion F-specific antibodies determine the
540   magnitude of RSV neutralizing activity in human sera. *Science translational medicine* 2015; **7**.

541   31 Crank MC, Ruckwardt TJ, Chen M *et al.* A proof of concept for structure-based vaccine design
542   targeting RSV in humans. *Science* 2019; **365**:505-509.

543   32 McLellan JS, Chen M, Joyce MG *et al.* Structure-based design of a fusion glycoprotein vaccine
544   for respiratory syncytial virus. *Science (New York, NY)* 2013; **342**:592-598.

545   33 Li T, Zhang Z, Zhang Z *et al.* Characterization of native-like HIV-1 gp140 glycoprotein
546   expressed in insect cells. *Vaccine* 2019; **37**:1418-1427.

547   34 Li Z, Wang D, Gu Y *et al.* Crystal Structures of Two Immune Complexes Identify Determinants
548   for Viral Infectivity and Type-Specific Neutralization of Human Papillomavirus. *mBio* 2017; **8**.

549   35 Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic
550   correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods*
551   2017; **14**:331.

552   36 Zhang K. Gctf: Real-time CTF determination and correction. *Journal of structural biology* 2016;
553   **193**:1-12.

554   37 Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised
555   cryo-EM structure determination. *Nat Methods* 2017; **14**:290-296.

556   38 Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps.
557   *Nature methods* 2014; **11**:63.

558   39 Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera—a visualization system for
559   exploratory research and analysis. *Journal of computational chemistry* 2004; **25**:1605-1612.

560

561
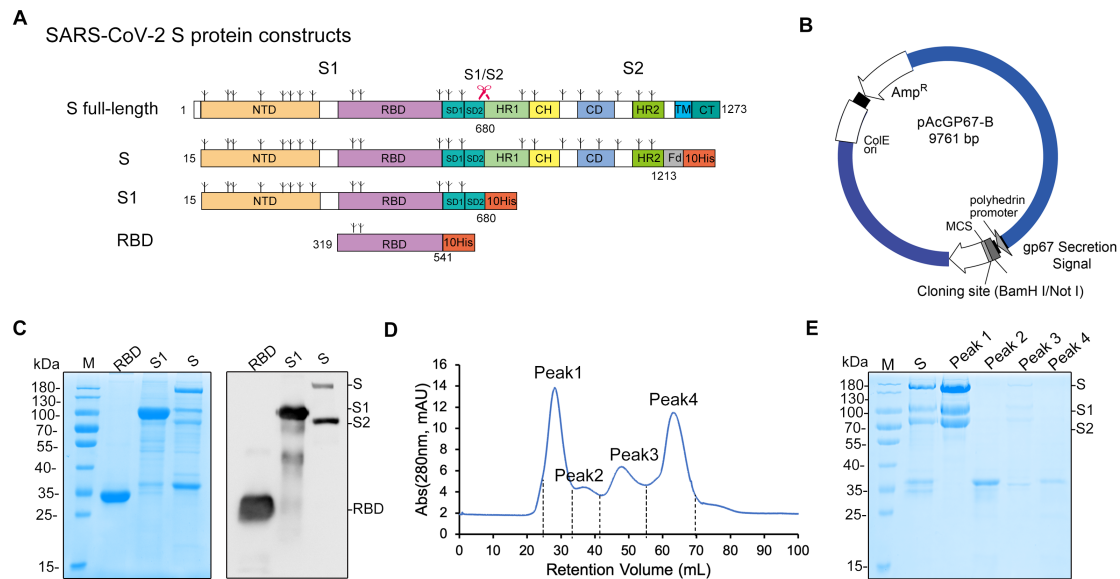
Figure 1. Schematic map of the SARS-CoV-2 S protein constructs. (A) Linear representations of the S protein primary structure and construct design. NTD, N-terminal domain; RBD, receptor binding domain; SD1, subdomain 1; SD2, subdomain 2; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail; FD, T4 foldon motif. The predicted glycosylation sites are indicated above the domain bars. (B) Map of the cloning vector pAcgp67B. The interest genes were cloned to plasmid pAcgp67B at BamH I/Not I site to generate transfer vectors. (C) SDS-PAGE and western blotting of the Ni-NTA purified proteins. RBS, S1 and S were eluted by 250 mM imidazole. Anti-His antibody was used as detection antibody in western blotting. (D) Size-exclusion chromatogram of the second-step purification of the S protein. (E) SDS-PAGE of the four fractions harvested from the chromatography purification as shown in (D).
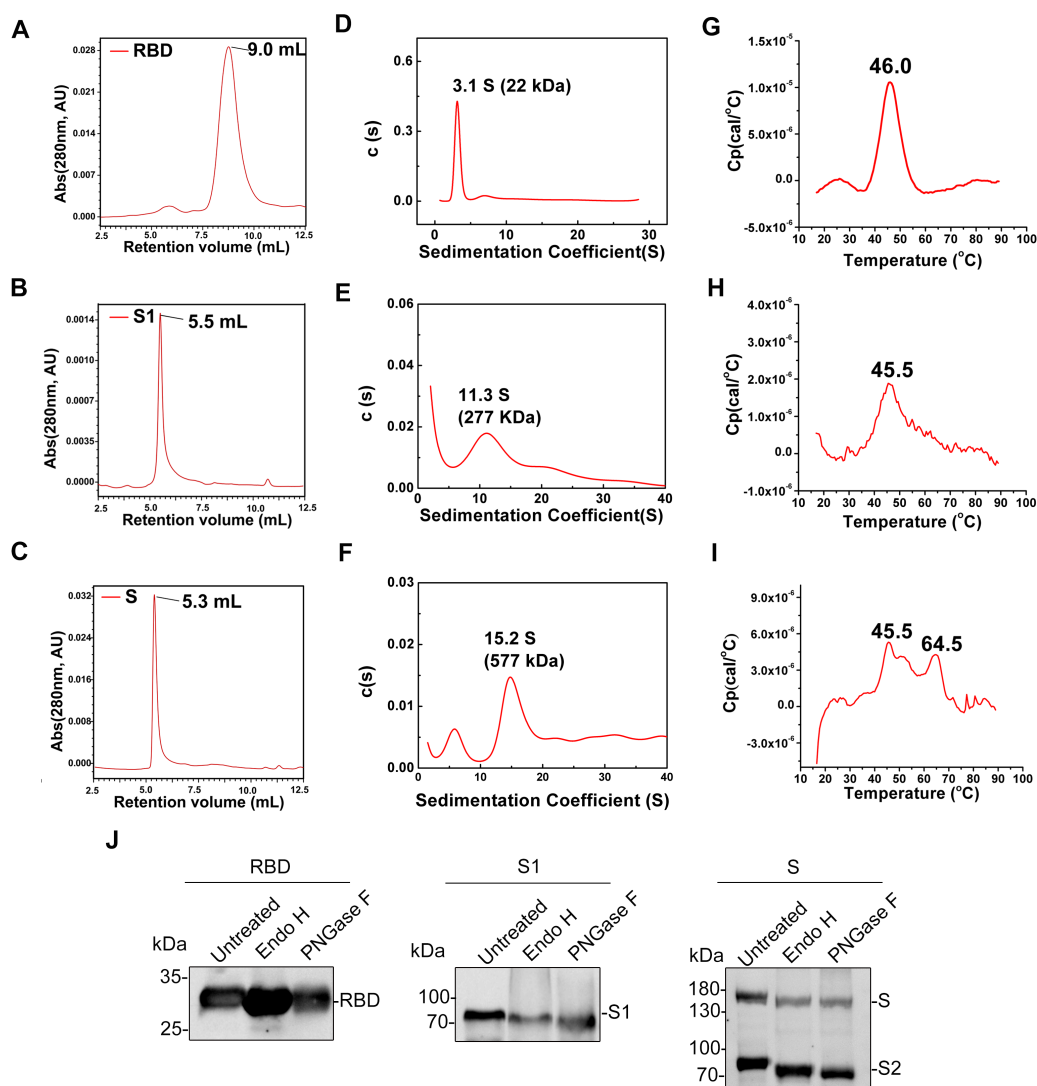
**Figure 2. Characterization of the purified RBD, S1 and S proteins.** (A-C) HPSEC profiles of the purified RBD, S1 and S proteins; (D-F) AUC profiles of RBD, S1 and S proteins; (G-I) DSC profiles of RBD, S1 and S proteins. (J) Western blotting of three purified proteins treated with Endo H and PNGase F or untreated as control. Anti-His antibody was used as detection antibody in western blotting.

582
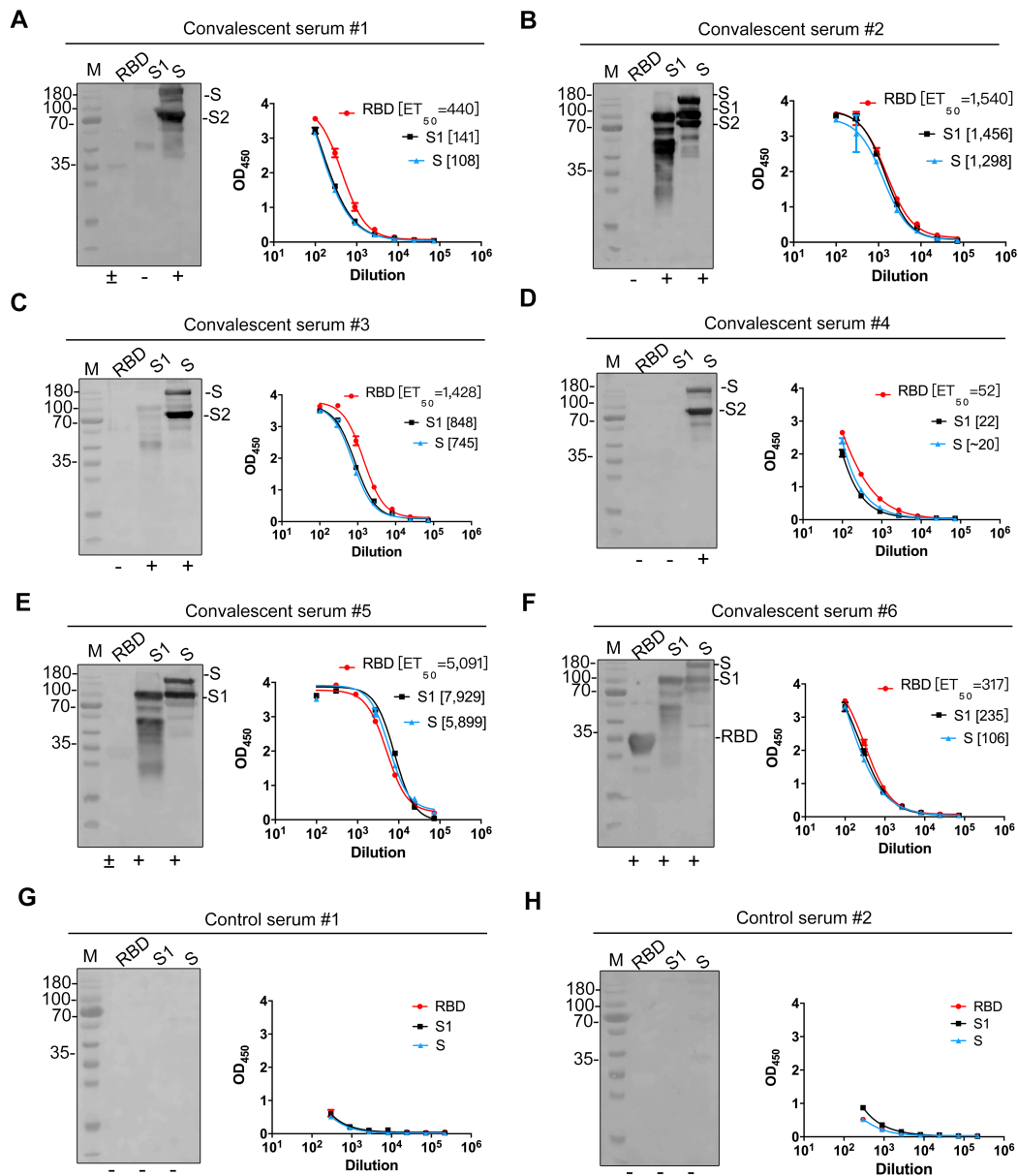
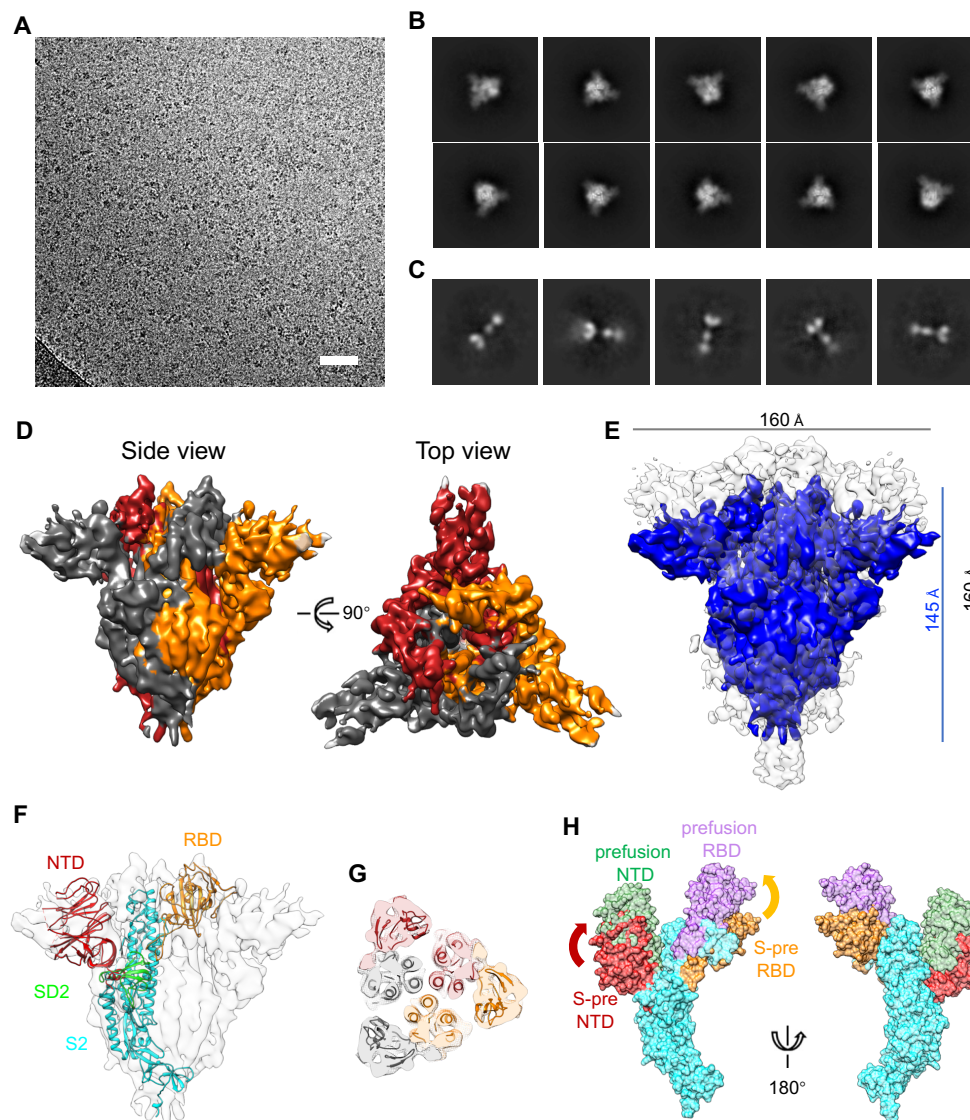**Figure 3. Antigenicity of RBD, S1 and S proteins against convalescent sera.** (A-F)

The reactivity of the RBD, S1 and S proteins against six COVID-19 convalescent

human sera (#1-#6) by western blotting (left panel) and ELISA (Right panel). (G, H)

Results of two control sera. The gels used for western blotting were duplicates of the

reducing SDS gels same as Fig. 1C.

588

589

Figure 4. Cryo-EM structure of the SARS-CoV-2 S trimer. (A) Representative

micrograph of frozen-hydrated SARS-CoV-2 S particles, Scale bar 50nm. (B, C) Ten and five

selected class averages showing the particles along different orientations belonging to prefusion (B)

and postfusion (C) S protein, respectively. (D) 5.43 Å density map of prefusion S trimer (S-pre) that

is colored by protomer. (E) Structural comparison to the reported prefusion SARS-CoV-2 S trimer

(EMD-21374, C3 symmetry, low-pass to 5.43Å) show a different conformation of S-pre (~15Å

shorter in height). (F, G) Each domain of the model of prefusion SARS-CoV-2 S monomer (F) or

trimer (G) (PDB no. 6VSB) were separately fitted in the density map of S-pre. (H) Schematic

598   diagram shows conformational diversities between NTD and RBD of S-pre (NTD: red, RBD:

599   orange) and reported prefusion S (NTD: light green, RBD: light purple).

600
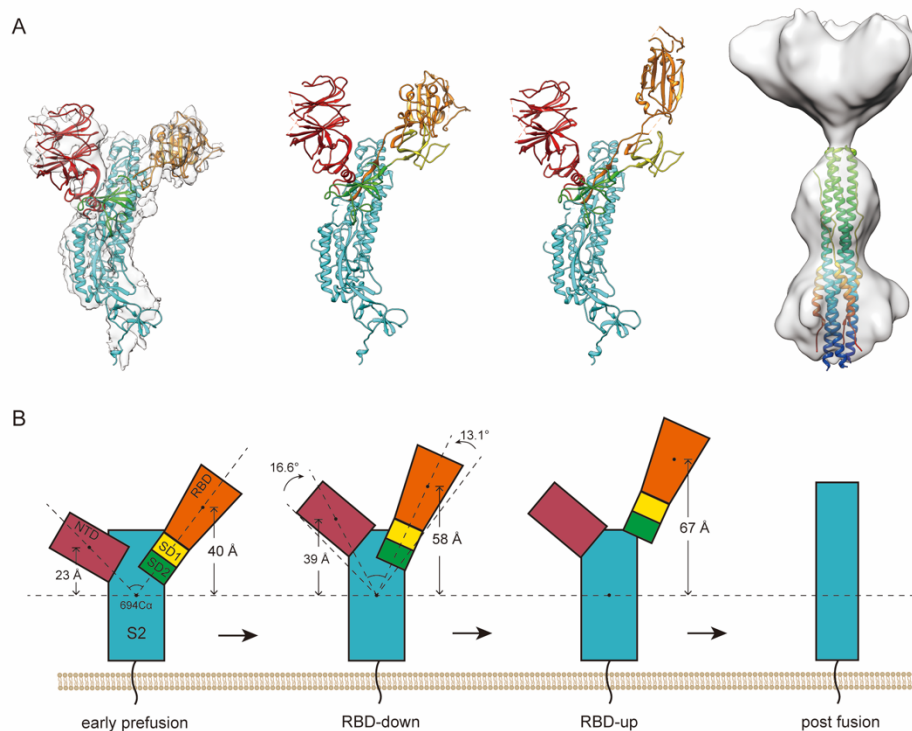
Figure 5. SARS-CoV-2 S spike at early pre-, pre- and post-fusion states and the proposed timeline for conformation change. (A) Four conformations of SARS-CoV-2 S spike. For clarity, only one monomer at early prefusion, RBD down prefusion and RBD up prefusion conformation is shown in ribbon mode. The model fitted in our S-pre map is a combined model comprised of individually fitted NTD, RBD, SD2 and S2 domain, which were evicted from the model of prefusion S trimer (PDB no. 6VSB). A deposited postfusion core of SARS-CoV-2 S2 subunit (PDB no. 6LXT) is fitted to our S-post map. Domains are designated by different colors, red: NTD, orange: RBD, yellow: SD1, green: SD2 and cyan: S2. (B) Simplified schematic diagram of the S monomer interpreting the conformation change across different states.

614 **Supplementary Information**

615 **Supplementary Figure 1**

616 **Supplementary Figure 2**

617 **Supplementary Figure 3**

618 **Supplementary Figure 4**

619 **Supplementary Movie 1**