What is a furin cleavage site, why is it important, and how might this have arisen in SARS-CoV-2?

Neil L Harrison[1], Jeffrey D Sachs[2]

[1]Columbia University, Department of Anesthesiology, New York, NY, USA
[2]Columbia University, Sustainability Institute, New York, NY, USA

*Corresponding authors: sachs@ei.columbia.edu, nh2298@columbia.edu

Keywords
SARS-CoV-2, furin cleavage site, furin

Summary
There is no consensus among scientists on the origin of SARS-CoV-2. One aspect of the virus that has been much discussed is the so-called “furin cleavage site” (FCS). Here we explain the structure and function of the FCS and its significance in SARS-CoV-2. The existing data suggest that the FCS of SARS-CoV-2, which remains unique among the hundreds of *sarbecoviruses* sampled from bats around the world, is fully functional and is consistent with the properties of FCS in many other substrates of this protease. Three possible routes have been proposed for how the FCS appeared in SARS-CoV-2: natural recombination, serial passage in cell culture or in an animal host and laboratory insertion via gene engineering. Here we review the merits and limitations of each proposal. All three proposals are limited by the absence to date of an immediate precursor virus. We renew our call that virus databases, lab notebooks and electronic communications be made available for independent scrutiny as part of a bipartisan investigation into the origins of COVID-19.

34 The “furin cleavage site” (FCS), is an important feature of the Spike protein
35 of the SARS-CoV-2 virus (**Figure 1a**). The FCS is present in the novel virus
36 SARS-CoV-2, but is absent in SARS-CoV-1 (**Figure 1b**) and in all other
37 members of the group of most closely related viruses, the subgenus
38 sarbecovirus (**Figure 1c**) (Coutard et al., 2020; Örd et al., 2020; Hoffmann
39 et al., 2020b; Temmam et al., 2023), which now includes several hundred
40 viruses, almost all of which were found in bats around the world.
41
42 A great deal has been written and said about the FCS in scientific journals,
43 in mainstream news outlets and on social media, yet there is still
44 considerable confusion about the structure, function, and importance of the
45 FCS in SARS-CoV-2. We aim to clarify key issues here.
46
47 The FCS is the site at the S1-S2 junction where the Spike protein of the
48 virus is cut by *furin,* an enzyme that is expressed in most human cells
49 (Thomas, 2002). This process is known as proteolysis. Processing of the
50 Spike protein by a combination of the two enzymes, furin and TMPRSS2
51 (Hoffmann et al., 2020a, Ou et al., 2021) (**Figure 2a**) is critical to the entry
52 of SARS-CoV-2 into cells in the lower respiratory tract (Coutard et al.,
53 2020) to infect the human lung, as well as for the transmission and
54 pathogenicity of the virus (Johnson et al., 2021, Peacock et al., 2021a).
55
56 The importance of furin cleavage for viral entry has been known since the
57 earliest work on *Sindbis* virus (Klimstra et al., 1999) but this is not a
58 universal feature for all viruses. For example, the replication of respiratory
59 syncytial virus (RSV) does not require furin (Zimmer et al., 2002). We
60 thought it a good idea to outline here some of the basic biology of furin and
61 the FCS, and to explain why this become so important to the study of how
62 the virus enters cells, as well as to discussions of the origins of the virus.
63
64 **What is furin?**
65 Furin is an enzyme that cuts proteins, i.e. it is a *protease*. To be more
66 specific, it is a member of a group called *proprotein covertases*, which is to
67 say it is an enzyme that is responsible for processing larger proteins into
68 their active final form (Seidah et al., 1998; Seidah & Prat, 2012).
69 Processing of larger precursors is especially important in endocrinology,
70 with many hormones such as insulin being derived by proteolysis of larger
71 protein precursors. Furin itself is produced by auto-proteolysis from a
72 precursor protein (Thomas, 2012).
73

74 Furin was first identified and originally designated as PACE, and the gene
75 was cloned in 1990 (Bresnahan et al., 1990). Unlike other proteases such
76 as TMPRSS2 that are found at the cell surface (Meng et al., 2020), furin is
77 mainly located inside the cell, specifically inside membrane-bound
78 organelles (**Figure 2b**), being synthesized on the endoplasmic reticulum
79 and then is modified in the presence of Ca^{2+}, moving through the ER-Golgi
80 intermediate complex (ER-GIC) to the trans-Golgi network (TGN) where the
81 mature form of the protein is stabilized at low pH (Thomas, 2002).
82
83 Once in the TGN, furin modifies proteins that have already been
84 synthesized, including those being packaged into vesicles for export to the
85 plasma membrane (**Figure 2b**). In the process, furin itself appears on the
86 cell surface, from where it is then recycled into endocytic vesicles (Molloy
87 et al., 1994). Viruses, including coronaviruses, make use of this cellular
88 sorting machinery to facilitate the manufacture of membrane-bound protein,
89 forming new virus particles. The Spike protein of SARS-CoV-2 is one
90 example, and it can be cleaved by furin inside the infected cell before being
91 packaged into newly synthesized virus containing other viral components.
92 These are essential steps in the process by which new virions exit the cell.
93
94 **What exactly is a furin cleavage site?**
95 Furin works by “cutting” the peptide strand of a precursor protein at one or
96 more locations, specifically by *catalyzing the hydrolysis of a specific*
97 *peptide bond*, typically between an arginine residue and its immediate
98 neighbour, which is variable but is typically a smaller (serine or valine)
99 residue. The hydrolysis of the peptide bond at R🡻X is most efficient when
100 a specific *furin recognition sequence* is present, which is enriched in basic
101 amino acid residues (R/K). The typical furin cleavage site contains at least
102 2 basic residues, separated by 2 amino acids, **RXXR** (this is RRAR in
103 SARS-CoV-2; **Figure 1a**), but there are exceptions to this rule. This
104 minimal consensus recognition sequence is what is usually referred to as a
105 *“furin cleavage site”,* although the proteolysis “cleavage” actually takes
106 place very specifically at what is termed the “scissile bond” (R🡻X), at the
107 C-terminal end of the RXXR recognition sequence.
108
109 Note that proteins can also be susceptible to proteolysis by other enzymes
110 found outside cells, such as trypsin and cathepsin B etc., and that the
111 importance of the specific furin recognition sequence is to enhance the
112 efficiency of proteolysis by one specific protease, furin, which is expressed
113 in most cells. There are examples of proteins that contain multiple FCS, so

114 that the protein is synthesized in its inactive form, and can then be
115 activated by the removal of an intervening auto-inhibitory domain, as
116 occurs with the epithelial sodium channel, ENaC (Kota et al., 2018) and in
117 the case of furin itself (Thomas, 2002). It is common for viral Spike proteins
118 to contain more than one FCS (Millet et al., 2014).
119
120 **What characterizes a furin cleavage site?**
121 Like all enzymes, furin has a well-characterized *catalytic domain*, which is
122 defined as the part of the enzyme that makes contact with the protein
123 substrate. This is where the chemical reaction "hydrolysis" occurs (Henrich
124 et al., 2003) and this is located within a hydrophilic region of the protein.
125 The structure of furin has been solved and the catalytic domain lies within a
126 substrate-exposed part of the furin structure. The substrate protein binds to
127 furin in a manner that activates the enzyme (Dahms et al., 2016), in an
128 example of "induced fit", interacting not just via the short furin recognition
129 sequence but over a more extended surface. The furin interaction domain
130 of the substrate is recognized to be part of an extended FCS strand that is
131 typically of around 20 amino acids in length (Tian, 2009), that stabilizes the
132 cleavage site by making close contacts within the catalytic domain of furin.
133
134 An extensive analysis of over a hundred proteins that are cleaved by furin
135 has shown that the furin recognition sequence has certain conserved
136 characteristics. The full-length furin cleavage site motif is typically
137 comprised of about 20 residues, here annotated P14-P6′ (Tian, 2009)
138 (**Figure 3a**). The FCS residues are numbered relative to the site (the
139 scissile bond) where the polypeptide is cut, and thus the arginine at 685 in
140 the SARS-CoV-2 spike protein is designated as the "P1" position (**Figure**
141 **3b**), with the serine being the "P′1 position". The other arginines in the
142 RRAR sequence are thus in the "P3" and "P4" positions, and in the case of
143 SARS-CoV-2, a proline occupies the "P5" position, a feature that has been
144 noted by several commentators (Holmes et al., 2021, Garry, 2022).
145
146 Specific physical properties such as volume, charge, and hydrophilicity are
147 required at specific positions in order to optimize the cleavage of substrate.
148 The furin cleavage site motif can be divided into two parts: a core region of
149 8 amino acids, (positions P6-P2′) packed inside the furin binding pocket,
150 and two flanking regions (**Figure 3a**) that are both solvent-accessible and
151 located outside the furin binding pocket – one of 8 polar amino acids,
152 (positions P7–P14), and another of 4 small amino acids, (positions P3′-
153 P6′). In the case of the SARS-CoV-2 Spike protein, some of the interactions

154 with the catalytic domain of furin are thought to be made in the core, with
155 the flanking regions providing stabilizing interactions (Venkadari, 2020).
156 More recent work has shown that the flanking regions can also be very
157 important, as in the case of the QTQTN motif (P7-P11) of the SARS-CoV-2
158 Spike (**Figure 3b**), which has been shown to be important for SARS-CoV-2
159 pathogenesis (Vu et al., 2022) as well as being highly antigenic (Haynes et
160 al., 2021).
161
162 The S1/S2 junction of the SARS-CoV-1 Spike lacks the FCS but forms a
163 short solvent-exposed loop (**Figure 4a**) and this loop is further extended in
164 SARS-CoV-2 (**Figure 4a, 4b**) due to the PRRA insert (Jaimes et al., 2020),
165 exposing the FCS for proteolysis. The intricate details of the SARS-CoV-2
166 FCS are only truly revealed in an elegant study of its interaction with the
167 catalytic domain of furin (**Figure 5a**) (Venkadari, 2020). The basic residues
168 within the polybasic RXXR sequence of a typical FCS make electrostatic
169 contacts with negatively charged residues in the catalytic domain of furin
170 (Venkadari, 2020) (**Figure 5b**), while in the case of SARS-CoV-2, the “P5”
171 residue is modeled with the proline side chain oriented away from the
172 catalytic domain (**Figure 5c)** so that there is no steric or electrostatic
173 hindrance. In fact, a variety of smaller amino acids can be tolerated at this
174 P5 position of FCS (Tian, 2009); there is thus no thermodynamic or steric
175 “prohibition” against a proline residue being located adjacent to the core
176 recognition sequence.
177
178 It follows from this survey of many FCS domains that a proline at the P5
179 position is neither unexpected nor unusual, contrary to some commentary
180 (Garry, 2022). Indeed, this point is emphasized by the fact that the Spike of
181 the MERS virus also has a proline at this P5 position (Millet et al., 2014,
182 Garry, 2022), although opinions may differ on whether the FCS of MERS is
183 itself fully functional (Millet et al., 2014).
184

185 **The furin cleavage site of SARS-CoV-2 is fully functional**

186 The argument has been advanced that the FCS in SARS-CoV-2 is “sub-
187 optimal”. This claim seems to be based on computer algorithms that predict
188 the functionality of FCS sequences. Two such algorithms are now in
189 widespread use (e.g. Duckert et al., 2004). Yet these algorithms are known
190 to give “false negatives”, i.e. the failure to predict a fully functional FCS.
191
192 One example of a predicted “sub-optimal” FCS (Holmes et al., 2021) is the
193 RRAR🡻SVAS sequence of SARS-CoV-2 itself. Although this FCS scores

lower than others according to prediction algorithms, it is fully and efficiently cleaved when cells are infected with SARS-CoV-2 (Hoffmann et al., 2020b, **Figure 6a**) or with pseudoviruses expressing the SARS-CoV-2 Spike protein (Walls et al., 2020; Peacock et al., 2021) (**Figure 6b**), in contrast to the SARS-CoV-1 Spike (**Figure 6c**) and this efficient proteolysis is abolished by deletion of the PRRA sequence (**Figure 6c**). It is worth noting that an identical FCS core sequence (RRAR🡻SVAS) is found in the α subunit of human ENaC (Anand et al., 2020) and that the ENaC α protein from mouse and rat is also completely and efficiently cleaved by furin in epithelial cells (Hughey et al., 2004, Kota et al., 2018).

Two algorithms predict that the FCS of SARS-CoV-2 is not "ideal", and this led several scientists to speculate that the FCS was "sub-optimal" (Holmes et al., 2021). Of course, the efficient proteolysis described above (**Figure 6a-d**) and the high human-to-human transmissibility of SARS-CoV-2 (R_0 ~2-3 for the original Wuhan-1 virus) would seem to argue otherwise.

Because of the assertions that the FCS is "sub-optimal", one might expect to find viral variants that show FCS mutations enhanced proteolysis. In fact, the RRAR (P4-P1) core sequence has remained remarkably stable, with very few mutations reported (Wolf et al., 2022; Cassari et al., 2023), suggesting that such mutations confer no evolutionary advantage. A variety of natural polymorphisms (point mutations) within the extended FCS region have now been tested in virology labs (Arora et al., 2022), and many of these mutations actually *resulted in a modest loss of proteolysis efficiency* relative to the original FCS (**Figure 6d**). While it has been suggested that the efficiency of proteolysis of the SARS-CoV-2 Spike protein would be further enhanced by mutation at the P5 residue, including the naturally observed mutations P681R and P681H (Peacock et al., 2021b), it is now clear that the experimental data do not in fact support this. In addition, it has been shown that the FCS of the original SARS-CoV-2 Spike protein can be imported into the SARS-CoV-1 Spike by engineering and that the resulting mutant Spike is then fully and efficiently cleaved by furin, in cells experimentally infected via a pseudovirus (Winstone et al., 2021). There is thus no convincing evidence that the FCS of SARS-CoV-2 is sub-optimal.

Insights from Structural Biology: studies of the Spike protein

Elegant work done following the emergence of SARS-CoV-1 had shown that this virus uses the human membrane protein angiotensin converting enzyme-2 (ACE2) as its primary receptor on human cells (Li et al, 2005).

234 Studies of SARS-CoV-2 confirmed that the newer virus also uses human
235 ACE2 as its primary receptor on susceptible cells (Shang et al., 2020;
236 Hoffmann et al., 2020a) and that the Spike protein forms a trimeric
237 assembly that binds with high affinity to its receptor (Walls et al., 2020).
238 The trimer exists in multiple conformational sates and binds to hACE2 with
239 one S^B domain of the Spike protein in the open conformation (Wrapp et al.,
240 2020 Walls et al., 2020). In one especially insightful experiment (Wrobel et
241 al., 2020), a comparison was made between the Spike proteins of SARS-
242 CoV-2 and a close relative termed RaTG13, which is known to infect bats
243 but is not thought to infect humans. Firstly, it was found that there were
244 small but significant differences between the structures of the receptor
245 binding domains (RBD) of the Spike proteins of these two viruses (**Figure
246 7a**). Secondly, it was noted that the Spike of RaTG13 binds very weakly to
247 hACE2, with as much as 1000 times lower affinity (**Figure 7b**), largely due
248 to steric constraints in the interaction between the RBD of the Spike and
249 hACE2 (**Figure 7b**). Crucially, these and other authors noted that cleavage
250 by furin destabilized the "closed" conformation of the Spike protein trimer
251 and hence promoted the "open" conformation, exposing the RBD that is
252 necessary for the binding of the Spike to hACE2. The lack of FCS in the
253 RaTG13 limits the Spike cleavage by proteolysis and stabilized the Spike in
254 the closed conformation associated with low affinity binding to hACE2
255 (Wrobel et al, 2020).
256
257 **The importance of FCS for viral transmission**
258 The importance of FCS for viral entry and transmissibility has been known
259 for some time. Extensive work has been performed in laboratories around
260 the world to insert FCS via genetic engineering into pathogenic viruses,
261 including influenza viruses (Schrauwen et al., 2011) and coronaviruses
262 such as SARS-CoV-1 (Millet et al., 2015) and the porcine epidemic
263 diarrhea virus, PEDV (Li et al., 2015). It should be pointed out that the vast
264 majority of this work was done in the context of *pseudovirus* experiments,
265 using an innocuous virus as a backbone, so that the chance of producing a
266 highly pathogenic virus with increased transmissibility escaping from a
267 laboratory was minimized (Belouzard et al., 2009). Work done in this format
268 is therefore considered safe and does not constitute what is known as
269 "Gain-of-Function" (GoF) research. In at least one case (the example given
270 above of PEDV), this type of work created a replication-competent novel
271 recombinant virus (Li et al., 2015), showing that engineering of functional
272 infectious viruses in this region is neither implausible nor novel.
273
274

275 **Research on the FCS in SARS-like viruses since SARS-CoV-1**
276 There has obviously been a long-standing interest among coronavirus
277 researchers in the role of spike protein proteolysis in enhancing viral entry.
278 Since the original epidemic of SARS and the identification of SARS-CoV-1
279 as a coronavirus that uses ACE2 as its receptor on epithelial cells, much
280 attention has been devoted to the study of the SARS-CoV-1 Spike protein
281 (Li et al., 2005), and FCS has been inserted into the spike protein of this
282 virus on multiple occasions (Belouzard et al., 2009; Millet et al., 2015).
283
284 Work on proteolysis as a determinant of viral entry was recently extended
285 to a range of viruses that include bat coronaviruses from the *sarbecovirus*
286 group (Menachery et al., 2020) some of which have been suggested to
287 show the potential for emergence (Menachery et al., 2015). Such work is
288 sometimes done under conditions of limited containment (BSL-2) because
289 of the perception that these bat *sarbecoviruses* lack pathogenic potential.
290 Work done on these bat viruses has proliferated, and is widely considered
291 to be a "grey area" that constitutes Gain-of-Function research of concern
292 (GoFRoC), since chimeric viruses of unknown function are created and the
293 pathogenicity of the resulting virus cannot be predicted. Laboratory
294 experiments using pseudoviruses showed that infection of lung cells by
295 SARS-CoV-2 and transmissibility between ferrets is strongly inhibited by
296 removal of the PRRA sequence (Peacock et al., 2021), and complementary
297 work in pseudovirus experiments confirmed that insertion of the PRRA
298 sequence into the Spike protein of SARS-CoV-1 confers high furin
299 sensitivity (Winstone et al., 2021) and enhances viral entry into cells.
300
301 **The possible origins of the FCS in SARS-CoV-2**
302 A combination of efficient human-to-human transmission with significant
303 pathogenicity is one of the hallmarks of a pathogen with pandemic potential
304 (PPP). In fact, the combination of factors that made SARS-CoV-2 a
305 pandemic virus involves a combination of its high affinity for human ACE2,
306 its processing by furin and TMPRSS2 (Essalmani et al., 2022) as well as its
307 ability to down-regulate the "innate immune response" in humans that is
308 mediated by interferons (Winstone et al., 2021), perhaps via one or more of
309 the "accessory" proteins of the virus, encoded by 3'-open reading frames
310 (Orf). All of these features of the virus may have arisen naturally, but the
311 unique nature of the FCS in SARS-CoV-2 among the SARS-related bat
312 viruses of the *sarbecovirus* clade (Coutard et al., 2020, Hoffmann et al.,
313 2020b) is quite remarkable (**Figure 1c**) and has given rise to speculation
314 about a possible anthropogenic origin (Chan and Zhan, 2022). Intensive

study on bats from around the world (Temmam et al., 2022, Sander et al., 2022; Apaa et al., 2023) has led to the identification of hundreds of novel sarbecoviruses, but not a single virus other than SARS-CoV-2 has been shown to have an FCS.

FCS sequences do, of course, exist in many of the common coronaviruses that infect humans and other animals (Wu and Zhao, 2021) including the endemic “common cold” viruses, OC43 and HKU-1, which are not pathogenic in healthy, immunocompetent individuals. The FCS may contribute to the high transmissibility of these respiratory viruses. In contrast, the sarbecoviruses are primarily *enteric* viruses in the bat and therefore remain confined to a limited host range. These viruses (BANAL-20-236 for example) efficiently infect human intestinal epithelial cells but do not infect cells of the mammalian respiratory tract (Temmam et al., 2023). The pathogenic viruses such as SARS-CoV-1 and SARS-CoV-2 are obviously an important exception to this rule. In this context, it is worth noting again that the SARS-CoV-1 virus, although highly pathogenic and capable of human-to-human transmission, lacks the FCS and this feature may have ultimately limited its pandemic potential. Experimental work undertaken to investigate the significance of the FCS for pathogenesis has included multiple examples of insertion of FCS-type sequences into SARS-CoV-1 spike, in the context of a pseudovirus (Watanabe et al., 2008; Belouzard et al., 2009, Winstone et al., 2021). For this reason, there is little question concerning the technical feasibility of such an insertion.

Did the FCS of SARS-CoV-2 Evolve Naturally via Recombination?

Sequence alignments suggest that the possibility that FCS can evolve in sarbecoviruses via a series of individual point mutations is low. Most of the viruses are insufficiently similar to permit a convincing alignment in the S1/S2 region (Holmes et al., 2021; Sander et al., 2022). Only the most highly similar viruses like RaTG13 and BANAL-20-52 provide the opportunity to align the amino acid or RNA sequences with SARS-CoV-2 in this region.

Well-articulated (but as yet unproven) arguments have been advanced that invoke processes of natural recombination in the acquisition of the FCS by SARS-CoV-2. Most notably, proposals from evolutionary biologists (Sander et al., 2022) and experimental virologists (Gallaher, 2020) invoke the process of “copy-choice” recombination. Such proposals are plausible but are not at this point supported by experimental evidence. Among the main

criticisms of these proposals has been the lack of a natural virus that is a sufficiently good match to the SARS-CoV-2 RNA sequence to have served as the immediate precursor for the proposed recombination event (Chan and Zhao, 2022). Even the closely related virus RaTG13, or the viruses identified in bats from Laos are not sufficiently similar at the RNA level to have served as the immediate ancestor to SARS-CoV-2 (Gallaher, 2020; Temmam et al., 2022). Less well-articulated arguments in favor of a natural origin of the FCS have been advanced (Garry, 2022) and clearly refuted (Harrison and Sachs, 2022b).

Did the FCS of SARS-CoV-2 Arise During Serial Passage?

An alternate explanation for the presence of the FCS in SARS-CoV-2 and its apparent adaptation to human cells, is that it might have arisen during the passage of a precursor virus in the laboratory, either during work done in human cells grown in cell culture or during serial passage in animal models such as humanized mice. Neither of these possibilities can be ruled out, but some evidence has accumulated to suggest that this may not have taken place. A bovine coronavirus was reported to have acquired a 12nt insert encoding the four amino acids SRRR during passage in human cells (Borucki et al., 2013), but on closer inspection of the data it emerged that this was not the case, as the variant carrying the insert was already present and had been selected for during passage, presumably because it conferred an advantage to the virus when grown in human cells.

An interesting and more recent study of serial passage looked at the bat virus BANAL-20-236, a *sarbecovirus* that was sampled from bats in Laos (Temmam et al., 2022) and is one of the closest known relatives of SARS-CoV-2. The sequence of the Spike protein around the S1/S2 junction is a close (but inexact) match to the sequence present in SARS-CoV-2. BANAL-20-236 is not able to infect human airway epithelial cells, but by growing this virus in human intestinal cells in culture, it was possible to propagate the virus and to look for evidence of adaptation to human cells. Although evidence was found for point mutations in the RBD during passage in cell culture, a FCS did not emerge from these experiments. Serial passage experiments performed with BANAL-20-236 in humanized mice produced similar results (Temmam et al., 2023). These results do not support the popular theory that the FCS arose during serial passage in culture, but are consistent with ideas that have been proposed regarding the adaptation of the RBD in a laboratory setting (Sirotkin & Sirotkin, 2020).

395 **Was the FCS of SARS-CoV-2 Engineered?**
396 It is clear from the DEFUSE grant proposal released in 2021 that the UNC-
397 WIV-EHA group was planning to extend their ongoing studies on
398 proteolysis (Menachery et al., 2020) as a factor in controlling viral entry,
399 and that they aimed to do so by identifying novel FCS sequences and
400 inserting these into newly identified coronaviruses (Lerner, 2021), and this
401 proposal might obviously include work on unreported viruses that have not
402 been disclosed. As with the arguments for natural origin, the main criticism
403 of the idea that lab manipulation was involved is that the precursor virus
404 (the "template") necessary for such engineering experiments has not been
405 identified. It is worth noting here that BANAL-20-52, BANAL-20-236 and
406 RaTG13, although the closest relatives to SARS-CoV-2*, are simply too*
407 *different at the nucleotide level* to have served as the precursor for either
408 route of origin (**Figure 8a**).
409
410 The experiments proposed by EHA and their partners in DEFUSE would
411 seem to represent a logical extension of the work of many virologists, in
412 particular the work done on proteolysis as a factor determining host range,
413 and their long-standing interest in the FCS as a critical determinant of viral
414 entry. A lack of transparency regarding this research has amplified
415 concerns that the FCS of the virus might have a laboratory origin (Segreto
416 and Deigin, 2020; Chan and Zhao, 2022). Insertion of the FCS by
417 engineering is technically very simple to achieve and has already been
418 performed many times, for example, with SARS-CoV-1 in pseudovirus
419 experiments (Watanabe et al., 2008; Belouazard et al., 2009; Winstone et
420 al., 2021).
421
422 An unusual *BsaX* I restriction site is found in SARS-CoV-2, bracketing the
423 P2-P12 residues, and contained within the extended FCS (**Figure 8b**). This
424 interesting observation, made by many observers, is consistent with the
425 idea that the FCS could have been inserted in a lab. In fact, the much-
426 maligned proline (P681) found at the P5 position, adjacent to RRAR, is
427 both consistent with, and obligatory for, the insertion of a *BsaX* I site at this
428 position. This unusual restriction site then provides for a potential
429 application of what is termed "Golden mutagenesis", in which *any ten*
430 *nucleotides* can be inserted 3'- to the CUCC sequence, resulting in *any*
431 *three amino acids* being inserted between P5 (P681) and the conserved P1
432 arginine (R685). Golden mutagenesis is one application of "Golden Gate"
433 cloning (Engler and Marillonet, 2014) using "type IIs" restriction enzymes
434 (REs), a group that includes not only *BsaX* I (Tengs et al., 2004), but also

Bsa I and *BsmB* I, which have been extensively used in the design and recovery of infectious clones of coronaviruses (Hou et al., 2020).

The use and application of type IIs REs was predicted and discussed by the world's most prolific coronavirus engineer several years ago, in relation to potential biowarfare and bioterrorist activities (Baric, 2007). There is some concern within the broader scientific community that similar experiments might have led to the creation of a virus closely related to SARS-CoV-2, a possibility that was foreseen by experts in biosecurity many years earlier (Klotz and Sylvester, 2014). The summation of these and other concerns about coronavirus engineering has led to calls by us and others for the full disclosure of sequences, email communications and laboratory notebooks, all as part of a detailed inquiry into the origins of the virus (Relman, 2020, VanHelden et al., 2021; Harrison and Sachs, 2022a).

Author contributions: Both authors conceived the paper. NLH wrote the first draft and both authors edited the final version. We are grateful to an anonymous card-carrying virologist for helpful comments and advice.

Competing Interests: JDS is Chair of the Lancet COVID-19 Commission. The authors declare no competing financial or scientific interests.

Figures

Figure 1a

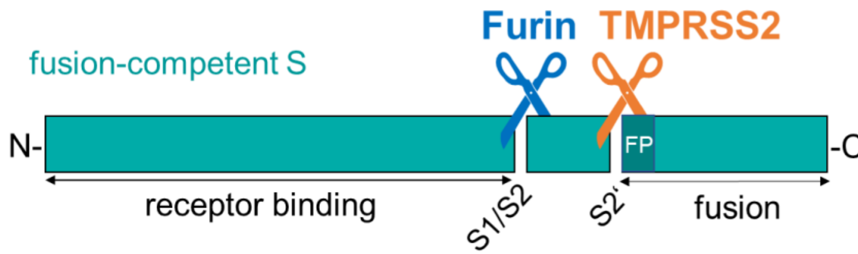


Figure 1b

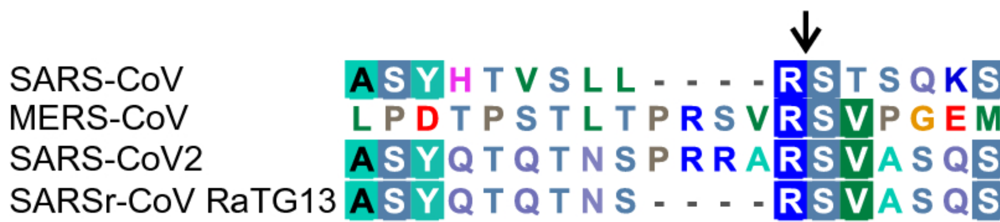
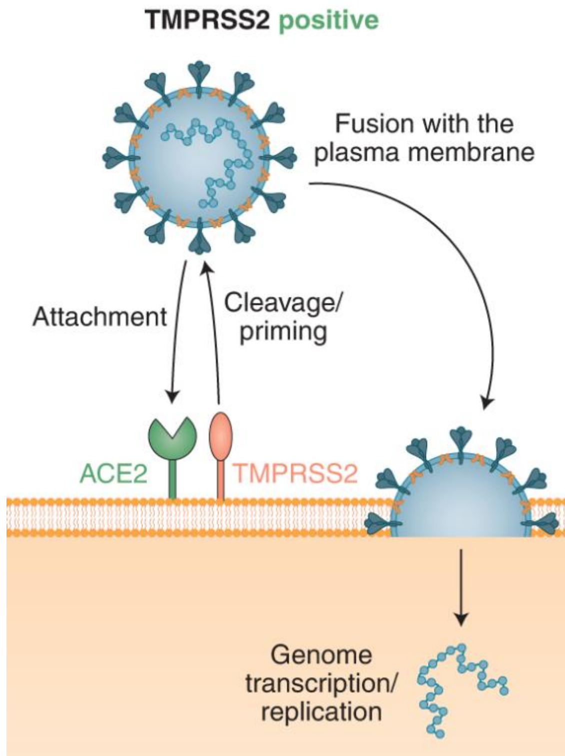


Figure 1c

	S1/S2	S2'
Human SARS-CoV BJ01	655 - GICASYHTVSLL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Human SARS-CoV CUHK-W1	655 - GICASYHTVSLL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Human SARS-CoV Tor2	655 - GICASYHTVSLL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Human SARS-CoV Frankfurt-1	655 - GICASYHTVSLL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Human SARS-CoV Urbani	655 - GICASYHTVSLL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Civet SARS-CoV civet020	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Civet SARS-CoV SZ16	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Raccoon dog SARS-CoV A030	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
SARS-CoV-2	669 - GICASYQTQTNSPRRARSVA - 688	808 - DPSKPSKRSFIED - 820
Pangolin CoV MP789	n/a - GICASYQTQTNS----RSVS - n/a	n/a - DPSKPSKRSFIED - n/a
Bat SARSr-CoV RaTG13	669 - GICASYQTQTNS----RSVA - 684	804 - DPSKPSKRSFIED - 816
Bat SARSr-CoV LYRa11	659 - GICASYHTASLL----RNTD - 674	794 - DPSKPTKRSFIED - 806
Bat SARSr-CoV LYRa3	659 - GICASYHTASLL----RNTG - 674	794 - DPSKPTKRSFIED - 806
Bat SARSr-CoV RsSHC014	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV Rs4084	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV WIV1	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV Rs3367	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV Rs7327	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV Rs9401	656 - GICASYHTVSSL----RSTS - 671	791 - DPLKPTKRSFIED - 803
Bat SARSr-CoV Rs4231	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Bat SARSr-CoV WIV16	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Bat SARSr-CoV Rs4874	655 - GICASYHTVSSL----RSTS - 670	790 - DPLKPTKRSFIED - 802
Bat SARSr-CoV ZC45	646 - GICASYHTASIL----RSTS - 661	781 - DPSKPSKRSFIED - 793
Bat SARSr-CoV ZXC21	645 - GICASYHTASIL----RSTG - 660	780 - DPSKPSKRSFIED - 792
Bat SARSr-CoV Rf4092	634 - GICASYHTASTL----RGVG - 649	769 - DPSKPTKRSFIED - 781
Bat SARSr-CoV Rf/JL2012	636 - GICASYHTASLL----RSTG - 651	771 - DPLKPTKRSFIED - 783
Bat SARSr-CoV JTMC15	636 - GICASYHTASLL----RSTG - 651	771 - DPLKPTKRSFIED - 783
Bat SARSr-CoV 16BO133	636 - GICASYHTASLL----RSTG - 651	771 - DPLKPTKRSFIED - 783
Bat SARSr-CoV B15-21	636 - GICASYHTASLL----RSTG - 651	771 - DPLKPTKRSFIED - 783
Bat SARSr-CoV YN2013	633 - GICASYHTASTL----RSIG - 648	768 - DPSKPTKRSFIED - 780
Bat SARSr-CoV Anlong-103	633 - GICASYHTASTL----RSVG - 648	768 - DPSKPTKRSFIED - 780
Bat SARSr-CoV Rp/Shaanxi2011	640 - GICASYHTASVL----RSTG - 655	775 - DPSKPTKRSFIED - 787
Bat SARSr-CoV Rs/HuB2013	641 - GICASYHTASVL----RSTG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV YNLF/34C	641 - GICASYHTASVL----RSTG - 656	776 - DPLKPTKRSFIED - 788
Bat SARSr-CoV YNLF/31C	641 - GICASYHTASVL----RSTG - 656	776 - DPLKPTKRSFIED - 788
Bat SARSr-CoV Rf1	641 - GICASYHTASHL----RSTG - 656	776 - DPLKPTKRSFIED - 788
Bat SARSr-CoV 273	641 - GICASYHTASHL----RSTG - 656	776 - DPLKPTKRSFIED - 788
Bat SARSr-CoV Rf/SX2013	639 - GICASYHTASLL----RSTG - 654	774 - DPLKPTKRSFIED - 786
Bat SARSr-CoV Rf/HeB2013	641 - GICASYHTASLL----RSTG - 656	776 - DPLKPTKRSFIED - 788
Bat SARSr-CoV Cp/Yunnan2011	641 - GICASYHTASLL----RNTG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rs672	641 - GICASYHTASTL----RSVG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rs4255	641 - GICASYHTASTL----RSVG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rs4081	641 - GICASYHTASTL----RSVG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rm1	641 - GICASYHTASVL----RSTG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV 279	641 - GICASYHTASVL----RSTG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rs/GX2013	642 - GICASYHTASVL----RSTG - 657	777 - DPSKPTKRSFIED - 789
Bat SARSr-CoV Rs806	641 - GICASYHTASLL----RSTG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV HKU3-1	642 - GICASYHTASVL----RSTG - 657	777 - DPSKPTKRSFIED - 789
Bat SARSr-CoV Longquan-140	642 - GICASYHTASVL----RSTG - 657	777 - DPSKPTKRSFIED - 789
Bat SARSr-CoV Rp3	641 - GICASYHTASTL----RSVG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV Rs4247	642 - GICASYHTASTL----RSVG - 657	777 - DPSKPTKRSFIED - 789
Bat SARSr-CoV Rs4237	641 - GICASYHTASTL----RSVG - 656	776 - DPSKPTKRSFIED - 788
Bat SARSr-CoV As6526	641 - GICASYHTASTL----RSVG - 656	777 - DPSKPTKRSFIED - 789
Bat SARSr-CoV BtKY72/KEN	660 - GICAKFGS---D---KIRMG - 673	793 - DPKKLSYRSFIED - 805
Bat SARSr-CoV BM48-31	658 - GICAKYTNVSST---LVRSG - 674	794 - DPAKPSSRSFIED - 806
	****.:	** * : ******

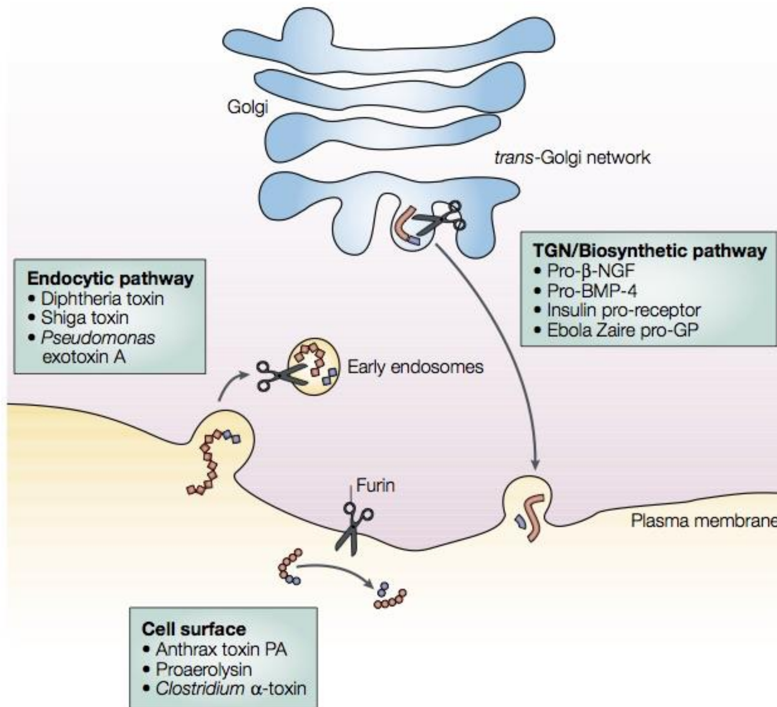
464

465 **Figure 2a**



466

467 **Figure 2b**



468

Figure 3a

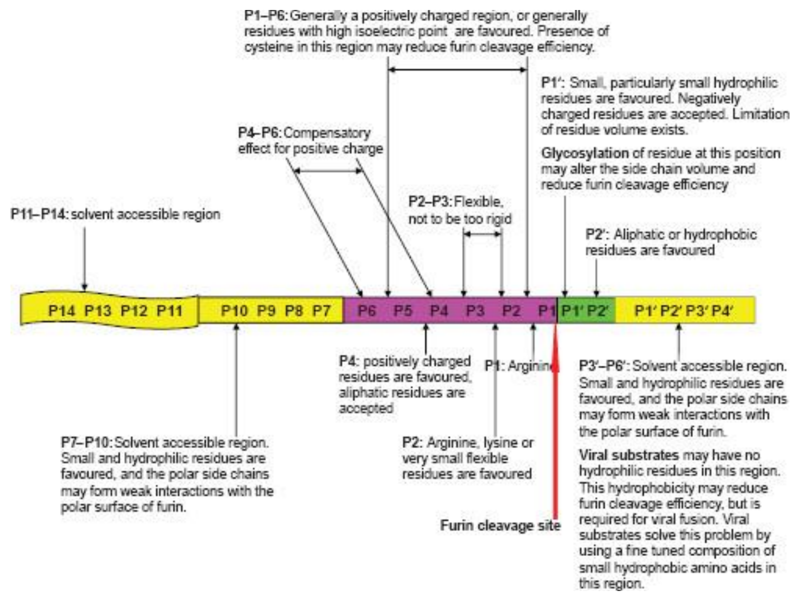
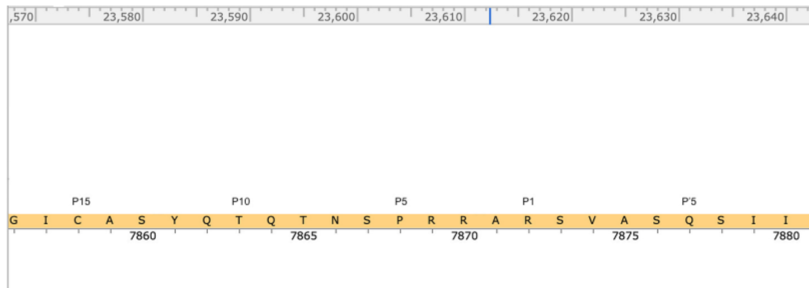
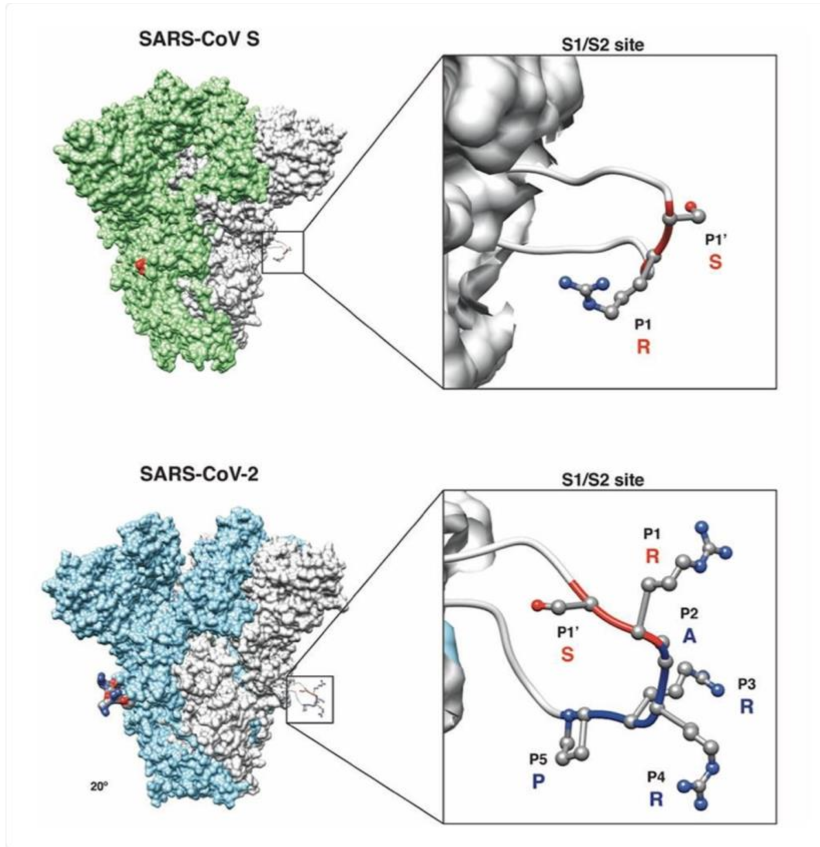


Figure 3b



476 # Figure 4a

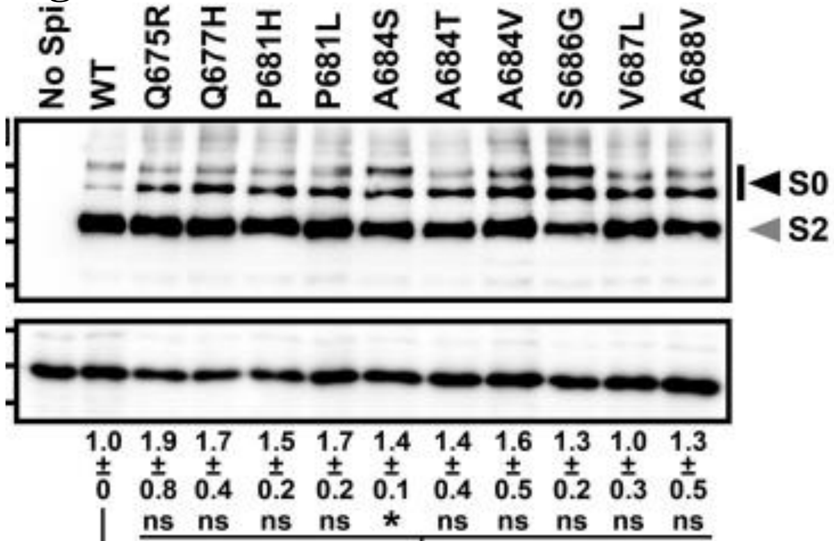
477



478

479

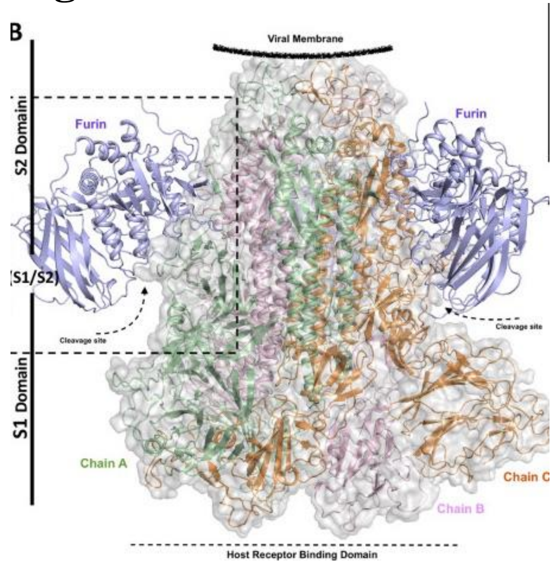
480 # Figure 4b



481

482

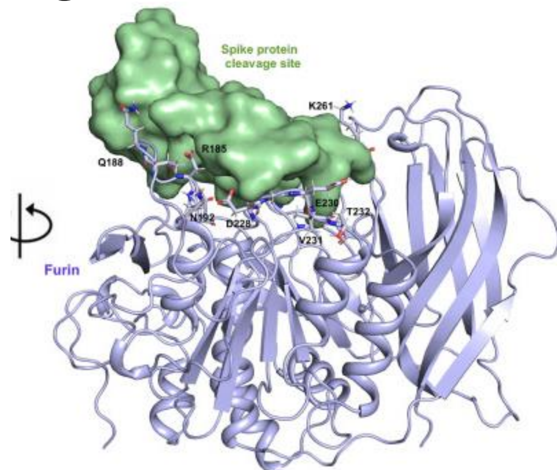
483 # Figure 5a



484

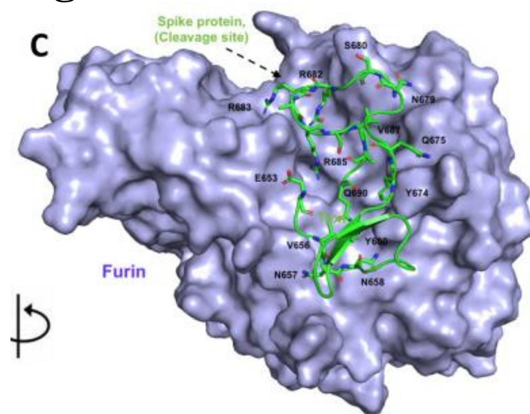
485

486 # Figure 5b



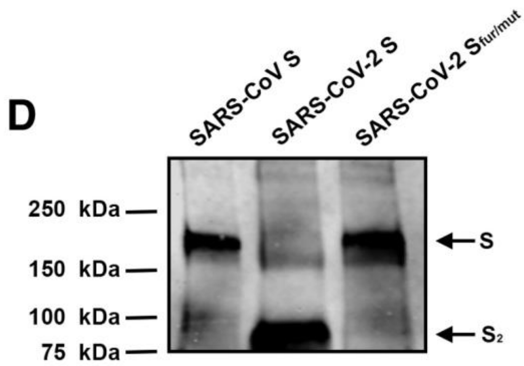
487

488 # Figure 5c



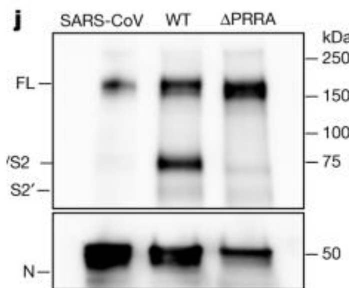
489

490 # Figure 6a

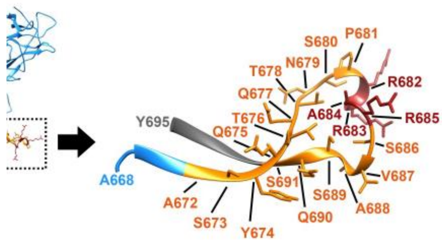


491
492
493 # Figure 6b
494

495
496
497 # Figure 6c



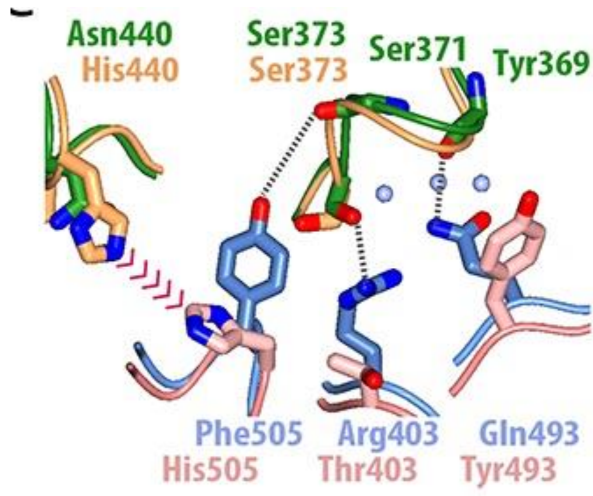
498
499
500 # Figure 6d



501

502 **Figure 7a**

503

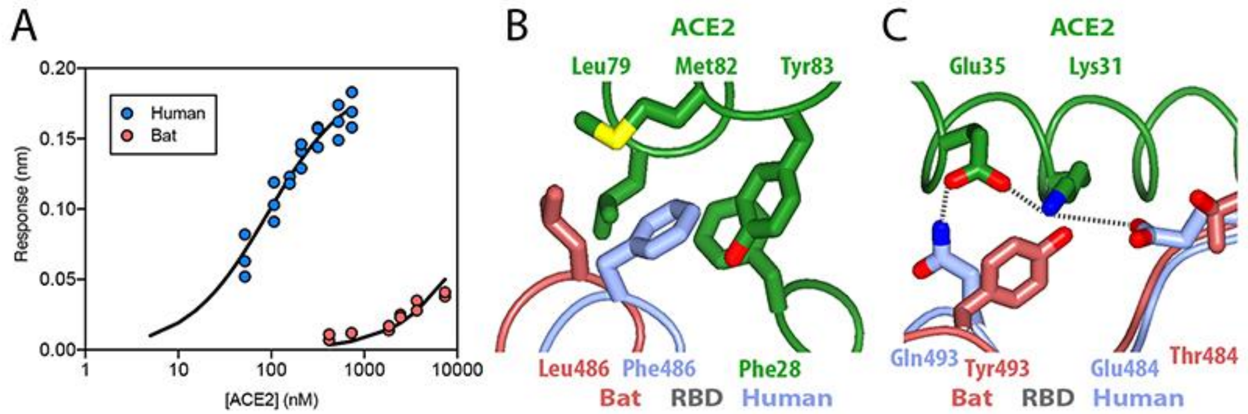


504

505

506 **Figure 7b**

507



508

509

510 **Figure 8a**

511

CONSERVED REGION   VARIABLE REGION

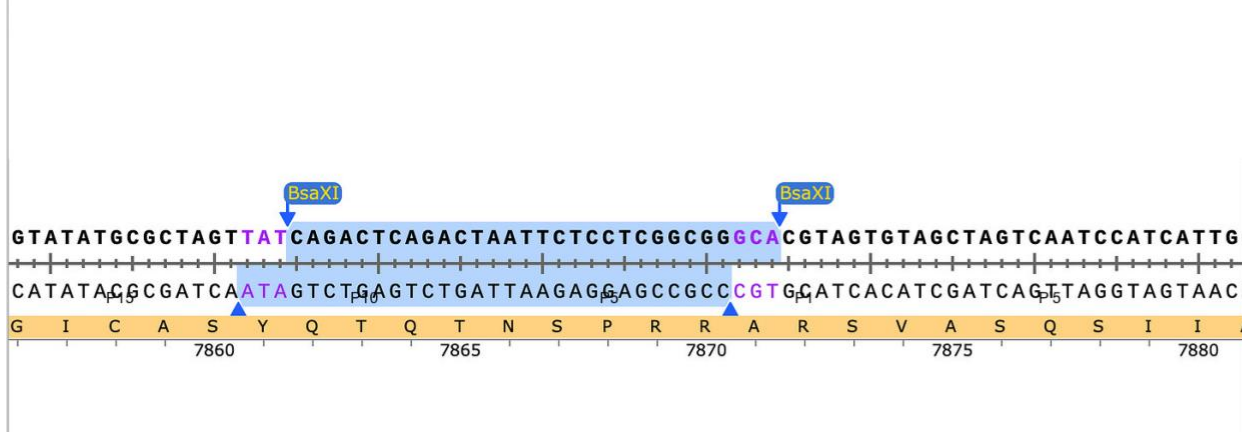
```
SARS-CoV-2   G   A   G   I   C   A   S   Y   Q   T   Q   T   N   S   P   R   R   A   R   S   V   A   S   Q   S   I   I
             ggt gca ggt ata tgc gct agt tat cag act cag act aat tct cct cgg cgg gca cgt agt gta gct agt caa tcc atc att

RaTG13       G   A   G   I   C   A   S   Y   Q   T   Q   T   N   S   -   -   -   -   R   S   V   A   S   Q   S   I   I
             ggt gca gga ata tgc gcc agt tat cag act caa act aat tca --- --- --- --- cgt agt gtg gcc agt caa tct att att
```

512

513

514 **Figure 8b**



515

516

517

518 **Figure Legends**
519
520 **Figure 1.** a. A schematic of the Spike protein of SARS-CoV-2, showing the
521 receptor binding domain and the two protease cleavage sites for furin and
522 TMPRSS2. From Keller et al. (2022). b. The amino acid sequences at the S1-S2
523 junction of the Spike proteins of four viruses, showing the FCS is present in
524 SARS-CoV-2 and MERS but not in SARS-CoV-1 or RaTG13. From Örd et al.
525 (2020). c. Amino acid sequence alignments around the S1-S2 junction of the Spike
526 proteins of many members of the sub-genus *Sarbecovirus*, with the FCS of SARS-
527 CoV-2 highlighted in red. Adapted from Hoffmann et al. (2020b).
528
529 **Figure 2.** a. The direct fusion pathway for SARS-CoV-2 entry is facilitated by
530 TMPRSS2. Adapted from Hoffmann and Pöhlmann (2021) b. The cell biology of
531 furin, showing its maturation in the Golgi, transport from the TGN to the plasma
532 membrane and subsequent uptake by endocytosis. Adapted from Thomas (2002).
533
534 **Figure 3.** a. A schematic of the generic extended furin cleavage site, illustrating
535 the core sequence flanked by two solvent-accessible regions. Adapted from Tian
536 (2009). b. The amino acid sequence of the extended FCS in the Spike protein of
537 SARS-CoV-2, labeled using the FCS numbering convention in Figure 3a.
538
539 **Figure 4.** The FCS of SARS-CoV-2 is an extended structure in a solvent-
540 accessible region of the Spike protein structure. a. A comparison between the
541 S1/S2 junction of SARS-CoV-1 and SARS-CoV-2. The insertion of PRRA extends
542 the loop structure relative to the analogous region of the SARS-CoV-1 Spike
543 protein, shown for comparison in these models. The peptide bond between R and S
544 is exposed to proteases. From Jaimes et al. (2020). b. A model of the extended
545 furin cleavage loop of SARS-CoV-2 from A668 to Y695, showing the highly
546 conserved *sarbecovirus* sequence C671-Y674 (CASY), with the R682-R685
547 (RRAR) recognition sequence highlighted. From Arora et al. (2022).
548
549
550
551

552 **Figure 5.** a. The interaction between molecules of the Spike protein and furin (in
553 lilac), showing the location of the cleavage site protruding into the solvent. From
554 Venkadari (2020). b. The catalytic domain of furin (lilac) has several
555 electronegative residues (D228, N193, E230 etc..) that can make electrostatic
556 interactions with the furin cleavage site of the spike protein. From Venkadari
557 (2020) c. The extended FCS (green) is overlaid over the surface of the furin
558 catalytic domain (lilac). Positively charged R682, R683 and R685 interact with
559 furin while P681 projects away from the enzyme. The peptide bond between R685
560 and S686 is where the enzyme catalyzes the proteolysis of the Spike protein. From
561 Venkadari (2020).
562
563 **Figure 6.** a. The Spike protein of SARS-CoV-2 is subject to efficient proteolysis in
564 cells infected with the virus, while the SARS-CoV-1 Spike is not. Mutation of the
565 FCS (Sfur/mut) abolishes proteolysis. From Hoffmann et al. (2020b). b. The Spike
566 protein of SARS-CoV-2 is subject to proteolysis in cells infected with VSV
567 pseudovirus, while the SARS-CoV-1 Spike is not. Mutation of the FCS (Sfur/mut)
568 abolishes proteolysis. From Walls et al. (2020). c. The Spike protein of SARS-
569 CoV-2 is subject to proteolysis in cells infected with a psedudovirus, while the
570 SARS-CoV-1 Spike is not. Deletion of the FCS (-PRRA) abolishes proteolysis.
571 From Peacock et al. (2021). d. Proteolysis of the Spike protein from SARS-CoV-2
572 and several natural variants or polymorphisms, showing that mutations within the
573 extended FCS decrease the efficiency of proteolysis. From Arora et al. (2022).
574
575 **Figure 7.** a. Subtle differences in the structure of the Spike protein RBD from
576 SARS-CoV-2 (blue) and the related bat virus RaTG13 (pink), illustrating
577 unfavorable interactions between His 505 of RaTG13 and its receptor (green).
578 From Wrobel et al. (2020) b. The RBD of the SARS-CoV-2 virus binds to hACE2
579 with high (nanoMolar) affinity, while the RBD of the bat virus binds weakly to
580 ACE2, due in part to the lack of Phe 486, Glu 484 and 493 in the bat virus. From
581 Wrobel et al. (2020).
582
583 **Figure 8.** a. Sequence alignment between the Spike proteins of SARS-CoV-2 and
584 RaTG13 in the region of the FCS. The enumeration of the FCS begins at C, the
585 P15 residue in the conserved domain. Despite the sequence identity of the amino
586 acid sequence the divergent nucleotide sequence suggests that RaTG13 was neither
587 the immediate evolutionary precursor of SARS-CoV-2, nor a laboratory template.
588 From Deigin and Segreto (2021). b. A restriction enzyme site for *BsaX* I is located
589 within the extended FCS region and flanks the conserved 6 amino-acid sequence
590 QTQTNS, as well as the 4 amino-acid insert PRRA.
591

592 **References**

593

594 1. Anand P, Puranik A, Aravamudan M, Venkatakrishnan AJ,
595 Soundararajan V. SARS-CoV-2 strategically mimics proteolytic
596 activation of human ENaC. Elife. 2020 May 26;9:e58603.

597 2. Apaa T, Withers AJ, Staley C, Blanchard A, Bennett M, Bremner-
598 Harrison S, CHadwicj EA, Hailer F, Harrison SWR, Loose M,
599 Mathews F, Tarlinton R (2023) Sarbecoviruses of British Horseshoe
600 Bats; Sequence Variation and Epidemiology.
601 bioRxiv 2023.02.14.528476; doi: https://doi.org/10.1101/2023.02.14.5
602 28476.

603 3. Arora, P.; Sidarovich, A.; Graichen, L.; Hörnich, B.; Hahn, A.;
604 Hoffmann, M.; Pöhlmann, S. (2022) Functional Analysis of
605 Polymorphisms at the S1/S2 Site of SARS-CoV-2 Spike
606 Protein. PLoS ONE 17: e0265453.

607 4. Baric RS. (2007). Synthetic Viral Genomics. In: *Working Papers for*
608 *Synthetic Genomics: Risks and Benefits for Science and Society*, pp.
609 35-81. Garfinkel MS, Endy D, Epstein GL, Friedman RM, eds.

610 5. Belouzard, S., Chu, V. C. & Whittaker, G. R. (2009) Activation of the
611 SARS coronavirus spike protein via sequential proteolytic cleavage at
612 two distinct sites. Proc. Natl Acad. Sci. USA 106: 5871–5876.

613 6. Borucki MK, Allen JE, Chen-Harris H, Zemla A, Vanier G, Mabery S,
614 et al. (2013) The Role of Viral Population Diversity in Adaptation of
615 Bovine Coronavirus to New Host Environments. PLoS ONE 8(1):
616 e52752.

617 7. Bresnahan PA, Leduc R, Thomas L, Thorner J, Gibson HL, Brake AJ,
618 Barr PJ, Thomas G. (1990) Human fur gene encodes a yeast KEX2-
619 like endoprotease that cleaves pro-beta-NGF in vivo. J Cell Biol.
620 111:2851-2859.

621 8. Cassari L, Pavan A, Zoia G, Chinellato M, Zeni E, Grinzato A,
622 Rothenberger S, Cendron L, Dettin M, Pasquato A. (2023) SARS-CoV-2 S
623 Mutations: A Lesson from the Viral World to Understand How Human
624 Furin Works. International Journal of Molecular Sciences 24:4791.

625
626 9. Chan YA, Zhan SH. (2022) The Emergence of the Spike Furin
627 Cleavage Site in SARS-CoV-2. Mol Biol Evol. 39: msab327.

628 10. Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah,
629 N.G., and Decroly, E. (2020). The spike glycoprotein of the new
630 coronavirus 2019-nCoV contains a furin-like cleavage site absent in
631 CoV of the same clade. Antiviral Res. 176:104742.

632 11. Dahms SO, Arciniega M, Steinmetzer T, Huber R, Than ME.
633 (2016) Structure of the unliganded form of the proprotein convertase
634 furin suggests activation by a substrate-induced mechanism. Proc
635 Natl Acad Sci U S A. 113: 11196-11201.

636 12. Deigin Y, Segreto R (2021) SARS-CoV-2′s claimed natural
637 origin is undermined by issues with genome sequences of its relative
638 strains. Coronavirus sequences RaTG13, MP789 and RmYN02 raise
639 multiple questions to be critically addressed by the scientific
640 community. BioEssays 43: 2100015.

641 13. Duckert P, Brunak S, Blom N (2004) Prediction of proprotein
642 convertase cleavage sites, Protein Engineering, Design and
643 Selection, 17: 107–112.

644 14. Essalmani R, Jain J, Susan-Resiga D, Andréo U, Evagelidis A,
645 Derbali RM, Huynh DN, Dallaire F, Laporte M, Delpal A, Sutto-Ortiz
646 P, Coutard B, Mapa C, Wilcoxen K, Decroly E, Nq Pham T, Cohen
647 ÉA, Seidah NG. (2022) Distinctive Roles of Furin and TMPRSS2 in
648 SARS-CoV-2 Infectivity. J Virol. 96:e0012822.

649 15. Engler C, Marillonnet S. (2014) Golden Gate cloning. Methods
650 Mol Biol. 1116:119-131.

651 16. Gallaher WR (2020) A palindromic RNA sequence as a
652 common breakpoint contributor to copy-choice recombination in
653 SARS-COV-2 Arch. Virol., 165: 2341-2348.

654 17. Garry RF. (2022) SARS-CoV-2 furin cleavage site was not
655 engineered. Proceedings of the National Academy of Sciences of the
656 United States of America. 119: e2211107119.

657 18. Harrison NL, Sachs JD. (2022a) A call for an independent
658 inquiry into the origin of the SARS-CoV-2 virus. Proc Natl Acad Sci U
659 S A. 119: e2202769119.

660 19. Harrison NL, Sachs JD. (2022b) The origin of SARS-CoV-2
661 remains unresolved. Proc Natl Acad Sci U S A. 119: e2215826119.

662 20. Haynes, W.A., Kamath, K., Bozekowski, J. *et al.* (2021) High-
663 resolution epitope mapping and characterization of SARS-CoV-2
664 antibodies in large cohorts of subjects with COVID-19.*Commun*
665 *Biol* **4**, 1317.

666 21. Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber
667 R, Lindberg I, Bode W, Than ME. (2003) The crystal structure of the
668 proprotein processing proteinase furin explains its stringent
669 specificity. Nat Struct Biol. 10:520-526.

670 22. Hoffmann M, Kleine-Weber H, Pöhlmann S. (2020a) A
671 Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is
672 Essential for Infection of Human Lung Cells. Mol Cell. 78:779-784.e5.

673 23. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler
674 T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller
675 MA, Drosten C, Pöhlmann S. (2020b) SARS-CoV-2 Cell Entry
676 Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically
677 Proven Protease Inhibitor. Cell 181:271-280.e8.

678 24. Hoffmann, M., Pöhlmann, S. (2021) How SARS-CoV-2 makes
679 the cut. *Nat Microbiol* **6**, 828–829.

680 25. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL,
681 Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF,
682 Doherty PC, Farrar J, Geoghegan JL, Jiang X, Leibowitz JL, Neil
683 SJD, Skern T, Weiss SR, Worobey M, Andersen KG, Garry RF,
684 Rambaut A. The origins of SARS-CoV-2: A critical review. Cell. 2021
685 Sep 16;184(19):4848-4856.

686 26. Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T,
687 Dinnon KH 3rd, Kato T, Lee RE, Yount BL, Mascenik TM, Chen G,
688 Olivier KN, Ghio A, Tse LV, Leist SR, Gralinski LE, Schäfer A, Dang
689 H, Gilmore R, Nakano S, Sun L, Fulcher ML, Livraghi-Butrico A,

690 Nicely NI, Cameron M, Cameron C, Kelvin DJ, de Silva A, Margolis
691 DM, Markmann A, Bartelt L, Zumwalt R, Martinez FJ, Salvatore SP,
692 Borczuk A, Tata PR, Sontake V, Kimple A, Jaspers I, O'Neal WK,
693 Randell SH, Boucher RC, Baric RS. (2020) SARS-CoV-2 Reverse
694 Genetics Reveals a Variable Infection Gradient in the Respiratory
695 Tract. Cell 182:429-446.e14.

696 27. Hughey RP, Bruns JB, Kinlough CL, Harkleroad KL, Tong Q,
697 Carattino MD, Johnson JP, Stockand JD, Kleyman TR. (2004)
698 Epithelial sodium channels are activated by furin-dependent
699 proteolysis. J Biol Chem. 279:18111-18114.

700 28. Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR
701 (2020). Phylogenetic Analysis and Structural Modeling of SARS-CoV-
702 2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically
703 Sensitive Activation Loop. J Mol Biol. 2020 May 1;432(10):3309-
704 3325.

705 29. Johnson BA, Xie X, Kalveram B, Lokugamage KG, Muruato A,
706 Zou J, Zhang X, Juelich T, Smith JK, Zhang L, Bopp N, Schindewolf
707 C, Vu M, Vanderheiden A, Swetnam D, Plante JA, Aguilar P, Plante
708 KS, Lee B, Weaver SC, Suthar MS, Routh AL, Ren P, Ku Z, An Z,
709 Debbink K, Shi PY, Freiberg AN, Menachery VD. (2021) Furin
710 Cleavage Site Is Key to SARS-CoV-2 Pathogenesis. Nature,
711 591:293-299.

712 30. Keller, C., Böttcher-Friebertshäuser, E. & Lohoff, M. (2022)
713 TMPRSS2, a novel host-directed drug target against SARS-CoV-
714 2. Sig Transduct Target Ther 7: 251.

715 31. Klimstra WB, Heidner HW, Johnston RE. (1999) The furin
716 protease cleavage recognition sequence of Sindbis virus PE2 can
717 mediate virion attachment to cell surface heparan sulfate. J Virol.
718 73:6299-6306.

719 32. Klotz LC, Sylvester EJ. (2014) The consequences of a lab
720 escape of a potential pandemic pathogen. Front Public Health. 2:116.

721 33. Kota P, Gentzsch M, Dang YL, Boucher RC, Stutts MJ. (2018)
722 The N terminus of α-ENaC mediates ENaC cleavage and activation
723 by furin. J Gen Physiol. 150:1179-1187.

34. Lerner S, Hvistendahl M. (2021) New details emerge about coronavirus research at Chinese lab. The Intercept (6 September 2021). https://theintercept.com/2021/09/06/new-details-emerge-about-coronavirus-research-at-chinese-lab/. Accessed 22 February 2022.

35. Li, F., Li, W., Farzan, M., Harrison, S. C. (2005) Structural biology: structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864–1868.

36. Li W, Wicht O, van Kuppeveld FJ, He Q, Rottier PJ, Bosch BJ. (2015) A Single Point Mutation Creating a Furin Cleavage Site in the Spike Protein Renders Porcine Epidemic Diarrhea Coronavirus Trypsin Independent for Cell Entry and Fusion. J Virol. 89:8077-81.

37. Menachery VD, Dinnon KH 3rd, Yount BL Jr, McAnarney ET, Gralinski LE, Hale A, Graham RL, Scobey T, Anthony SJ, Wang L, Graham B, Randell SH, Lipkin WI, Baric RS. (2020) Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection. J Virol. 2020 94:e01774-19.

38. Menachery VD, Yount BL Jr, Sims AC, Debbink K, Agnihothram SS, Gralinski LE, Graham RL, Scobey T, Plante JA, Royal SR, Swanstrom J, Sheahan TP, Pickles RJ, Corti D, Randell SH, Lanzavecchia A, Marasco WA, Baric RS. (2016) SARS-like WIV1-CoV poised for human emergence. Proc Natl Acad Sci U S A. 113:3048-53.

39. Meng B, Abdullahi A, Ferreira IATM, Goonawardane N, Saito A, Kimura I, Yamasoba D, Gerber PP, Fatihi S, Rathore S, Zepeda SK, Papa G, Kemp SA, Ikeda T, Toyoda M, Tan TS, Kuramochi J, Mitsunaga S, Ueno T, Shirakawa K, Takaori-Kondo A, Brevini T, Mallery DL, Charles OJ; CITIID-NIHR BioResource COVID-19 Collaboration; Genotype to Phenotype Japan (G2P-Japan) Consortium; Ecuador-COVID19 Consortium; Bowen JE, Joshi A, Walls AC, Jackson L, Martin D, Smith KGC, Bradley J, Briggs JAG, Choi J, Madissoon E, Meyer KB, Mlcochova P, Ceron-Gutierrez L, Doffinger R, Teichmann SA, Fisher AJ, Pizzuto MS, de Marco A, Corti D, Hosmillo M, Lee JH, James LC, Thukral L, Veesler D, Sigal A, Sampaziotis F, Goodfellow IG, Matheson NJ, Sato K, Gupta RK.

759 (2022) Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts
760 infectivity and fusogenicity. Nature. 603:706-714.

761 40. Millet, J. K. & Whittaker, G. R. (2014) Host cell entry of Middle
762 East respiratory syndrome coronavirus after two-step, furin-mediated
763 activation of the spike protein. Proc. Natl Acad. Sci. USA 111, 15214–
764 15219.

765 41. Millet, J. K. & Whittaker, G. R. (2015) Host cell proteases:
766 critical determinants of coronavirus tropism and pathogenesis. Virus
767 Res. 202, 120–134.

768 42. Molloy SS, Thomas L, VanSlyke JK, Stenberg PE, Thomas G.
769 (1994) Intracellular trafficking and activation of the furin proprotein
770 convertase: localization to the TGN and recycling from the cell
771 surface. EMBO J. 13:18-33.

772 43. Örd, M., Faustova, I. & Loog, M. (2020) The sequence at Spike
773 S1/S2 site enables cleavage by furin and phospho-regulation in SARS-
774 CoV2 but not in SARS-CoV1 or MERS-CoV. Sci Rep 10: 16944.

775 44. Ou T, Mou H, Zhang L, Ojha A, Choe H, Farzan M. (2021)
776 Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is
777 attenuated by TMPRSS2. PLoS Pathog. 2021 17:e1009212.

778 45. Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann
779 OC, Kugathasan R, Penn R, Brown JC, Sanchez-David RY, Braga L,
780 Williamson MK, Hassard JA, Staller E, Hanley B, Osborn M, Giacca
781 M, Davidson AD, Matthews DA, Barclay WS. (2021a) The furin
782 cleavage site in the SARS-CoV-2 spike protein is required for
783 transmission in ferrets. Nat Microbiol. 6:899-909.

784 46. Peacock TP, Sheppard CM, Brown JC, Goonawardane N, Zhou
785 J, Whiteley M, et al. (2021b) The SARS-CoV-2 variants associated
786 with infections in India, B.1.617, show enhanced spike cleavage by
787 furin. bioRxiv. 2021:2021.05.28.446163.

788 47. Relman DA. (2020) To stop the next pandemic, we need to
789 unravel the origins of COVID-19. Proc Natl Acad Sci U S A.
790 117:29246-29248.

791 48. Sander AL, Moreira-Soto A., Yordanov S. *et al.* (2022) Genomic
792 determinants of Furin cleavage in diverse European SARS-related
793 bat coronaviruses. Commun Biol 5: 491.

794 49. Schrauwen EJA, Bestebroer TM, Munster VJ, de Wit E, Herfst
795 S, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. (2011)
796 Insertion of a multibasic cleavage site in the haemagglutinin of human
797 influenza H3N2 virus does not increase pathogenicity in ferrets. J
798 Gen Virol. 2011 92:1410-1415.

799 50. Segreto R, Deigin Y. The genetic structure of SARS-CoV-2
800 does not rule out a laboratory origin: SARS-COV-2 chimeric structure
801 and furin cleavage site might be the result of genetic manipulation.
802 Bioessays. 2021 Mar;43(3):e2000240.

803 51. Seidah NG, Day R, Marcinkiewicz M, Chrétien M. (1998)
804 Precursor convertases: an evolutionary ancient, cell-specific,
805 combinatorial mechanism yielding diverse bioactive peptides and
806 proteins. Ann N Y Acad Sci. 839:9-24.

807 52. Seidah NG & Prat A. (2012) The biology and therapeutic
808 targeting of the proprotein convertases. Nature Reviews Drug
809 Discovery, 11: 367-383.

810 53. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q,
811 Auerbach A, Li F. (2020) Structural basis of receptor recognition by
812 SARS-CoV-2. Nature. 581:221-224.

813 54. Sirotkin K, Sirotkin D. Might SARS-CoV-2 Have Arisen via
814 Serial Passage through an Animal Host or Cell Culture?: A potential
815 explanation for much of the novel coronavirus' distinctive genome.
816 Bioessays. 2020 Oct;42(10):e2000091.

817 55. Temmam S, Vongphayloth K, Salazar EB, Munier S, Bonomi M,
818 Regnault B, Douangboubpha B, Karami Y, Chrétien D, Sanamxay D
819 et al (2022) Bat coronaviruses related to SARS-CoV-2 and infectious
820 for human cells. Nature 604: 330-336

821 56. Temmam S, Montagutelli X, Herate C, Donati F, Regnault B,
822 Attia M, Baquero Salazar E, Chretien D, Conquet L, Jouvion G, Pipoli
823 Da Fonseca J, Cokelaer T, Amara F, Relouzat F, Naninck T,

824 Lemaitre J, Derreudre-Bosquet N, Pascal Q, Bonomi M, Bigot T,
825 Munier S, Rey FA, Le Grand R, van der Werf S, Eloit M. (2023)
826 SARS-CoV-2-related bat virus behavior in human-relevant models
827 sheds light on the origin of COVID-19. EMBO Rep. 6:e56055.

828 57. Tengs T, LaFramboise T, Den RB, Hayes DN, Zhang J,
829 DebRoy S, Gentleman RC, O'Neill K, Birren B, Meyerson M. (2004)
830 Genomic representations using concatenates of Type IIB restriction
831 endonuclease digestion fragments. Nucleic Acids Res. 32:e121.

832 58. Thomas G. (2002) Furin at the cutting edge: from protein traffic
833 to embryogenesis and disease. Nat Rev Mol Cell Biol. 3:753-766

834 59. Tian S. (2009) A 20 residues motif delineates the furin cleavage
835 site and its physical properties may influence viral
836 fusion. Biochemistry Insights; 2. doi:10.4137/BCI.S2049.

837 60. Van Helden J, Butler CD, Achaz G, Canard B, Casane D,
838 Claverie JM, Colombo F, Courtier V, Ebright RH, Graner F,
839 Leitenberg M, Morand S, Petrovsky N, Segreto R, Decroly E, Halloy
840 J. (2021) An appeal for an objective, open, and transparent scientific
841 debate about the origin of SARS-CoV-2. Lancet. 398:1402-1404.

842 61. Venkadari N (2020). Structure of Furin Protease Binding to
843 SARS-CoV-2 Spike Glycoprotein and Implications for Potential
844 Targets and Virulence. The Journal of Physical Chemistry Letters 11:
845 6655-6663.

846 62. Vu MN, Lokugamage KG, Plante JA, Scharton D, Bailey AO,
847 Sotcheff S, Swetnam DM, Johnson BA, Schindewolf C, Alvarado RE,
848 Crocquet-Valdes PA, Debbink K, Weaver SC, Walker DH, Russell
849 WK, Routh AL, Plante KS, Menachery VD. (2022) QTQTN motif
850 upstream of the furin-cleavage site plays a key role in SARS-CoV-2
851 infection and pathogenesis. Proc Natl Acad Sci U S A. 119:
852 e2205690119.

853 63. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler
854 D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2
855 Spike Glycoprotein. Cell. 181:281-292.e6

856 64. Watanabe R, Matsuyama S, Shirato K, Maejima M, Fukushi S,
857 et al. (2008) Entry from the Cell Surface of Severe Acute Respiratory
858 Syndrome Coronavirus with Cleaved S Protein as Revealed by
859 Pseudotype Virus Bearing Cleaved S Protein. J Virol 82: 11985–
860 11991.

861 65. Winstone H, Lista MJ, Reid AC, Bouton C, Pickering S, Galao
862 RP, Kerridge C, Doores KJ, Swanson CM, Neil SJD. (2021) The
863 Polybasic Cleavage Site in SARS-CoV-2 Spike Modulates Viral
864 Sensitivity to Type I Interferon and IFITM2. J Virol. 95:e02422-20.

865 66. Wolf KA, Kwan JC, Kamil JP. (2022) Structural Dynamics and
866 Molecular Evolution of the SARS-CoV-2 Spike Protein. mBio.
867 13:e0203021.

868 67. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL,
869 Abiona O, Graham BS, McLellan JS. (2020) Cryo-EM structure of the
870 2019-nCoV spike in the prefusion conformation. Science. 367:1260-
871 1263.

872 68. Wrobel AG, Benton DJ, Xu P, Roustan C, Martin SR, Rosenthal
873 PB, Skehel JJ, Gamblin SJ. (2020) SARS-CoV-2 and bat RaTG13
874 spike glycoprotein structures inform on virus evolution and furin-
875 cleavage effects. Nat Struct Mol Biol. 27:763-767.

876 69. Wu Y, Zhao S (2021) Furin cleavage sites naturally occur in
877 coronaviruses. Stem Cell Research, 50, 102115.

878 70. Zimmer G, Conzelmann KK, Herrler G. (2002) Cleavage at the
879 furin consensus sequence RAR/KR(109) and presence of the
880 intervening peptide of the respiratory syncytial virus fusion protein are
881 dispensable for virus replication in cell culture. J Virol. 76:9218-9224.

882
883