1

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.

The "furin cleavage site" (FCS), is an important feature of the Spike protein of the SARS-CoV-2 virus (Figure 1a). The FCS is present in the novel virus SARS-CoV-2, but is absent in SARS-CoV-1 (Figure 1b) and in all other members of the group of most closely related viruses, the subgenus sarbecovirus (Figure 1c) (Coutard et al., 2020; Hoffmann et al., 2020b; Temmam et al., 2023), which now includes several hundred viruses, almost all of which were found in bats around the world.

A great deal has been written and said about the FCS in scientific journals, in mainstream news outlets and on social media media, yet there is still considerable confusion about the structure, function, and importance of the FCS in SARS-CoV-2. We aim to clarify key issues here.

The FCS is the site at the S1-S2 junction where the Spike protein of the virus is cut by *furin,* an enzyme that is expressed in most human cells (Thomas, 2002). This process is known as proteolysis. Processing of the Spike protein by a combination of the two enzymes, furin and TMPRSS2 (Hoffmann et al., 2020a, Ou et al., 2021) (**Figure 2a**) is critical to the entry of SARS-CoV-2 into cells in the lower respiratory tract (Coutard et al., 2020) to infect the human lung - as well as for the transmission and pathogenicity of the virus (Johnson et al., 2021, Peacock et al., 2021).

The importance of furin cleavage for viral entry has been known since the earliest work on *Sindbis* virus (Klimstra et al., 1999) but this is not a universal feature for all viruses. For example, the replication of respiratory syncytial virus (RSV) does not require furin (Zimmer et al., 2002). We thought it a good idea to outline here some of the basic biology of furin and the FCS, and to explain why this become so important to the study of how the virus enters cells, as well as to discussions of the origins of the virus.

What is furin?
Furin is an enzyme that cuts proteins, i.e. it is a *protease*. To be more specific, it is a member of a group called *proprotein covertases*, which is to say it is an enzyme that is responsible for processing larger proteins into their active final form (Seidah et al., 1998; Seidah & Prat, 2012). Processing of larger precursors is especially important in endocrinology, with many hormones such as insulin being derived by proteolysis of larger protein precursors. Furin itself is produced by auto-proteolysis from a precursor protein (Thomas, 2012).

Furin was first identified and originally designated as PACE, and the gene was cloned in 1990 (Bresnahan et al., 1990). Unlike other proteases such as TMPRSS2 that are found at the cell surface (Meng et al., 2020), furin is mainly located inside the cell, specifically inside membrane-bound organelles (**Figure 2b**), being synthesized on the endoplasmic reticulum and then is modified in the presence of Ca^{2+}, moving through the ER-Golgi intermediate complex (ER-GIC) to the trans-Golgi network (TGN) where the mature form of the protein is stabilized at low pH (Thomas, 2002).

Once in the TGN, furin modifies proteins that have already been synthesized, including those being packaged into vesicles for export to the plasma membrane (**Figure 2b**). In the process, furin itself appears on the cell surface, from where it is then recycled into endocytic vesicles (Molloy et al., 1994). Viruses, including coronaviruses, make use of this cellular sorting machinery to facilitate the manufacture of membrane-bound protein, forming new virus particles. The Spike protein of SARS-CoV-2 is one example, and it can be cleaved by furin inside the infected cell before being packaged into newly synthesized virus containing other viral components. These are essential steps in the process by which new virions exit the cell.

What exactly is a furin cleavage site?
Furin works by "cutting" the peptide strand of a precursor protein at one or more locations, specifically by *catalyzing the hydrolysis of a specific peptide bond*, typically between an arginine residue and its immediate neighbour, which is variable but is typically a smaller (serine or valine) residue (6). The hydrolysis of the peptide bond at R↓X is most efficient when a specific *furin recognition sequence* is present, which is enriched in basic amino acid residues (R/K). The typical furin cleavage site contains at least 2 basic residues, separated by 2 amino acids, **RXXR** (this is RRAR in SARS-CoV-2; **Figure 1a**), but there are exceptions to this rule. This minimal consensus recognition sequence is what is usually referred to as a *"furin cleavage site",* although the proteolysis "cleavage" actually takes place very specifically at what is termed the "scissile bond" (R↓X), at the C-terminal end of the RXXR recognition sequence.

Note that proteins can also be susceptible to proteolysis by other enzymes found outside cells, such as trypsin and cathepsin B etc., and that the importance of the specific furin recognition sequence is to enhance the efficiency of proteolysis by one specific protease, furin, which is expressed in most cells. There are examples of proteins that contain multiple FCS, so

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

4

157 with the catalytic domain of furin are thought to be made in the core, with
158 the flanking regions providing stabilizing interactions (Venkadari, 2020).
159 More recent work has shown that the flanking regions can also be very
160 important, as in the case of the QTQTN motif (P7-P11) of the SARS-CoV-2
161 Spike (**Figure 3b**), which has been shown to be important for SARS-CoV-2
162 pathogenesis (Vu et al., 2022) as well as being highly antigenic (Haynes et
163 al., 2021).
164
165 The S1/S2 junction of the SARS-CoV-1 Spike lacks the FCS but forms a
166 short solvent-exposed loop (**Figure 4a**) and this loop is further extended in
167 SARS-CoV-2 (**Figure 4a, 4b**) due to the PRRA insert (Jaimes et al., 2020),
168 exposing the FCS for proteolysis. The intricate details of the SARS-CoV-2
169 FCS are only truly revealed in an elegant study of its interaction with the
170 catalytic domain of furin (**Figure 5a**) (Venkadari, 2020). The basic residues
171 within the polybasic RXXR sequence of a typical FCS make electrostatic
172 contacts with negatively charged residues in the catalytic domain of furin
173 (Venkadari, 2020) (**Figure 5b**), while in the case of SARS-CoV-2, the "P5"
174 residue is modeled with the proline side chain oriented away from the
175 catalytic domain (**Figure 5c)** so that there is no steric or electrostatic
176 hindrance. In fact, a variety of smaller amino acids can be tolerated at this
177 P5 position of FCS (Tian, 2009); there is thus no thermodynamic or steric
178 "prohibition" against a proline residue being located adjacent to the core
179 recognition sequence.
180
181 It follows from this survey of many FCS domains that a proline at the P5
182 position is neither unexpected nor unusual, contrary to some commentary
183 (Garry, 2022). Indeed, this point is emphasized by the fact that the Spike of
184 the MERS virus also has a proline at this P5 position (Millet et al., 2014,
185 Garry, 2022), although opinions may differ on whether the FCS of MERS is
186 itself fully functional (Millet et al., 2014).
187
188 **The furin cleavage site of SARS-CoV-2 is fully functional**
189 The argument has been advanced that the FCS in SARS-CoV-2 is "sub-
190 optimal". This claim seems to be based on computer algorithms that predict
191 the functionality of FCS sequences. Two such algorithms are now in
192 widespread use (e.g. Duckert et al., 2004). Yet these algorithms are known
193 to give "false negatives", i.e. the failure to predict a fully functional FCS.
194
195 One example of a predicted "sub-optimal" FCS (Holmes et al., 2021) is the
196 RRAR⬇SVAS sequence of SARS-CoV-2 itself. Although this FCS scores

197 lower than others according to prediction algorithms, it is fully and efficiently
198 cleaved when cells are infected with SARS-CoV-2 (Hoffmann et al., 2020b,
199 **Figure 6a**) or with pseudoviruses expressing the SARS-CoV-2 Spike
200 protein (Walls et al., 2020; Peacock et al., 2021) (**Figure 6b**), in contrast to
201 the SARS-CoV-1 Spike (**Figure 6c**) and this efficient proteolysis is
202 abolished by deletion of the PRRA sequence (**Figure 6c**). It is worth noting
203 that an identical FCS sequence is found in the α subunit of human ENaC
204 (Anand et al., 2020) and that the ENaC α protein from mouse and rat is
205 also completely and efficiently cleaved by furin in epithelial cells (Hughey et
206 al., 2004, Kota et al., 2018).
207
208 Two algorithms predict that the FCS of SARS-CoV-2 is not "ideal", and this
209 led several scientists to speculate that the FCS was "sub-optimal" (Holmes
210 et al., 2021). Of course, the efficient proteolysis described above (**Figure
211 6a-d**) and the high human-to-human transmissibility of SARS-CoV-2 (R_0
212 ~2-3 for the original Wuhan-1 virus) would seem to argue otherwise.
213
214 Because of the assertions that the FCS is "sub-optimal", one might expect
215 to find viral variants that show FCS mutations enhanced proteolysis. In fact,
216 the RRAR (P4-P1) core sequence has remained remarkably stable (Wolf et
217 al., 2022), suggesting that such mutations would confer no evolutionary
218 advantage. A variety of natural polymorphisms (point mutations) within the
219 extended FCS region have now been tested (Arora et al., 2022), and these
220 mutations actually *resulted in a modest loss of proteolysis efficiency*
221 relative to the original FCS (**Figure 6d**). While it has been suggested that
222 the efficiency of proteolysis of the SARS-CoV-2 Spike protein would be
223 further enhanced by mutation at the P5 residue, including the naturally
224 observed mutations P681R and P681H (Peacock et al., 2021b), it is now
225 clear that the experimental data do not in fact support this. In addition, it
226 has been shown that the FCS of the original SARS-CoV-2 Spike protein
227 can be imported into the SARS-CoV-1 Spike by engineering and that the
228 resulting mutant Spike is then fully and efficiently cleaved by furin, in cells
229 experimentally infected via a pseudovirus (Winstone et al., 2021). There is
230 thus no convincing evidence that the FCS of SARS-CoV-2 is sub-optimal.
231
232 **Insights from Structural Biology: studies of the Spike protein**
233 Elegant work done following the emergence of SARS-CoV-1 had shown
234 that this virus uses the human membrane protein angiotensin converting
235 enzyme-2 (ACE2) as its primary receptor on human cells (Li et al, 2005).
236 Studies of SARS-CoV-2 confirmed that the newer virus also uses human

237 ACE2 as its primary receptor on susceptible cells (Shang et al., 2020;
238 Hoffmann et al., 2020a) and that the Spike protein forms a trimeric
239 assembly that binds with high affinity to its receptor (Walls et al., 2020).
240 The trimer exists in multiple conformational sates and binds to hACE2 with
241 one S^B domain of the Spike protein in the open conformation (Wrapp et al.,
242 2020 Walls et al., 2020). In one especially insightful experiment (Wrobel et
243 al., 2020), a comparison was made between the Spike proteins of SARS-
244 CoV-2 and a close relative termed RaTG13, which is known to infect bats
245 but is not thought to infect humans. Firstly, it was found that there were
246 small but significant differences between the structures of the receptor
247 binding domains (RBD) of the Spike proteins of these two viruses (**Figure**
248 **7a**). Secondly, it was noted that the Spike of RaTG13 binds very weakly to
249 hACE2, with as much as 1000 times lower affinity (**Figure 7b**), largely due
250 to steric constraints in the interaction between the RBD of the Spike and
251 hACE2 (**Figure 7b**). Crucially, these and other authors noted that cleavage
252 by furin destabilized the "closed" conformation of the Spike protein trimer
253 and hence promoted the "open" conformation, exposing the RBD that is
254 necessary for the binding of the Spike to hACE2. The lack of FCS in the
255 RaTG13 limits the Spike cleavage by proteolysis and stabilized the Spike in
256 the closed conformation associated with low affinity binding to hACE2
257 (Wrobel et al, 2020).
258
259 **The importance of FCS for viral transmission**
260 The importance of FCS for viral entry and transmissibility has been known
261 for some time. Extensive work has been performed in laboratories around
262 the world to insert FCS via genetic engineering into pathogenic viruses,
263 including influenza viruses (Schrauwen et al., 2011) and coronaviruses
264 such as SARS-CoV-1 (Millet et al., 2015) and the porcine epidemic
265 diarrhea virus, PEDV (Li et al., 2015). It should be pointed out that the vast
266 majority of this work was done in the context of *pseudovirus* experiments,
267 using an innocuous virus as a backbone, so that the chance of producing a
268 highly pathogenic virus with increased transmissibility escaping from a
269 laboratory was minimized (Belouzard et al., 2009). Work done in this format
270 is therefore considered safe and does not constitute what is known as
271 "Gain-of-Function" (GoF) research. In at least one case (the example given
272 above of PEDV), this type of work created a replication-competent novel
273 recombinant virus (Li et al., 2015), showing that engineering of functional
274 infectious viruses in this region is neither implausible nor novel.
275
276

Research on the FCS in SARS-like viruses since SARS-CoV-1

There has obviously been a long-standing interest among coronavirus researchers in the role of spike protein proteolysis in enhancing viral entry. Since the original epidemic of SARS and the identification of SARS-CoV-1 as a coronavirus that uses ACE2 as its receptor on epithelial cells, much attention has been devoted to the study of the SARS-CoV-1 Spike protein (Li et al., 2005), and FCS has been inserted into the spike protein of this virus on multiple occasions (Belouzard et al., 2009; Millet et al., 2015).

Work on proteolysis as a determinant of viral entry was recently extended to a range of viruses that include bat coronaviruses from the *sarbecovirus* group (Menachery et al., 2020) some of which have been suggested to show the potential for emergence (Menachery et al., 2015). Such work is sometimes done under conditions of limited containment (BSL-2) because of the perception that these bat *sarbecoviruses* lack pathogenic potential. Work done on these bat viruses has proliferated, and is widely considered to be a "grey area" that constitutes Gain-of-Function research of concern (GoFRoC), since chimeric viruses of unknown function are created and the pathogenicity of the resulting virus cannot be predicted. Laboratory experiments using pseudoviruses showed that infection of lung cells by SARS-CoV-2 and transmissibility between ferrets is strongly inhibited by removal of the PRRA sequence (Peacock et al., 2021), and complementary work in pseudovirus experiments confirmed that insertion of the PRRA sequence into the Spike protein of SARS-CoV-1 confers high furin sensitivity (Winstone et al., 2021) and enhances viral entry into cells.

The possible origins of the FCS in SARS-CoV-2

A combination of efficient human-to-human transmission with significant pathogenicity is one of the hallmarks of a pathogen with pandemic potential (PPP). In fact, the combination of factors that made SARS-CoV-2 a pandemic virus involves a combination of its high affinity for human ACE2, its processing by furin and TMPRSS2 (Essalmani et al., 2022) as well as its ability to down-regulate the "innate immune response" in humans that is mediated by interferons (Winstone et al., 2021), perhaps via one or more of the "accessory" proteins of the virus, encoded by 3'-open reading frames (Orf). All of these features of the virus may have arisen naturally, but the unique nature of the FCS in SARS-CoV-2 among the SARS-related bat viruses of the *sarbecovirus* clade (Coutard et al., 2020, Hoffmann et al., 2020b) is quite remarkable (**Figure 1c**) and has given rise to speculation 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

357 criticisms of these proposals has been the lack of a natural virus that is a
358 sufficiently good match to the SARS-CoV-2 RNA sequence to have served
359 as the immediate precursor for the proposed recombination event (Chan
360 and Zhao, 2022). Even the closely related virus RaTG13, or the viruses
361 identified in bats from Laos are not sufficiently similar at the RNA level to
362 have served as the immediate ancestor to SARS-CoV-2 (Gallaher, 2020;
363 Temmam et al., 2022). Less well-articulated arguments in favor of a natural
364 origin of the FCS have been advanced (Garry, 2022) and clearly refuted
365 (Harrison and Sachs, 2022b).
366

367 **Did the FCS of SARS-CoV-2 Arise During Serial Passage?**
368 An alternate explanation for the presence of the FCS in SARS-CoV-2 and
369 its apparent adaptation to human cells, is that it might have arisen during
370 the passage of a precursor virus in the laboratory, either during work done
371 in human cells grown in cell culture or during serial passage in animal
372 models such as humanized mice. Neither of these possibilities can be ruled
373 out, but some evidence has accumulated to suggest that this may not have
374 taken place. A bovine coronavirus was reported to have acquired a 12nt
375 insert encoding the four amino acids SRRR during passage in human cells
376 (Borucki et al., 2013), but on closer inspection of the data it emerged that
377 this was not the case, as the variant carrying the insert was already present
378 and had been selected for during passage, presumably because it
379 conferred an advantage to the virus when grown in human cells.
380

381 An interesting and more recent study of serial passage looked at the bat
382 virus BANAL-20-236, a *sarbecovirus* that was sampled from bats in Laos
383 (Temmam et al., 2022) and is one of the closest known relatives of SARS-
384 CoV-2. The sequence of the Spike protein around the S1/S2 junction is a
385 close (but inexact) match to the sequence present in SARS-CoV-2.
386 BANAL-20-236 is not able to infect human airway epithelial cells, but by
387 growing this virus in human intestinal cells in culture, it was possible to
388 propagate the virus and to look for evidence of adaptation to human cells.
389 Although evidence was found for point mutations in the RBD during
390 passage in cell culture, a FCS did not emerge from these experiments.
391 Serial passage experiments performed with BANAL-20-236 in humanized
392 mice produced similar results (Temmam et al., 2023). These results do not
393 support the popular theory that the FCS arose during serial passage in
394 culture, but are consistent with ideas that have been proposed regarding
395 the adaptation of the RBD in a laboratory setting (Sirotkin & Sirotkin, 2020).
396

397

Was the FCS of SARS-CoV-2 Engineered?

It is clear from the DEFUSE grant proposal released in 2021 that the UNC-WIV-EHA group was planning to extend their ongoing studies on proteolysis (Menachery et al., 2020) as a factor in controlling viral entry, and that they aimed to do so by identifying novel FCS sequences and inserting these into newly identified coronaviruses (Lerner, 2021), and this proposal might obviously include work on unreported viruses that have not been disclosed. As with the arguments for natural origin, the main criticism of the idea that lab manipulation was involved is that the precursor virus (the "template") necessary for such engineering experiments has not been identified. It is worth noting here that BANAL-20-52, BANAL-20-236 and RaTG13, although the closest relatives to SARS-CoV-2*, are simply too different at the nucleotide level* to have served as the precursor for either route of origin (**Figure 8a**).

The experiments proposed by EHA and their partners in DEFUSE would seem to represent a logical extension of the work of many virologists, in particular the work done on proteolysis as a factor determining host range, and their long-standing interest in the FCS as a critical determinant of viral entry. A lack of transparency regarding this research has amplified concerns that the FCS of the virus might have a laboratory origin (Segreto and Deigin, 2020; Chan and Zhao, 2022). Insertion of the FCS by engineering is technically very simple to achieve and has already been performed many times, for example, with SARS-CoV-1 in pseudovirus experiments (Watanabe et al., 2008; Belouazard et al., 2009; Winstone et al., 2021).

An unusual *BsaX* I restriction site is found in SARS-CoV-2, bracketing the P2-P12 residues, and contained within the extended FCS (**Figure 8b**). This interesting observation, made by many observers, is consistent with the idea that the FCS could have been inserted in a lab. In fact, the much-maligned proline (P681) found at the P5 position, adjacent to RRAR, is both consistent with, and obligatory for, the insertion of a *BsaX* I site at this position. This unusual restriction site then provides for a potential application of what is termed "Golden mutagenesis", in which *any ten nucleotides* can be inserted 3'- to the CUCC sequence, resulting in *any three amino acids* being inserted between P5 (P681) and the conserved P1 arginine (R685). Golden mutagenesis is one application of "Golden Gate" cloning (Engler and Marillonet, 2014) using "type II" restriction enzymes (REs), a group that includes not only *BsaX* I (Tengs et al., 2004), but also

438 *Bsa* I and *BsmB* I, which have been extensively used in the design and
439 recovery of infectious clones of coronaviruses (Hou et al., 2020).
440
441 The use and application of type II REs was predicted and discussed by the
442 world's most accomplished coronavirus biologist several years ago, in
443 relation to potential biowarfare and bioterrorist activities (Baric, 2007).
444 There is some concern within the broader scientific community that similar
445 experiments might have led to the creation of a virus closely related to
446 SARS-CoV-2, a possibility that was foreseen by experts in biosecurity
447 many years earlier (Klotz and Sylvester, 2014). The summation of these
448 and other concerns about coronavirus engineering has led to calls by us
449 and others for the full disclosure of sequences, email communications and
450 laboratory notebooks, all as part of a detailed inquiry into the origins of the
451 virus (Relman, 2020, VanHelden et al., 2021; Harrison and Sachs, 2022a).
452

460 **Figures**

461 **Figure 1a**

462

463 **Figure 1b**

464

465 **Figure 1c**

466

467

468 **Figure 2a**

TMPRSS2 positive

Fusion with the plasma membrane

Attachment

Cleavage/priming

ACE2

TMPRSS2

Genome transcription/replication

469

470 **Figure 2b**

Golgi

trans-Golgi network

Endocytic pathway
• Diphtheria toxin
• Shiga toxin
• *Pseudomonas* exotoxin A

TGN/Biosynthetic pathway
• Pro-β-NGF
• Pro-BMP-4
• Insulin pro-receptor
• Ebola Zaire pro-GP

Early endosomes

Furin

Plasma membrane

Cell surface
• Anthrax toxin PA
• Proaerolysin
• *Clostridium* α-toxin

471

472 **Figure 3a**

473

474
475

476 **Figure 3b**

477

478
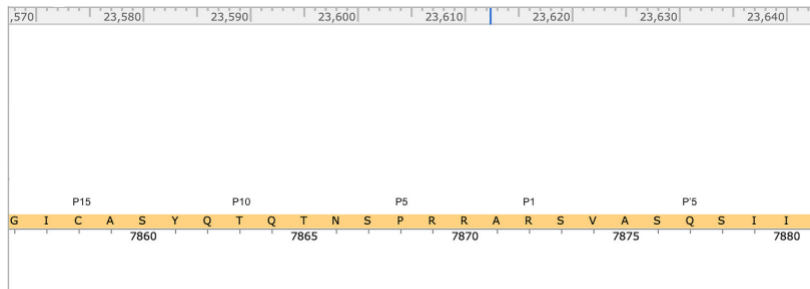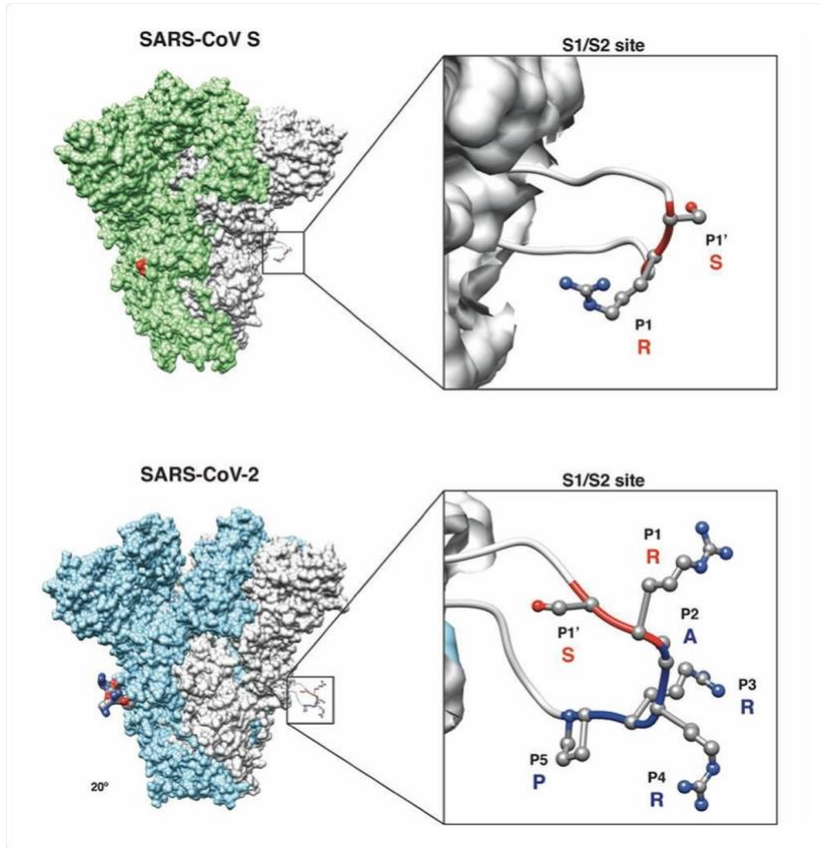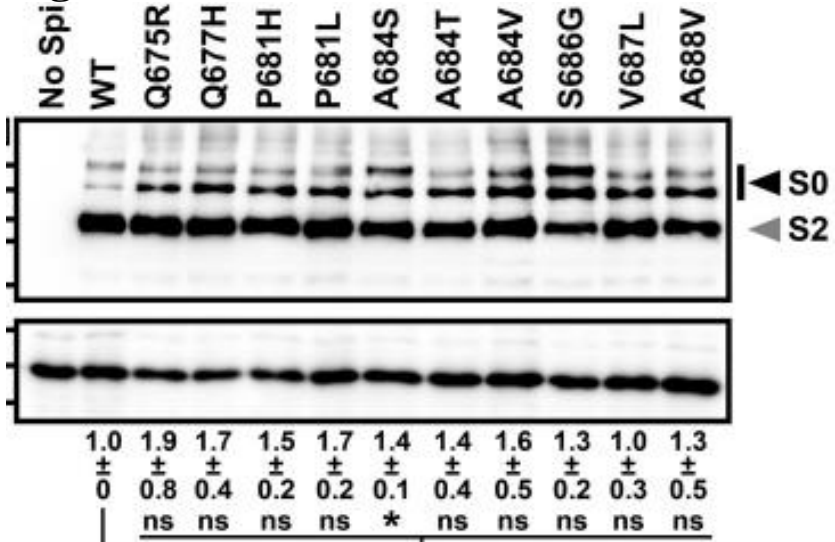
Figure 4a

Figure 4b
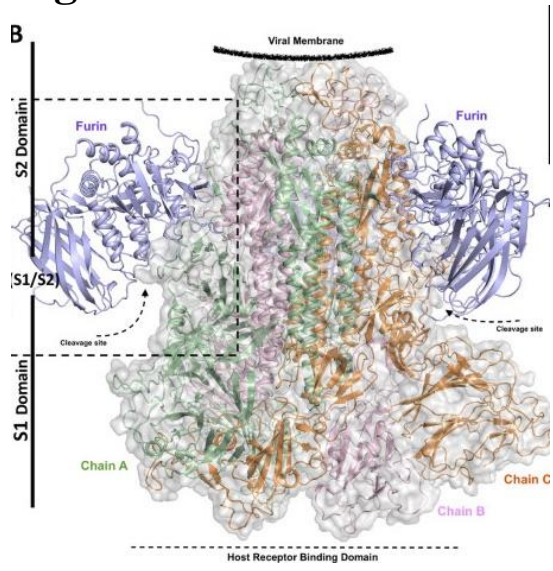
486 **Figure 5a**

487
488

489 **Figure 5b**
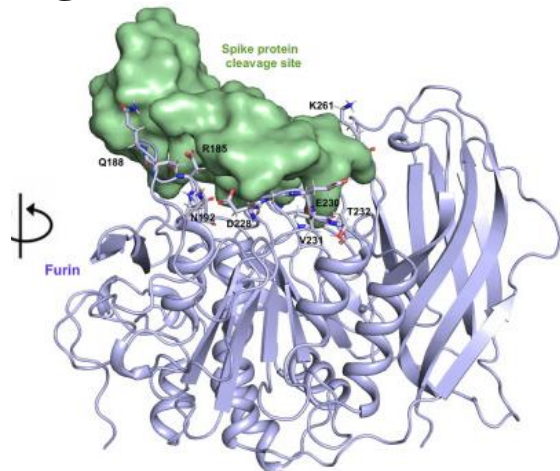
490
491 **Figure 5c**
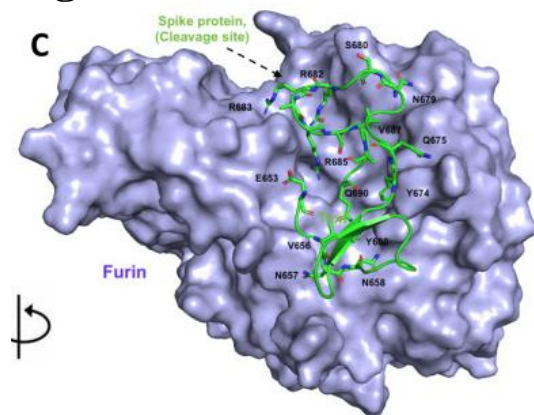
492

493 **Figure 6a**
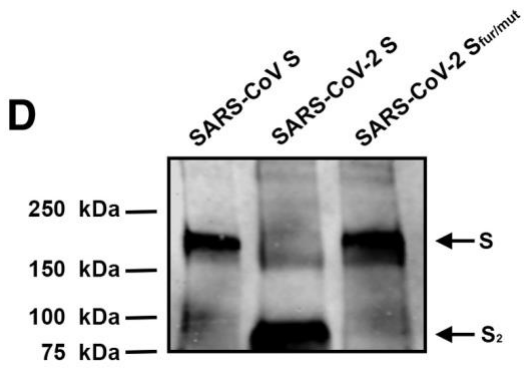
494
495

496 **Figure 6b**
497
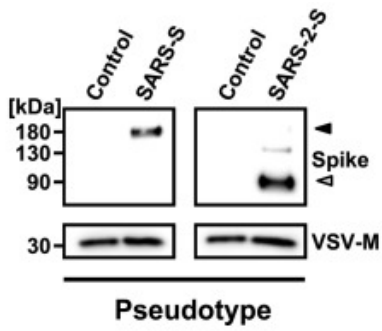
498
499

500 **Figure 6c**
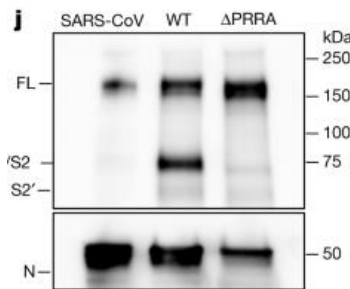
501
502

503 **Figure 6d**

504

505 **Figure 7a**

506

Asn440 / His440   Ser373 / Ser373   Ser371   Tyr369

Phe505 / His505   Arg403 / Thr403   Gln493 / Tyr493

507

508

509 **Figure 7b**

510

A   Response (nm) vs [ACE2] (nM)   Human   Bat

B   ACE2   Leu79   Met82   Tyr83   Leu486 / Phe486   Phe28   Bat   RBD   Human

C   ACE2   Glu35   Lys31   Gln493 / Tyr493   Glu484 / Thr484   Bat   RBD   Human

511

512

513   # Figure 8a

514

SARS-CoV-2 · RaTG13 sequence alignment showing CONSERVED REGION and VARIABLE REGION
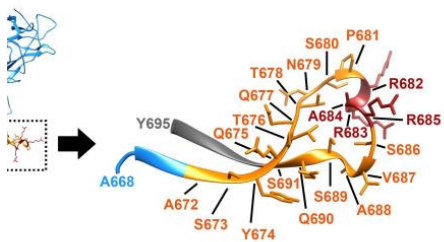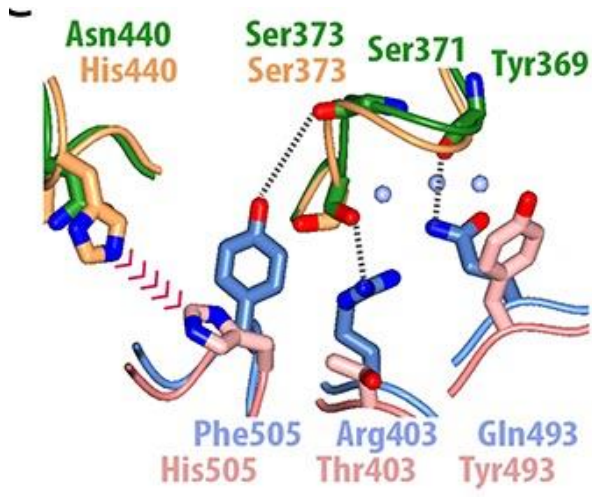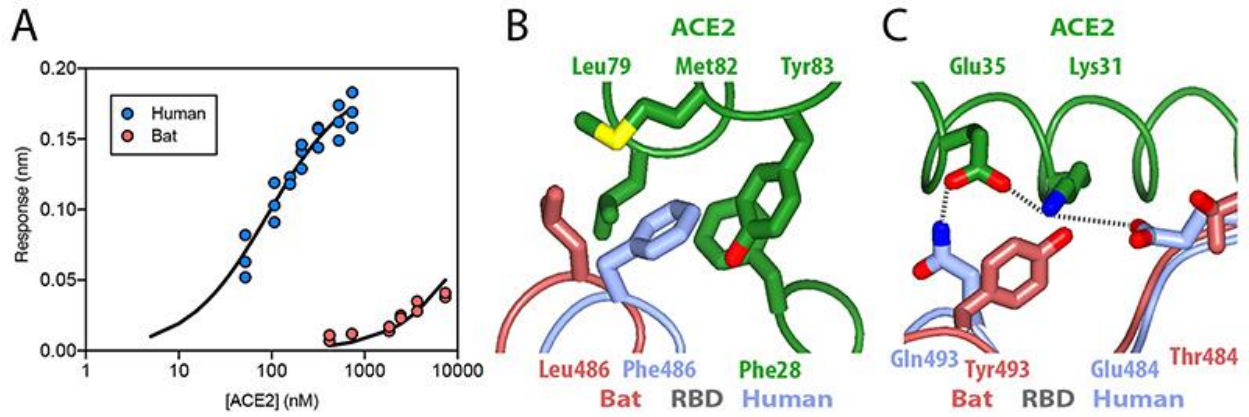
	CONSERVED REGION	VARIABLE REGION

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

515

516

517   # Figure 8b

518
519
520
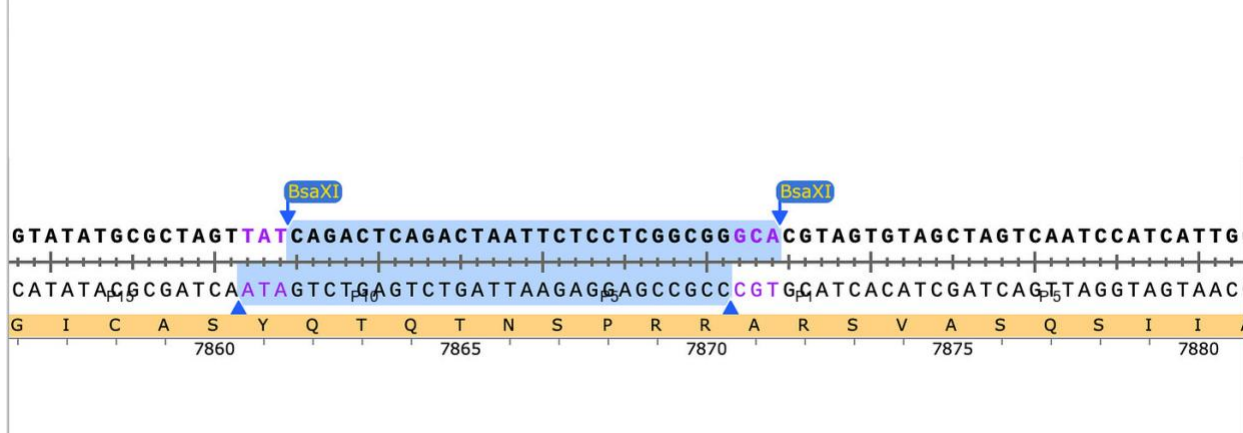
Figure Legends

Figure 1. a. A schematic of the Spike protein of SARS-CoV-2, showing the receptor binding domain and the two protease cleavage sites for furin and TMPRSS2. From Keller et al. (2022). b. The amino acid sequences at the S1-S2 junction of the Spike proteins of five viruses, with the FCS of SARS-CoV-2 highlighted in red. From Coutard et al. (2020). c. Amino acid sequence alignments around the S1-S2 junction of the Spike proteins of many members of the sub-genus *Sarbecovirus*, with the FCS of SARS-CoV-2 highlighted in red. Adapted from Hoffmann et al. (2020b).

Figure 2. a. The direct fusion pathway for SARS-CoV-2 entry is facilitated by TMPRSS2. Adapted from Hoffmann and Pöhlmann (2021) b. The cell biology of furin, showing its maturation in the Golgi, transport from the TGN to the plasma membrane and subsequent uptake by endocytosis. Adapted from Thomas (2002).

Figure 3. a. A schematic of the generic extended furin cleavage site, illustrating the core sequence flanked by two solvent-accessible regions. Adapted from Tian (2009). b. The amino acid sequence of the extended FCS in the Spike protein of SARS-CoV-2, labeled using the FCS numbering convention in Figure 3a.

Figure 4. The FCS of SARS-CoV-2 is an extended structure in a solvent-accessible region of the Spike protein structure. a. A comparison between the S1/S2 junction of SARS-CoV-1 and SARS-CoV-2. The insertion of PRRA extends the loop structure relative to the analogous region of the SARS-CoV-1 Spike protein, shown for comparison in these models. The peptide bond between R and S is exposed to proteases. From Jaimes et al. (2020). b. A model of the extended furin cleavage loop of SARS-CoV-2 from A668 to Y695, showing the highly conserved *sarbecovirus* sequence C671-Y674 (CASY), with the R682-R685 (RRAR) recognition sequence highlighted. From Arora et al. (2022).

555 **Figure 5.** a. The interaction between molecules of the Spike protein and furin (in
556 lilac), showing the location of the cleavage site protruding into the solvent. From
557 Venkadari (2020). b. The catalytic domain of furin (lilac) has several
558 electronegative residues (D228, N193, E230 etc..) that can make electrostatic
559 interactions with the furin cleavage site of the spike protein. From Venkadari
560 (2020) c. The extended FCS (green) is overlaid over the surface of the furin
561 catalytic domain (lilac). Positively charged R682, R683 and R685 interact with
562 furin while P681 projects away from the enzyme. The peptide bond between R685
563 and S686 is where the enzyme catalyzes the proteolysis of the Spike protein. From
564 Venkadari (2020).

565

566 **Figure 6.** a. The Spike protein of SARS-CoV-2 is subject to efficient proteolysis in
567 cells infected with the virus, while the SARS-CoV-1 Spike is not. Mutation of the
568 FCS (Sfur/mut) abolishes proteolysis. From Hoffmann et al. (2020b). b. The Spike
569 protein of SARS-CoV-2 is subject to proteolysis in cells infected with VSV
570 pseudovirus, while the SARS-CoV-1 Spike is not. Mutation of the FCS (Sfur/mut)
571 abolishes proteolysis. From Walls et al. (2020). c. The Spike protein of SARS-
572 CoV-2 is subject to proteolysis in cells infected with a pseududovirus, while the
573 SARS-CoV-1 Spike is not. Deletion of the FCS (-PRRA) abolishes proteolysis.
574 From Peacock et al. (2021).

575

576 **Figure 7.** a. Subtle differences in the structure of the Spike protein RBD from
577 SARS-CoV-2 (blue) and the related bat virus RaTG13 (pink), illustrating
578 unfavorable interactions between His 505 of RaTG13 and its receptor (green).
579 From Wrobel et al. (2020) b. The RBD of the SARS-CoV-2 virus binds to hACE2
580 with high (nanoMolar) affinity, while the RBD of the bat virus binds weakly to
581 ACE2, due in part to the lack of Phe 486, Glu 484 and 493 in the bat virus. From
582 Wrobel et al. (2020).

583

584 **Figure 8.** a. Sequence alignment between the Spike proteins of SARS-CoV-2 and
585 RaTG13 in the region of the FCS. The enumeration of the FCS begins at C, the
586 P15 residue in the conserved domain. Despite the sequence identity of the amino
587 acid sequence the divergent nucleotide sequence suggests that RaTG13 was neither
588 the immediate evolutionary precursor of SARS-CoV-2, nor a laboratory template.
589 From Deigin and Segreto (2021). b. A restriction enzyme site for *BsaX* I is located
590 within the extended FCS region and flanks the conserved 6 amino-acid sequence
591 QTQTNS, as well as the 4 amino-acid insert PRRA.

592

593

594

References

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

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

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

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

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

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

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

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

9. Dahms SO, Arciniega M, Steinmetzer T, Huber R, Than ME. (2016) Structure of the unliganded form of the proprotein convertase furin

suggests activation by a substrate-induced mechanism. Proc Natl Acad Sci U S A. 113: 11196-11201.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

24. Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T, Dinnon KH 3rd, Kato T, Lee RE, Yount BL, Mascenik TM, Chen G, Olivier KN, Ghio A, Tse LV, Leist SR, Gralinski LE, Schäfer A, Dang H, Gilmore R, Nakano S, Sun L, Fulcher ML, Livraghi-Butrico A, Nicely NI, Cameron M, Cameron C, Kelvin DJ, de Silva A, Margolis DM, Markmann A, Bartelt L, Zumwalt R, Martinez FJ, Salvatore SP, Borczuk A, Tata PR, Sontake V, Kimple A, Jaspers I, O'Neal WK, Randell SH, Boucher RC, Baric RS. (2020) SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. Cell 182:429-446.e14.

25. Hughey RP, Bruns JB, Kinlough CL, Harkleroad KL, Tong Q, Carattino MD, Johnson JP, Stockand JD, Kleyman TR. (2004) Epithelial sodium

channels are activated by furin-dependent proteolysis. J Biol Chem. 279:18111-18114.

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

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

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

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

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

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

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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. (2022) Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. Nature. 603:706-714.

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

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

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

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

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

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

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

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

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

787 47. Segreto R, Deigin Y. The genetic structure of SARS-CoV-2 does not
788   rule out a laboratory origin: SARS-COV-2 chimeric structure and furin
789   cleavage site might be the result of genetic manipulation. Bioessays.
790   2021 Mar;43(3):e2000240.

791 48. Seidah NG, Day R, Marcinkiewicz M, Chrétien M. (1998) Precursor
792   convertases: an evolutionary ancient, cell-specific, combinatorial
793   mechanism yielding diverse bioactive peptides and proteins. Ann N Y
794   Acad Sci. 839:9-24.

795 49. Seidah NG & Prat A. (2012) The biology and therapeutic targeting of
796   the proprotein convertases. Nature Reviews Drug Discovery, 11: 367-
797   383.

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

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

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

809 53. Temmam S, Montagutelli X, Herate C, Donati F, Regnault B, Attia M,
810   Baquero Salazar E, Chretien D, Conquet L, Jouvion G, Pipoli Da
811   Fonseca J, Cokelaer T, Amara F, Relouzat F, Naninck T, Lemaitre J,
812   Derreudre-Bosquet N, Pascal Q, Bonomi M, Bigot T, Munier S, Rey
813   FA, Le Grand R, van der Werf S, Eloit M. (2023) SARS-CoV-2-related
814   bat virus behavior in human-relevant models sheds light on the origin
815   of COVID-19. EMBO Rep. 6:e56055.

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

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

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

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

830 58. Venkadari N (2020). Structure of Furin Protease Binding to SARS-
831   CoV-2 Spike Glycoprotein and Implications for Potential Targets and
832   Virulence. The Journal of Physical Chemistry Letters 11: 6655-6663.

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

840 60. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D.
841   (2020) Structure, Function, and Antigenicity of the SARS-CoV-2
842   Spike Glycoprotein. Cell. 181:281-292.e6

843 61. Watanabe R, Matsuyama S, Shirato K, Maejima M, Fukushi S, et al.
844   (2008) Entry from the Cell Surface of Severe Acute Respiratory
845   Syndrome Coronavirus with Cleaved S Protein as Revealed by
846   Pseudotype Virus Bearing Cleaved S Protein. J Virol 82: 11985–
847   11991.

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

852 63. Wolf KA, Kwan JC, Kamil JP. (2022) Structural Dynamics and
853   Molecular Evolution of the SARS-CoV-2 Spike Protein. mBio.
854   13:e0203021.

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

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

862 66. Wu Y, Zhao S (2021) Furin cleavage sites naturally occur in
863   coronaviruses. Stem Cell Research, 50, 102115.

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

868
869