



Evolutionary insights into the furin cleavage sites of SARS-CoV-2 variants from humans and animals

Abdou Nagy¹ · Shereen Basiouni² · Rokshana Parvin³ · Hafez M. Hafez⁴ · Awad A. Shehata^{5,6}

Received: 28 April 2021 / Accepted: 19 May 2021 / Published online: 13 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

The SARS-CoV-2 spike protein Q677P/H mutation and furin cleavage site (FCS) have been shown to affect cell tropism and virus transmissibility. Here, we analyzed the frequency of Q677P/H and FCS point mutations in 1,144,793 human and 1042 animal spike protein sequences and from those of the emergent variants B.1.1.7, B.1.351, P.1, B.1.429 + B.1.427, and B.1.525, which were deposited in the database of the GISAID Initiative. Different genetic polymorphisms, particularly P681H and A688V, were detected in the FCS, mainly in human isolates, and otherwise, only pangolin and bat sequences had these mutations. Multiple FCS amino acid deletions such as $\Delta^{680}\text{SPRRA}^{684}$ and $\Delta^{685}\text{RSVA}^{688}$ were only detected in eight and four human isolates, respectively. Surprisingly, deletion of the entire FCS motif as $\Delta^{680}\text{SPRRARSVA}^{688}$ and $\Delta^{680}\text{SPRRARSVAS}^{689}$ was detected only in three human isolates. On the other hand, analysis of FCS from emergent variants showed no deletions in the FCS except for spike P681del, which was detected in seven B.1.1.7 isolates from the USA. Spike Q677P was detected only once in variant, B.1.1.7, whereas Q677H was detected in all variants, i.e., B.1.1.7 (n = 1938), B.1.351 (n = 28), P.1 (n = 9), B.1.429 + B.1.427 (n = 132), and B.1.525 (n = 1584). Structural modeling predicted that mutations or deletions at or near the FCS significantly alter the cleavage loop structure and would presumably affect furin binding. Taken together, our results show that Q677H and FCS point mutations are prevalent and may have various biological effects on the circulating variants. Therefore, we recommend urgent monitoring and surveillance of the investigated mutations, as well as laboratory assessment of their pathogenicity and transmissibility.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). The virus emerged in Wuhan, China, in December 2019 and has had a serious impact on global health. The World Health Organization (WHO) in early 2020 declared COVID-19 to be a pandemic [1]. At the time of writing, more than 3,100,121 deaths and 124,068,643 million confirmed positive cases in humans have been reported globally, making it the most contagious pandemic in the past decade. SARS-CoV-2 is a member of the family *Coronaviridae*, subfamily *Orthocoronavirinae*, genus *Betacoronavirus* [33]. In 2002 and 2003, another betacoronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in China and spread to 29 countries worldwide, affecting more than 8000 people. Ten years later, Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in the Arabian Peninsula and spread mainly in the Middle East, affecting more than 2000 people [2, 3].

Handling Editor: Pablo Pineyro.

✉ Abdou Nagy
abdonagy@zu.edu.eg

✉ Awad A. Shehata
awad.shehata@vet.usc.edu.eg

¹ Department of Virology, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Sharkia 44511, Egypt

² Clinical Pathology Department, Faculty of Veterinary Medicine, Benha University, Benha, Egypt

³ Department of Pathology, Faculty of Veterinary Science, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh

⁴ Institute of Poultry Diseases, Faculty of Veterinary Medicine, Free University, Berlin, Germany

⁵ Avian and Rabbit Diseases Department, Faculty of Veterinary Medicine, Sadat City University, Sadat City, Egypt

⁶ Research and Development Section, PerNaturam GmbH, Gödenroth, Germany

SARS-CoV-2 is an enveloped, single-stranded positive-sense RNA virus with a genome of 29,903 nucleotides that contains 11 open reading frames (ORFs) encoding 27 proteins. ORF1ab constitutes the first two-thirds of the genome (21,290 nucleotides) and encodes 16 non-structural proteins (nsp1–nsp16). The last third of the genome encodes the structural proteins (spike [S], envelope [E], matrix [M], and nucleocapsid [N]) and six accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10) [4].

The spike (S) glycoprotein plays a pivotal role in viral infection and pathogenesis, and it is composed of two functional subunits: S1 and S2. The S1 subunit consists of an N-terminal domain (NTD) and a receptor-binding domain (RBD), which binds to human angiotensin-converting enzyme 2 (ACE-2) [5, 6]. The S2 subunit contains a fusion peptide, heptad repeat 1 (HR1), a central helix, a connector domain, heptad repeat 2 (HR2), a transmembrane domain (TM), and a cytoplasmic domain (CD), and it mediates fusion of the viral envelope with the host cell membrane [7, 8]. At the boundary between the S1 and S2 subunits, there is a furin cleavage site (FCS) containing multiple basic residues [8–11] that are not found in SARS-CoV-related coronaviruses but are present in the human coronaviruses OC43, HKU1, and MERS-CoV [12]. Cleavage of the S protein by the cellular protease furin, which is ubiquitously expressed in multiple organs and tissues in humans, mediates the entry of SARS-CoV-2 into host cells and expands the tropism of the virus [8].

Like other coronaviruses and RNA viruses, SARS-CoV-2 quickly acquires new mutations, leading to the emergence of new SARS-CoV-2 variants [12–14]. Recently, the SARS-CoV-2 variants B.1.1.7, B.1.351, and B.1.1.207 emerged independently in the United Kingdom, South Africa, and Nigeria, respectively [15]. Mutations in the S glycoprotein can alter the efficiency of binding to the ACE-2 receptor and viral fitness. The D614G mutation is thought to increase virus transmissibility [16, 17], while the N501Y mutation has also been implicated in the rapid spread of SARS-CoV-2 [15, 18]. Additionally, some mutations may enable immune evasion and reduce vaccine efficacy [19]. Moreover, mutations in or close to the FCS, such as Q677P/H, are common in the course of coronaviruses evolution and adaptation, particularly when the virus adapts to a different host. Interestingly, up to April 2020, the FCS region of the 122 SARS-CoV-2 genome sequences available in the GISAID Initiative database were strictly conserved [8]. However, in June 2020, a total of 103 out of 45,828 SARS-CoV-2 genome sequences had mutations in the FCS [20]. Therefore, to understand the pathogenicity of SARS-CoV-2, it is important to characterize these mutations and investigate their potential impact on virus tropism and transmissibility. Here, we analyzed the evolutionary patterns of Q677P/H and FCS amino acids residues (amino acid positions 680–689)

from all animal and human variants and five major emergent variants for which whole-genome SARS-CoV-2 sequences had been deposited in the GISAID Initiative database since the start of the COVID-19 pandemic.

Materials and methods

Frequency of spike Q677P/Q677H and FCS mutations

Human and animal isolates

Data on the spread of mutants containing amino acid (aa) changes at positions 680–689 in the FCS of SARS-CoV-2 were collected as described previously [21]. All human and animal SARS-CoV-2 sequences that were deposited at the GISAID Initiative (<https://www.gisaid.org/help/search/>) during the period of January 10, 2020 to April 18, 2021 were analyzed. The search for each FCS mutation in human and animal SARS-CoV-2 isolates was done separately using the GISAID Initiative default search criteria, including the name of the mutation (e.g., Spike_P681H in section Variant/Subs) and the host name, e.g., Human and/or *Neovison vison* and *Mustela lutreola* for mink). The GISAID Initiative was selected for data collection, as it is the only database with all up-to-date sequences from different animal species and humans from every country. Some reference strains (Wuhan-Hu-1, Bat-RaTG13, and Bat-HKU9-1) were downloaded from the NCBI database for comparison. In addition, data on multiple FCS motif mutations, particularly those with a high frequency in human isolates worldwide, were also investigated.

Emergent variants

Major emergent variants of SARS-CoV-2, including B.1.1.7 (n = 353,925), B.1.351 (n = 9,271), P.1 (n = 3,546), B.1.429 + B.1.427 (n = 27,757), and B.1.525 (n = 1,604), with sequences available in the GISAID Initiative database were analyzed for single and multiple deletions. Moreover, the spike mutations Q677H and Q677P were also analyzed in the five variants.

Structural modeling of the S protein

Following the GISAID Initiative workflow, multiple amino acid sequences of selected human and animal SARS-CoV-2 strains were aligned. The aligned region consisted of 1273 aa, including the signal peptide (SP), NTD, RBD, and RBM of the S1 subunit and the fusion peptide (FP), HR1, HR2, TM, and CD of the S2 subunit. To better understand the structural effects of FCS

variations, a previously published and validated model of the full-length S glycoprotein that was refined and modeled based on a cryo-EM structure was selected from the Swiss-Model repository [PODTC2 -SPIKE_SARS COV 2; 7DK6 "S-2H2-F2 structure, two RBDs are up and one RBD is down, each up RBD binds with a 2H2 Fab."] (<https://swissmodel.expasy.org/repository/uniprot/PODTC2>). Furthermore, the structure of the FCS region (aa positions 541-788) was modeled using the aa sequences of a human strain (NC_045512.2 SARS CoV2- isolate Wuhan-Hu-1) and three animal strains from pangolin and bat (EPI_ISL_410721, hCoV-19-like/pangolin/Guangdong/1/2019; ABN10911.1, bat coronavirus HKU9-1; QHR63300.2, bat coronavirus RaTG13). Strains with other mutations at or near the FCS (position 681 and 677) were also used for modeling. Models were generated using the online Swiss-Model tools (<https://swissmodel.expasy.org/interactive/58Ub4Q/models/>). The models were viewed and decorated using the DeepView software incorporated in the Swiss-Model package.

Results

Frequency and tracing of spike FCS mutations and Q677P/Q677H

Human and animal isolates

Non-synonymous mutations, including substitutions and/or deletions, in the spike FCS motif (aa position 680-689) of human isolates ($n = 1,144,793$) and animal isolates ($n = 1042$) of SARS-CoV-2 with sequences available in the GISAID Initiative database were investigated. Interestingly, the recently detected P681H substitution was the most prevalent FCS mutation, detected in 395,711 human isolates worldwide (Fig. 1). Furthermore, P681H was detected in isolates from different animal species, including dog ($n = 3$), cat ($n = 2$), monkey ($n = 2$), lion ($n = 1$), tiger ($n = 1$), and leopard ($n = 1$), and was found in environmental samples ($n = 167$) (Table 1). A688V was the second-most prevalent FCS mutation among human isolates ($n = 3273$), with wide geographic distribution (Fig. 1), whereas it was detected only once in



Fig. 1 World map shows geographic spread of highly prevalent FCS mutations that were detected in SARS-CoV-2 human isolates. The map was created with BioRender.com.

Table 1 Frequency of critical SARS-CoV-2 S mutations and deletions in the FCS

S mutation	Human isolates	Animal isolates
S680P*	14	0.0
S680del	14	2 from pangolins (China) 1 from a bat (China)
S680F	298	2 from pangolins (China)
P681H	395,711	Dog (n = 3), cat (n = 2), monkey (n = 2), lion (n = 1), tiger (n = 1), leopard (n = 1), environment (n = 167)
P681R	3,023	0.0
P681S	71	0.0
P681L	443	Cat (n = 1)
P681del	21	22 (18 from pangolins, 4 from bats)
R682G	2	0.0
R682Q	11	0.0
R682W	20	0.0
R682del	12	23 (18 from pangolins, 5 from bats)
R683G	1	0.0
R683P	2	0.0
R683del	8	22 (18 from pangolins, 4 from bats)
A684E	3	0.0
A684S	75	1 from mink (Denmark)
A684T	112	0.0
A684V	684	0.0
A684del	10	17 (14 from pangolins, 3 from bats)
S686R	6	1 from bat China
S686del	7	0.0
V687I	265	0.0
V687del	4	0.0
A688S	308	12 from pangolins, bat (n = 1), environment (n = 2)
A688V	3273	Environment (n = 1)
A688del	4	0.0
S689I	619	Environment (n = 2)
$\Delta^{680}\text{SP}^{681}$	8	4
$\Delta^{682}\text{RR}^{683}$	8	18
$\Delta^{680}\text{SPR}^{682}$	8	4
$\Delta^{680}\text{SPRR}^{683}$	8	4
$\Delta^6\text{68 A R685}$	8	0.0
$\Delta\text{684 AdRS686}$	8	0.0
$\Delta\text{687VA688}$	4	0.0
$\Delta\text{687 VAS689}$	1	0.0
$\Delta^{680}\text{SPRRA}^{684}$	8	0.0
$\Delta^{685}\text{RSVA}^{688}$	4	0.0
$\Delta^{680}\text{SPRRARSVA}^{688}$	3	0.0
$\Delta^{680}\text{SPRRARSVAS}^{689}$	3	0.0

*The covered mutations are located at amino acid positions 680-688 of the spike protein.

an environmental sample and was not detected in animal isolates (Table 1). Other genetic polymorphisms, S680F, S680P, P681L, P681R, P681S, A684V, A684T, V687I, A688S, and S689I, were found in isolates 298, 14, 443, 3,023, 71, 684, 112, 265, 308, and 619, respectively, from humans, with wide geographic distribution (Fig. 1).

Major FCS mutations in animal isolates were mainly detected in pangolins and bats, whereas very few mutations were detected in other animal species, such as A684S, which was detected once in a Danish mink isolate. On the other hand, 13 FCS mutations, namely, S680P, S680F, P681H, P681R, P681S, P681L, R682W, A684T, A684V, V687I,

A688S, A688V, and S689I, among over 1,144,793 human isolates had a high frequency, with wide geographic distribution (Fig. 1). In addition, deletions (Δ) at different positions in the FCS motif were also detected in both human and animal SARS-CoV-2 isolates, such as Δ S680, which was detected in eight human isolates from the USA, Europe, Asia, and Australia, whereas it was only detected in pangolin isolates (n = 4) and bat isolates (n = 1) from China. The Δ P681, Δ R682, Δ R683del, and Δ A684 deletions were also detected in human isolates (n = 8) from the same geographic locations (Fig. 1), whereas four of these deletions were detected in animal isolates (Table 1). Interestingly, Q677P and Q677H have not been detected in animal isolates, whereas Q677H has been detected twice in environmental samples (USA and Europe).

Emergent variants

Single and/or multiple deletions in the FCS of major currently circulating SARS-CoV-2 variants, including B.1.1.7 (n = 353,925), B.1.351 (n = 9,271), P.1 (n = 3,546), B.1.429 + B.1.427 (n = 27,757), and B.1.525 (n = 1,604), were analyzed. Δ P681 was detected in seven B.1.1.7 isolates from the USA, whereas no FCS deletions were detected in other

variants. Q677P was detected once in an isolate of variant B.1.1.7, whereas mutation Q677H was detected in all variants: B.1.1.7 (n = 1938), B.1.351 (n = 28), P.1 (n = 9), B.1.429 + B.1.427 (n = 132), and B.1.525 (n = 1584).

Conformational changes in the cleavage loop due to mutations and deletions

A diagram of the functional organization of the S protein and its FCS region is shown in Fig. 2. Different regions in the FCS, particularly F1, located between the S1 and S2 domains of the spike protein, are highlighted in a separate box showing the positions of FCS mutations in SARS-CoV-2 strains from humans and animals (Fig. 2, Table 1). Structural models were made to examine the effects of a few selected F1 FCS mutations and nearby polymorphisms (Q677P/H) in human and animal strains (Fig. 3). Furin binding occurs in a clamp-like manner, where it clips to the cleavage site of the S glycoprotein. The overall complex structure shows the S glycoprotein homotrimer with the middle or equatorial S1/S2 region at the adjacent side of the spike trimer (Fig. 3A). The F1 FCS region (aa 541-788) with different mutations at positions Q681-P/H/R and Q677-P/H detected in human SARS-CoV-2 strains, including the emergent variant strains described above as well

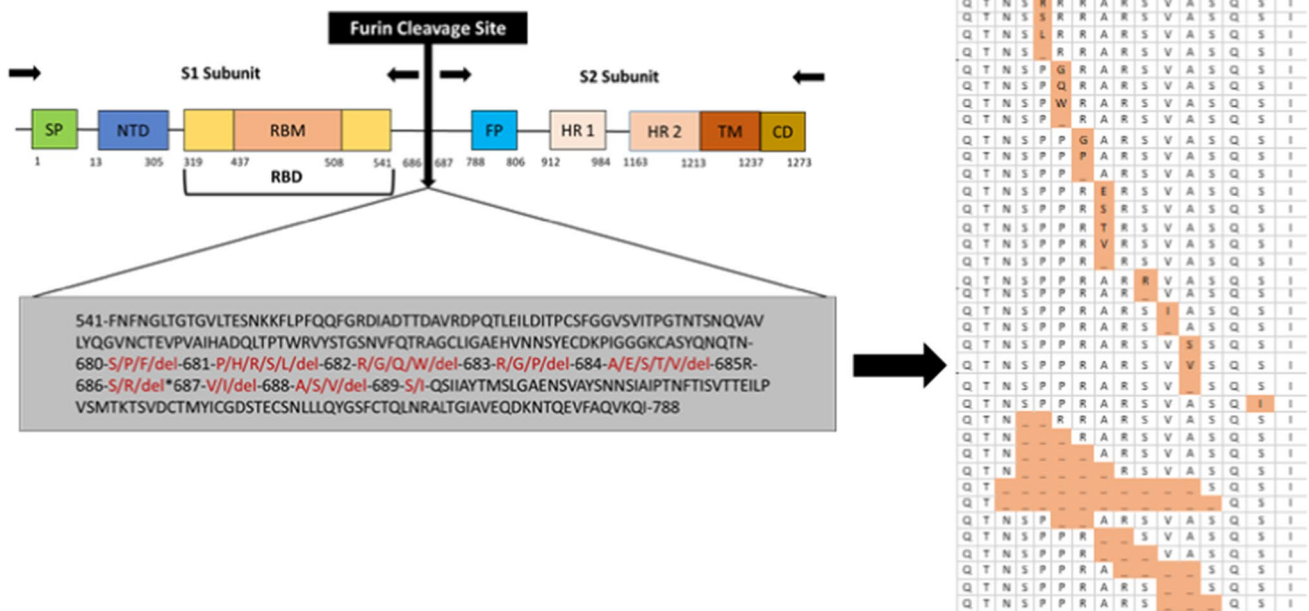
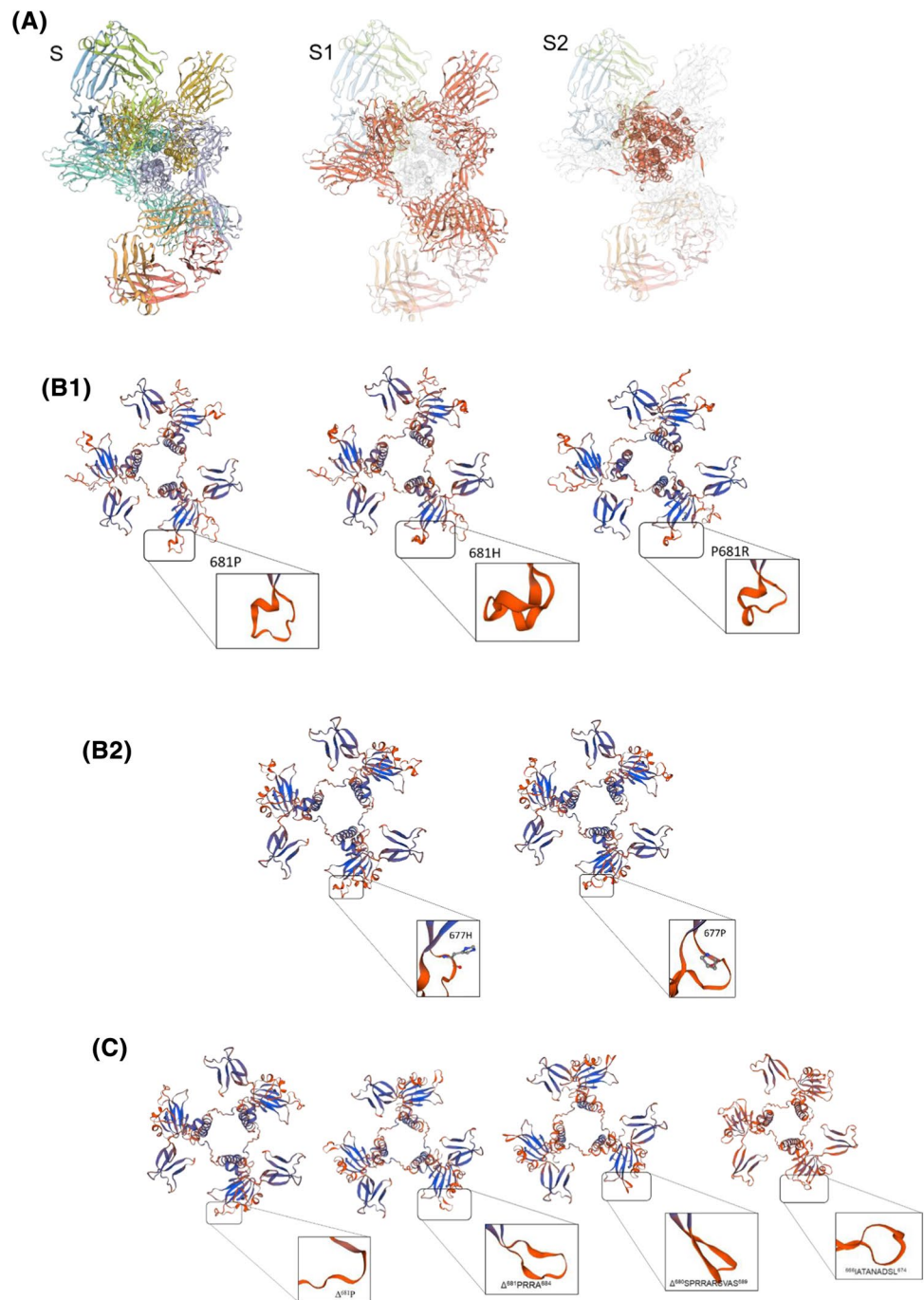


Fig. 2 A) Genomic sketch of spike protein (1-1273aa) with highlighted close view of FCS region (aa position 541-788) and the reported mutations at FCS (680-689; red color) detected in sequences of both human and animal isolates/strains available at GISAID data-

base. B) Polymorphisms in the FCS of the SARS-CoV-2 spike glycoprotein. The colored sequences indicate mutations in the FCS. Green colored sequences indicating the sequence of the FCS of the reference SARS-CoV-2 virus.

Fig. 3 QMEAN and cartoon representation of crystal structure of Furin cleaved SARS-CoV-2 spike glycoprotein homotrimer as unbound or closed conformation. **A** Full-length spike protein with S1/S2 conformation. **B** Cleavage loop (FCS) variations due to mutation at position 681-P/H/R found in human strains. **C** Differences in tertiary structure of the FCS loop (magnified within the box) due to polymorphisms in animal strains



as an environmental sample (GISAID ID EPI_ISL_1036658), showed an extended loop structure with a different conformation (Fig. 3B1 and B2). In addition, aa deletions near or within the FCS found in U.S. variant B1.1.7 strains (Δ^{681P}), pangolin strains ($\Delta^{680SPRRARSVAS^{689}}$), and two bat strains (RaTG13: $\Delta^{681PRRA^{684}}$ /HUK9-1: $^{666}IATANADSL^{674}$) changed its structural conformation, resulting in an altered loop structure (Fig. 3C). These models of different FCS variants reveal that mutations affect the depth, shape, and charge of the FCS, which may have a biological and clinical impact.

Moreover, the docking structure reveals that SARS-CoV-2 spike glycoprotein aa residues N657–Q690 are the key residues interacting with furin protease.

Discussion

The FCS ($^{680SPRRARSVAS^{689}}$) in the S glycoprotein is important for the pathogenesis and evolution of SARS-CoV-2, and changes in this region might enable immune

evasion and reduce vaccine efficacy. Therefore, analysis of the novel FCS mutations in sequenced SARS-CoV-2 genomes and emergent variants is urgently needed for studying their potential impact on the virulence and pathogenesis. In the present study, SARS-CoV-2 S protein sequences from human, animal, and emergent variants that were deposited in the GISAID Initiative database from January 10, 2020 to April 18, 2021 were analyzed to assess the ongoing molecular evolution of the virus.

Most of SARS-CoV-2 sequences isolated from humans and animals have the polybasic motif $^{682}\text{RRAR}^{685}$ in the FCS. This polybasic motif has been shown to broaden cell tropism, increase virus transmissibility, and promote efficient virus entry into host cells [6, 22, 23]. Cleavage of the S glycoprotein into the S1 and S2 subunits is mediated by host proteases, such as furin, which is widely expressed in a variety of organs and tissues, including brain, liver, pancreas, and epithelial cells of the respiratory, gastrointestinal, and reproductive tracts [24, 25]. Initial analysis of the FCS region of the available SARS-CoV-2 sequences revealed the presence of naturally occurring polymorphisms, four of which were deletions, including single, double, and multiple aa deletions (Fig. 1 and Table 1).

Deep analysis of SARS-CoV-2 S protein sequences revealed multiple substitutions, including S680P/F, P681H/R/S/L, R682G/Q/W, R683G/P, A684E/S/T/V, S686R, V687I, and A688S/V (Table 1). In June 2020, 103 out of 45,828 SARS-CoV-2 sequences contained mutations in the FCS, including 18 unique non-synonymous point mutations, one deletion, and six insertions of a premature stop codon [20]. We did not find any point mutations at Arg685, which is in accordance with a previous report [20]. In addition, seven single deletions (ΔS680 , ΔP681 , ΔR682 , ΔR683 , ΔA684 , ΔS686 , ΔV687 , and ΔA688) were also found in the FCS of SARS-CoV-2 isolated from humans and animals. It is expected that mutations around Arg685 and Ser686 may affect the recognition of the cleavage site [20]. Although the P681H mutation was the most prevalent FCS mutation in human SARS-CoV-2 isolates, it was also detected in different animal species, including dog ($n = 3$), cat ($n = 2$), monkey ($n = 2$), lion ($n = 1$), tiger ($n = 1$), and leopard ($n = 1$), as well as in environmental samples ($n = 167$), indicating that this mutation may have an impact on virus transmissibility and fitness. Moreover, this mutation is also a characteristic of the newly emerged SARS-CoV-2 variants [26], which are defined by multiple mutations in the S glycoprotein ($\Delta\text{69-70}$, Δ144 , N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H). Although no accumulated mutations were detected in the FCS of the sequences analyzed, different polymorphisms in the FCS due to double or multiple mutations were detected in SARS-CoV-2 sequences isolated from animals and humans, with wide geographic distribution (Fig. 1).

It has been shown that the FCS mutant $\Delta\text{681PRRA684}$ has faster replication kinetics but reduced overall replication efficiency in a human respiratory cell line, and it is attenuated in both hamster and K18-hACE2 transgenic mouse models of SARS-CoV-2 pathogenesis when compared with the parental SARS-CoV-2 [27]. Interestingly, we found that the $\Delta^{680}\text{SP}^{681}$, $\Delta^{682}\text{RR}^{683}$, $\Delta^{680}\text{SPR}^{682}$, and $\Delta^{680}\text{SPRR}^{683}$ deletions were detected in eight SARS-CoV-2 sequences out of 1,144,793 whole-genome sequence from humans. However, $\Delta\text{680SP681}$ and $\Delta^{682}\text{RR}^{683}$ were detected in pangolin isolates, while $\Delta^{680}\text{SPR}^{682}$ and $\Delta^{680}\text{SPRR}^{683}$ were found in bat isolates. In addition, the natural $\Delta^{680}\text{SPRRARSVAS}^{689}$, $\Delta^{680}\text{SPRRARSVA}^{688}$, $\Delta^{685}\text{RSVA}^{688}$, and $\Delta^{680}\text{SPRRA}^{684}$ deletions were detected only in SARS-CoV-2 sequences from humans.

Furin is a well-studied and essential serine protease with a minimum target sequence of Arg(R)-X-X-Arg (R). It is necessary for infection by some influenza virus and coronavirus strains [28, 29]. Furin cleavage is a critical initial step in the entry of SARS-CoV-2 virus into susceptible cells. Changes in the furin recognition site can affect the pathogenicity of the virus, and previous studies have demonstrated their effect on SARS-CoV and MERS-CoV pathogenesis [28, 30]. Three FCSs have been identified in SARS-CoV-2 (F1-3). Cleavage at F1 converts the S protein precursor to the S1 and S2 subunits and facilitates virus-cell fusion. Cleavage at F2 modifies the S2 subunit and participates in the pathogenicity of the virus after cell entry. F3 functions through the NTD and promotes adhesion between the virus and the cell surface [31]. Biophysical analysis has shown that interactions between the viral spike glycoprotein and furin protease are mediated by multiple hydrogen bonds. The cleavage loop inserts into the canyon-like substrate-binding pocket of furin protease, and this determines the orientation of binding of the enzyme to the spike glycoprotein [32]. Our modeling results suggest that the mutations 681P/H/R and 677P/H result in conformational changes in the cleavage loop (Fig. 3). In addition, significant structural deviations were observed due to multiple deletions in FCS ($\Delta^{666}\text{IATANADSL}^{674}$; $\Delta^{681}\text{PRRA}^{684}$; $\Delta^{680}\text{SPRRARSVAS}^{689}$) detected from animal isolates, which would be expected to affect the interaction of furin with the S protein.

The insertion of multiple amino acids can result in changes in host specificity and pathogenic potential. Therefore, FCS polymorphisms may play an important role in SARS CoV-2 evolution. We attempted to establish a timeline of the origins of FCS mutations by tracing the emergence of mutations in sequenced genomes. However, we observed a slight increase in the frequency of some randomly chosen mutations (S680F, P681S, A684T, and V687I), particularly during the period of October 2020 to January 2021 compared to other time periods in 2020 (Fig. 2). In addition, aa substitutions at different FCS positions from human

SARS-CoV-2 isolates were not consistent over the time period of January 2020 to January 2021. This inconsistency might be due to variations in sequencing activity in different countries and in the number of genomic sequences submitted to the GISAID Initiative, particularly in the first two quarters of 2020. However, this point needs further investigation.

In conclusion, several genetic polymorphisms were detected in the FCS of SARS-CoV-2 isolated from humans and animals that were due to point mutations or deletions. Structural modeling revealed differences in the predicted three-dimensional structure of the cleavage loop both in human and animal strains that might affect the interaction of furin with the spike protein. The rapid ongoing molecular evolution of SARS-CoV-2 highlights the urgent need to assess the impact of these mutations on pathogenicity and transmissibility.

Funding This research did not receive any funding.

Declarations

Conflict of interest Authors declare that they have no competing interests.

Ethical approval The authors declare that they have complied with the ethical standards for collecting and analyzing data.

References

- Cucinotta D, Vanelli M (2020) WHO declares COVID-19 a pandemic. *Acta Biomed* 91:157–160. <https://doi.org/10.23750/abm.v91i1.9397>
- Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E et al (2020) Clinical, laboratory and imaging features of COVID-19: a systematic review and meta-analysis. *Travel Med Infect Dis* 34:101623. <https://doi.org/10.1016/j.tmaid.2020.101623>
- Yan Y, Chang L, Wang L (2020) Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): current status, challenges, and countermeasures. *Rev Med Virol* 30:e2106. <https://doi.org/10.1002/rmv.2106>
- Helmy YA, Fawzy M, Elasad A et al (2020) The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J Clin Med*. <https://doi.org/10.3390/jcm9041225>
- Conceicao C, Thakur N, Human S et al (2020) The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol* 18:e3001016. <https://doi.org/10.1371/journal.pbio.3001016>
- Shang J, Wan Y, Luo C et al (2020) Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* 117:11727–11734. <https://doi.org/10.1073/pnas.2003138117>
- Elasad A, Fawzy M, Basiouni S, Shehata AA (2020) Mutational spectra of SARS-CoV-2 isolated from animals. *PeerJ* 8:e10609. <https://doi.org/10.7717/peerj.10609>
- Walls AC, Park Y-J, Tortorici MA et al (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181:281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Andersen KG, Rambaut A, Lipkin WI et al (2020) The proximal origin of SARS-CoV-2. *Nat Med* 26:450–452. <https://doi.org/10.1038/s41591-020-0820-9>
- Li W (2020) Delving deep into the structural aspects of a furin cleavage site inserted into the spike protein of SARS-CoV-2: a structural biophysical perspective. *Biophys Chem* 264:106420. <https://doi.org/10.1016/j.bpc.2020.106420>
- Wrapp D, Wang N, Corbett KS et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367:1260–1263. <https://doi.org/10.1126/science.abb2507>
- Mercatelli D, Giorgi FM (2020) Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol* 11:1800. <https://doi.org/10.3389/fmicb.2020.01800>
- Pater AA, Bosmeny MS, Barkau CL et al (2021) Emergence and evolution of a prevalent new SARS-CoV-2 variant in the United States. *Genomics* 91:157
- Tegally H, Wilkinson E, Giovanetti M et al (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. <https://doi.org/10.1101/2020.12.21.20248640>
- Leung K, Shum MH, Leung GM et al (2021) Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill*. <https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106>
- Korber B, Fischer WM, Gnanakaran S et al (2020) Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182:812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>
- Plante JA, Liu Y, Liu J et al (2020) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. <https://doi.org/10.1038/s41586-020-2895-3>
- Davies NG, Abbott S, Barnard RC et al (2021) Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. <https://doi.org/10.1126/science.abg3055>
- Dumonteil E, Herrera C (2020) Polymorphism and selection pressure of SARS-CoV-2 vaccine and diagnostic antigens: implications for immune evasion and serologic diagnostic performance. *Pathogens*. <https://doi.org/10.3390/pathogens9070584>
- Xing Y, Li X, Gao X, Dong Q (2020) Natural polymorphisms are present in the furin cleavage site of the SARS-CoV-2 spike glycoprotein. *Front Genet* 11:783. <https://doi.org/10.3389/fgene.2020.00783>
- Rambaut A, Holmes EC, O’Toole Á et al (2021) Addendum: a dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. <https://doi.org/10.1038/s41564-021-00872-5>
- Hoffmann M, Kleine-Weber H, Pöhlmann S (2020) 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. <https://doi.org/10.1016/j.molcel.2020.04.022>
- Sasaki M, Uemura K, Sato A et al (2021) SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells. *PLoS Pathog* 17:e1009233. <https://doi.org/10.1371/journal.ppat.1009233>
- Bestle D, Heindl MR, Limburg H et al (2020) TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance*. <https://doi.org/10.26508/lsa.202000786>
- Jaimes JA, André NM, Chappie JS et al (2020) Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *J Mol Biol* 432:3309–3325. <https://doi.org/10.1016/j.jmb.2020.04.009>
- Maison DP, Ching LL, Shikuma CM, Nerurkar VR (2021) Genetic characteristics and phylogeny of 969-bp S gene sequence

- of SARS-CoV-2 from Hawaii reveals the worldwide emerging P681H mutation. *bioRxiv*. <https://doi.org/10.1101/2021.01.06.425497>
27. Johnson BA, Xie X, Kalveram B et al (2020) Furin cleavage site is key to SARS-CoV-2 pathogenesis. *Microbiology* 581:221
 28. Simmons G, Bertram S, Glowacka I et al (2011) Different host cell proteases activate the SARS-coronavirus spike-protein for cell–cell and virus-cell fusion. *Virology* 413:265–274. <https://doi.org/10.1016/j.virol.2011.02.020>
 29. Tse LV, Hamilton AM, Friling T, Whittaker GR (2014) A novel activation mechanism of avian influenza virus H9N2 by furin. *J Virol* 88:1673–1683. <https://doi.org/10.1128/JVI.02648-13>
 30. Millet JK, Whittaker GR (2014) Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A* 111:15214–15219. <https://doi.org/10.1073/pnas.1407087111>
 31. Jin X, Xu K, Jiang P et al (2020) Virus strain from a mild COVID-19 patient in Hangzhou represents a new trend in SARS-CoV-2 evolution potentially related to Furin cleavage site. *Emerg Microbes Infect* 9:1474–1488. <https://doi.org/10.1080/22221751.2020.1781551>
 32. Vankadari N (2020) Structure of furin protease binding to SARS-CoV-2 spike glycoprotein and implications for potential targets and virulence. *J Phys Chem Lett* 11:6655–6663. <https://doi.org/10.1021/acs.jpcclett.0c01698>
 33. Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.