



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



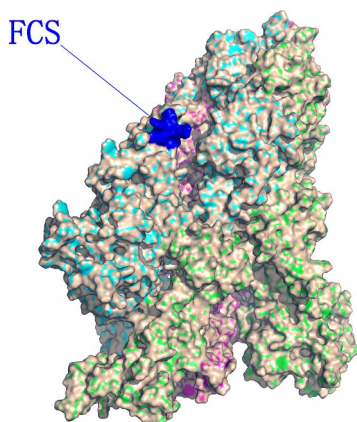
# Delving deep into the structural aspects of a furin cleavage site inserted into the spike protein of SARS-CoV-2: A structural biophysical perspective



Wei Li

*Institute of Special Environmental Medicine, Nantong University, No. 9, Seyuan Road, Nantong City, Jiangsu Province, People's Republic of China*

## GRAPHICAL ABSTRACT



## ABSTRACT

One notable feature of the SARS-CoV-2 genome, the spike (S) protein of SARS-CoV-2 has a polybasic furin cleavage site (FCS) at its S1-S2 boundary through the insertion of 12 nucleotides encoding four amino acid residues PRRA. Quite intriguingly, this polybasic FCS is absent in coronaviruses of the same clade as SARS-CoV-2. Thus, with currently available experimental structural data for S protein, this short article presents a set of comprehensive structural characterization of the insertion of FCS into S protein, and argues against a hypothesis of the origin of SARS-CoV-2 from purposeful manipulation: (1), the inserted FCS is spatially located at a random coil loop region, mostly distantly solvent-exposed (instead of deeply buried), with no structural proximity to the other part of the S protein; (2), the insertion of FCS itself does not alter, neither stabilize nor de-stabilize, the three-dimensional structure of S; (3), the net result here is the insertion of a furin cleavage site into S protein, whose S1 and S2 subunits will still be strongly electrostatically bonded together from a structural and biophysical point of view, even if the polybasic FCS is actually cleaved by furin protease before or after viral cell entry.

## 1. Introduction

The membrane of SARS-CoV-2 harbours a homotrimeric ([1]) transmembrane spike (S) glycoprotein, which is essential for the entry of virus particles into the cell. S protein contains two functional domains: a receptor binding domain (RBD) [2], and a second domain which contains sequences that mediate fusion of viral and cell

membranes [3–7]. Recently, it was reported that S protein contains a potential cleavage site for furin protease [8], including four residues (Pro681 (P681), Arg682 (R682), Arg683 (R683) and Ala684 (A684)) [9–13]. Functionally, R682, R683, A684 and Arg685 (R685) constitute the minimal polybasic furin cleavage site (FCS), i.e., RXYR, where X or Y is to be a positively charged arginine or lysine. With respect to the origin of this COVID-19 pandemic, of further interest is the fact that the

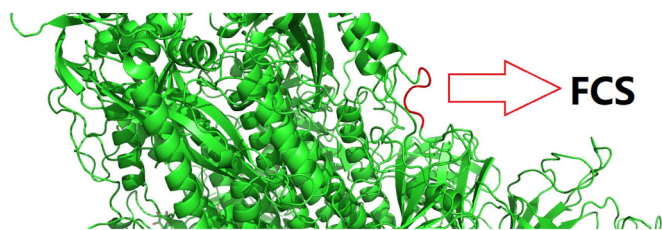
*E-mail address:* [liweiqidong@ntu.edu.cn](mailto:liweiqidong@ntu.edu.cn).

<https://doi.org/10.1016/j.bpc.2020.106420>

Received 9 May 2020; Received in revised form 27 May 2020; Accepted 13 June 2020

Available online 29 June 2020

0301-4622/ © 2020 Elsevier B.V. All rights reserved.



**Fig. 1.** An overall structure of the spike protein of SARS-CoV-2 in green cartoon. In this figure, the inserted FCS fragment from Chain B is coloured red. This figure is prepared using PyMol [30] with supplementary file model.pdb as an input. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

S protein has a specific FCS that is absent in coronaviruses of the same clade as SARS-CoV-2 [8,14–16].

In the midst of this COVID-19 pandemic, FCS is reportedly linked to a natural-selection, instead of purposeful-manipulation, –based hypothesis of the origin of this COVID-19 outbreak [9,17–21]. Regardless of this COVID-19-origin hypothesis [9], to date (Thu Jun 25 09:39:392020), it still remains not clear what the actual structural consequence is of the polybasic FCS's insertion into the S protein of SARS-CoV-2. Thus, this article delves into all currently available (as of Thu Jun 25 09:39:392020) structural data [22] of S protein and aims to uncover the structural impact of the insertion of FCS into the S protein of SARS-CoV-2.

## 2. Materials and methods

To begin with, QHD43416.1 is the GenBank access code of the surface glycoprotein of SARS-CoV-2 [Severe acute respiratory syndrome coronavirus 2]. With QHD43416.1, the amino acid sequence of S protein was retrieved and listed below,

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLH-STQDLFLPFSSNVTWFHAIHVSNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDLSDKTSLLLVNNAITNVVIVCEFCNDPFLGVYHYKNNKSWMESEFRVYSSANNCTFEYVSQPFMLDLEGGKQGNFKNLRFEVFNKNDGYFKIYSKHTPINLVRDLPGQFSALEPLVDLPIGINITRFQTTLLALHRSYLTGPDSSSGWTAGAAAAYVGYLQPRFTLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYA WNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRIAPGQTGKIADYNYKLPDDFTGCVIAWNSNLDLDSKVGNYNYLYRFRKSNLKPFFERDISTEIQAGSTPCNGVEGFNCYFPPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFFNGLTGGVLTESNKKFLPFQFGFRDIADTTDAVRDPQTLLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNSYECDIPIGAGICASYQTQNSPRRARSVASQSIAYTMSLGAENSVAYSNNIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDIESTECSNLLQYGSFCTQLNRALTGLAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSPKSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLCAQKFNGLTVLPLLTDEMIAYTSALLAGTITSGWTFGAGALQIPFAMQMA YRFNGIGVQTNVLYENQKLIANQFNSAIGKIQDLSSTASALGKLDQVNVNQAQALNLT VKQLSSNFGAIVSVDLNDLRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGGSKRVDFCGKGYHLSMFPQSAPHGVVFLHVTVYVPAQEKNFITAPAICHGDKAHFPREGVVFVSNNGTHWFVTQRNFYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFNKHTSPDVLGDISGINASVNIQKEIDRLNEVAKNLNESLIDLQELGKYEYQIKWPWYIWLGFIAGLIAIVMTIMLCCMTSCCSCLKGCSCGSCCKFDEDDSEPVLGKVKLHYT.

First, the sequence above was plugged into the SwissModel homology modelling [23] server in search of an experimental structural model of the S protein of SARS-CoV-2. In principle, it is expected that the experimental S protein structure is complete, i.e., with no experimentally uncharted territories (EUTs) [24]. In fact, however, the structural search led to a Cryo-EM structure (PDB ID: 6VSB) with a

range of EUTs [24], representing the three-dimensional structure of the prefusion 2019-nCoV spike glycoprotein with a single receptor-binding domain up [1]. From an amino acid sequence alignment (Fig. 1.2 in supplementary file supplement.pdf), it is obvious that those EUTs are widely scattered throughout the homotrimeric structure (PDB ID 6VSB) [1]. Yet, with another amino acid sequence alignment (Fig. 1.1 in supplementary file supplement.pdf), it was revealed that the sequence similarity between the Cryo-EM structure (PDB ID 6VSB) and QHD43416.1 is as high as 99.26%, making PDB ID 6VSB rather suitable to be used as a structural template for the subsequent homology structural modelling of the S protein of the SARS-CoV-2. Therefore, the Cryo-EM PDB ID 6VSB structure was used as the template by the SwissModel [23] homology modelling server to build a structural model with as less EUTs [24] as possible of the S protein of SARS-CoV-2.

Afterwards, the UCSF Chimera software [25] was employed to add hydrogen atoms to the homology homotrimeric structural model of the S protein, which was subsequently subject to a comprehensive set of electrostatic interaction analysis as described in [26] previously, and also a set of solvent accessible surface area (SASA) analysis by DSSP [27,28]. According to the DSSP-calculated SASA values, an amino acid residue of a protein can be classified as either buried or exposed according to a comparison with its standard SASA value contained in the “standard.data” file (supplementary file standard.data) available with the Naccess [29] software distribution.

## 3. Results

### 3.1. A structural electrostatic analysis of the homotrimeric structural model of the S protein of SARS-CoV-2

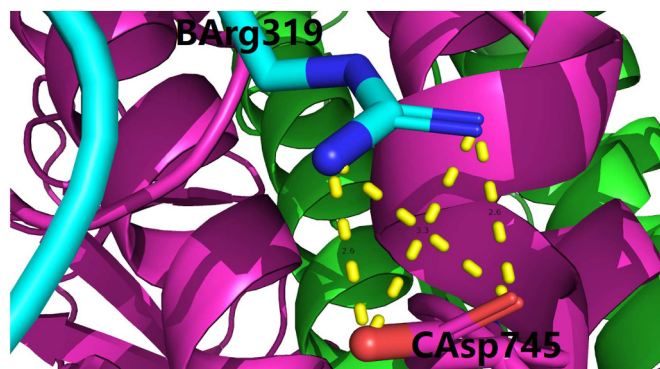
With SwissModel [23] and UCSF Chimera [25], a homology homotrimeric structural model (supplementary file model.pdb) of the spike protein of SARS-CoV-2 was built with only two EUTs (Fig. 1.3 in supplementary file supplement.pdf) [24], one at its N-terminal and another at its C-terminal, as compared (Fig. 1.2 in supplementary file supplement.pdf) with the widely scattered EUTs [24] inside its Cryo-EM structural template PDB ID 6VSB [1]. Subsequently, with the electrostatic analysis as described previously in [26], this short article puts forward a comprehensive set of structural electrostatic interaction analysis (supplementary file supplementary.pdf) for the homotrimeric structural model (supplementary file model.pdb) of the spike protein of SARS-CoV-2.

With a close inspection of all tables in the supplementary file supplementary.pdf,

1. no salt bridge or hydrogen bond was structurally identified for the three basic residues at FCS.
2. no hydrogen bond was structurally identified for Arg682 or Pro681.
3. one hydrogen bond was structurally identified for Ala684 with Arg685.
4. two hydrogen bonds were structurally identified for Arg683 with Thr604 and Asn679.
5. four hydrogen bonds were structurally identified for Arg685 with Val687, Lys310 and Ala684.

Further details of the hydrogen bonds were included in Table 1, where the first four (1, 2, 3 and 4 in Table 1) hydrogen bonds were formed between the side chains of the FCS residues and the main chain (backbone) oxygen atoms of the other part of the S protein, thus only affecting the backbone oxygen atoms in the structure of the S protein. In light of the double covalent bond formed between those affected backbone oxygen atoms and their neighbouring backbone carbon atoms, it is unlikely that the first four (1, 2, 3 and 4 in Table 1) hydrogen bonds are able to cause a major conformational perturbation of the structure of the S protein of SARS-CoV-2.

On the other hand, however, the last (hydrogen bond No. 5 in



**Fig. 2.** Four inter-chain salt bridges formed between basic residue pair B\_ARG\_319 and C\_ASP\_745 (Table 5). This figure is prepared using PyMol [30] with supplementary file model.pdb as an input. In this figure, the three (A, B and C) chains of the spike protein of SARS-CoV-2 is shown as green, cyan and purple cartoons, the residue naming scheme is Chain ID, residue name, residue number, the four salt bridges are shown here with four dotted yellow lines, with the lengths of the four inter-chain salt bridges being 2.6, 2.6, 3.3 and 3.5 Å, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1) hydrogen bond was formed between the main chain (backbone) oxygen atom of the FCS residue (Arg685) and the positively charged side chain of Lys310, for which a comprehensive set of structural electrostatic interaction analysis is included in Tables 2 and 3 below.

From the structurally identified salt bridges and hydrogen bonds (Tables 2 and 3), it is clear that Lys310 and Asp663 forms a set of strong electrostatic interactions for all three chains (A, B and C), making it unlikely that the 5th (Table 1) hydrogen bond is able to disrupt the basic residue pair Lys310-Asp663 and induce a major conformational change for the structure of the S protein of SARS-CoV-2.

To sum up, the inserted FCS is only involved in a set of weak electrostatic interactions (five hydrogen bonds, Table 1) within the S protein of SARS-CoV-2, whose overall scaffold is not to be altered, neither stabilized nor de-stabilized, by the insertion of the polybasic FCS.

### 3.2. An SASA analysis of the homotrimeric structural model of the S protein of SARS-CoV-2

To delve deep into the structural impact of FCS's insertion into the S protein of SARS-CoV-2, this short article takes a close look at the spatial location of FCS in the overall scaffold (supplementary file model.pdb) of the S protein of SARS-CoV-2.

As shown in Fig. 1, the inserted FCS is spatially located at a random coil loop region with no structural proximity to other part of the structure of the S protein. In light of the structural electrostatic analysis above, this visual observation (Fig. 1) constitutes further structural evidence that the inserted FCS does not alter, neither stabilize nor de-stabilize, the overall scaffold of the S protein of SARS-CoV-2.

To further investigate the structural impact of the inserted FCS, a set

**Table 1**

A summary of side chain and main chain hydrogen bonding analysis between the polybasic FCS and other part of the S protein of SARS-CoV-2. In this table, the residue naming scheme is Chain ID\_residue name\_residue number,  $\angle ADH$  represents the angle formed by acceptor (A), donor (D) and hydrogen (H) ( $\angle ADH$ ).

Hydrogen bond No.	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH$ (°)
1	O, A_THR_604	NE, A_ARG_683	HE, A_ARG_683	2.90	1.95	15.53
2	O, C_ASN_679	NH1, C_ARG_683	HH11, C_ARG_683	2.73	1.73	8.46
3	O, B_VAL_687	NH1, B_ARG_685	HH11, B_ARG_685	2.75	1.91	27.04
4	O, B_VAL_687	NE, B_ARG_685	HE, B_ARG_685	2.86	1.99	25.09
5	O, C_ARG_685	NZ, C_LYS_310	HZ3, C_LYS_310	2.70	1.83	24.75

**Table 2**

Lys310-specific salt bridging analysis of the structure of the S protein. In this table, the residue naming scheme is Chain ID\_residue name\_residue number.

Salt bridge	Residue A	Atom A	Residue B	Atom B	Distance (Å)
1	A_LYS_310	NZ	A_ASP_663	OD1	2.628
2	A_LYS_310	NZ	A_ASP_663	OD2	2.667
3	B_LYS_310	NZ	B_ASP_663	OD2	2.615
4	B_LYS_310	NZ	B_ASP_663	OD1	2.691
5	C_LYS_310	NZ	C_ASP_663	OD1	3.129

of quantitative SASA analysis (supplementary file model.dssp) by DSSP [27,28] was conducted for all three chains (A, B and C) in the structure of the S protein (supplementary file model.pdb). Specific details are included in Table 4 for residues at FCS as below,

In addition to Table 4, subsequent quantitative analysis demonstrated that the average relative accessibilities of the FCS are 56.54%, 64.55% and 58.93% for all three (A, B and C) chains in the homotrimeric [1] structure of the S protein of SARS-CoV-2, respectively. By and large, this result classifies the polybasic FCS as exposed, instead of buried, in all three chains of the homology structural model (supplementary file model.pdb). This quantitative relative accessibility analysis, along with the visual inspection of Fig. 1, indicates that the net structural consequence of FCS is the insertion of a furin cleavage site into SARS-CoV-2 S protein, leading to a question further: what if the inserted polybasic FCS is actually cleaved by furin protease?

### 3.3. What if the inserted polybasic FCS is actually cleaved by furin protease?

To answer this question, a similar set of electrostatic interaction analysis [26] was conducted for the trimeric [1] structural model of the S protein, with the difference here being that the whole S protein structure was splitted into two parts: the first part consists of three (chains A, B and C) S1 structural fragments, i.e., before the polybasic FCS, while the second part consists of three (chains A, B and C) S2 structural fragments, i.e., after the polybasic FCS. The results of this FCS-specific electrostatic interaction analysis are included in Tables 5, 6 and 7 as below.

From Tables 5, 6 and 7, it is clear that a series of electrostatic interactions (including both salt bridges and hydrogen bonds) still exist at the interface of the two parts of the S protein generated from the hypothesized cleavage by furin proteases. For instance, a total of four inter-chain salt bridges still exist at the interface of chains B and C of the S protein, as shown in Fig. 2. For another instance, not only does Arg319 of chain A form three inter-chain salt bridges with Asp737 and Asp745 of chain B (Table 5), Arg319 of chain A also forms a side chain hydrogen bond with Asp737 of chain B (Table 7). Among all electrostatic interactions listed in Tables 5, 6 and 7, these seven inter-chain salt bridges and one side chain hydrogen bond are merely two examples, which constitute a set of strong electrostatic forces towards the structural stabilization of the overall homotrimeric [1] structure of the S protein, even if the inserted polybasic FCS is actually cleaved by furin protease either before or after viral cell entry.



**Table 3**

Lys310-specific hydrogen bonding network analysis of the structure of the S protein. In this table, the residue naming scheme is Chain ID\_residue name\_residue number,  $\angle ADH$  represents the angle formed by acceptor (A), donor (D) and hydrogen (H) ( $\angle ADH$ ).

Hydrogen bond No.	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH$ (°)
1	OD1, A_ASP_663	NZ, A_LYS_310	HZ2, A_LYS_310	2.63	1.74	22.20
2	OD2, B_ASP_663	NZ, B_LYS_310	HZ3, B_LYS_310	2.62	1.67	16.87
3	O, C_ARG_685	NZ, C_LYS_310	HZ3, C_LYS_310	2.70	1.83	24.75
4	OD1, A_ASP_663	NZ, A_LYS_310	HZ2, A_LYS_310	2.63	1.74	22.20
5	OD2, B_ASP_663	NZ, B_LYS_310	HZ3, B_LYS_310	2.62	1.67	16.87

**Table 4**

A quantitative analysis by DSSP [27] of the SASA values and the relative SASA values of the FCS residues. The relative SASAs are calculated for each amino acid in the protein by expressing the various summed residue accessible surfaces as a ratio (Value/Reference) of that observed in a ALA-X-ALA tripeptide built using the QUANTA molecular graphics software package [29].

Residue ID	Chain ID	Residue name	Residue name	Value (Å <sup>2</sup> )	Reference (Å <sup>2</sup> )	Value/reference	Notes
681	A	P	PRO	40	136.13	0.29	Buried
682	A	R	ARG	214	238.76	0.90	Solvent-exposed
683	A	R	ARG	193	238.76	0.81	Solvent-exposed
684	A	A	ALA	17	107.95	0.16	Buried
685	A	R	ARG	79	238.76	0.33	Buried
681	B	P	PRO	112	136.13	0.82	Solvent-exposed
682	B	R	ARG	175	238.76	0.73	Solvent-exposed
683	B	R	ARG	214	238.76	0.90	Solvent-exposed
684	B	A	ALA	59	107.95	0.55	Solvent-exposed
685	B	R	ARG	60	238.76	0.25	Buried
681	C	P	PRO	135	136.13	0.99	Solvent-exposed
682	C	R	ARG	93	238.76	0.39	Buried
683	C	R	ARG	168	238.76	0.70	Solvent-exposed
684	C	A	ALA	21	107.95	0.19	Buried
685	C	R	ARG	149	238.76	0.62	Solvent-exposed

#### 4. Conclusion

Incorporating currently available structural data (as of Thu Jun 25 09:39:402020) of the S protein [22], and in the hope of pushing a little bit forward the boundary of our scientific knowledge on COVID-19, this short article presents a comprehensive structural characterization of the FCS inserted into the S protein of SARS-CoV-2, and puts forward a set of structural analysis as below,

1. the polybasic FCS is only involved in set of weak electrostatic interactions, and is therefore not able to alter, neither stabilize nor destabilize, the overall scaffold of the S protein of SARS-CoV-2.
2. the polybasic FCS is spatially located at a random coil loop region, mostly distantly solvent-exposed (Fig. 1, instead of deeply buried), with no structural proximity to the other part of the S protein of SARS-CoV-2.
3. the S1 and S2 subunits of the S protein will still be strongly bonded together, at least electrostatically [31] from a structural and biophysical point of view, even if the FCS is actually cleaved by furin protease.

#### 5. Discussion

Quite recently, it has been reported that this polybasic FCS is essential for SARS-CoV-2 to infect human lung cells, and that campaigns to develop therapeutics against SARS-CoV-2 should include the evaluation of furin inhibitors [14,16]. In view of the reported in vitro functional relevance of FCS [16], it is postulated here that the polybasic FCS does need to be cleaved by furin protease, leading to a different action mechanism of SARS-CoV-2 from its siblings, where its S1 and S2 subunits (while still strongly bonded together) undergoes a major structural rearrangement before or after viral cell entry. Nevertheless,

**Table 5**

FCS-specific inter-part salt bridging analysis of the S protein of SARS-CoV-2 with FCS at the boundary of its S1 and S2 units. In this table, the residue naming scheme is Chain ID\_residue name\_residue number. In this table, for residues A and B, one of them is located upstream (residue ID smaller than 681) of the polybasic FCS, while the other is located downstream (residue ID larger than 685) of the polybasic FCS.

PDB File name	Residue A	Atom A	Residue B	Atom B	Distance (Å)
Model.pdb	A_ARG_319	NH1	B_ASP_737	OD2	2.735
Model.pdb	A_ARG_319	NH1	B_ASP_745	OD1	3.587
Model.pdb	A_ARG_319	NH2	B_ASP_737	OD2	3.982
Model.pdb	B_ARG_319	NH1	C_ASP_745	OD1	2.606
Model.pdb	B_ARG_319	NH1	C_ASP_745	OD2	3.565
Model.pdb	B_ARG_319	NH2	C_ASP_745	OD1	3.325
Model.pdb	B_ARG_319	NH2	C_ASP_745	OD2	2.636
Model.pdb	B_LYS_854	NZ	A_ASP_614	OD1	3.487
Model.pdb	B_LYS_986	NZ	C_ASP_427	OD2	3.985
Model.pdb	C_ARG_319	NH1	A_ASP_745	OD1	3.819
Model.pdb	C_ARG_646	NH1	A_ASP_848	OD2	3.643
Model.pdb	C_ARG_646	NH2	A_ASP_848	OD1	3.695
Model.pdb	C_ARG_646	NH2	A_ASP_848	OD2	2.627
Model.pdb	C_ARG_847	NH1	B_GLU_619	OE1	3.034
Model.pdb	C_LYS_854	NZ	B_ASP_614	OD1	2.686

in the midst of this COVID-19 pandemic [32] with more than 5.5 million (as of Thu Jun 25 09:39:402020) confirmed cases globally, this short article puts forward a set of analysis that the net structural consequence of FCS here is the insertion of a furin cleavage site into the S protein of SARS-CoV-2, and is thus of only limited structural biophysical relevance here. Finally, along with [9], the structural biophysical analysis here makes even more unlikelier a purposeful-manipulation-based hypothesis of the origin of SARS-CoV-2.

**Table 6**

A summary of side chain and main chain hydrogen bonding analysis between the two structural fragments generated by the hypothesized cleavage at the polybasic FCS of the S protein of SARS-CoV-2. In this table, the residue naming scheme is Chain ID\_residue name\_residue number,  $\angle ADH$  represents the angle formed by acceptor (A), donor (D) and hydrogen (H) ( $\angle ADH$ ). In this table, for residues A and B, one of them is to be located upstream (residue ID smaller than 681) of the polybasic FCS, while the other is to be located downstream (residue ID larger than 685) of the polybasic FCS.

PDB file name	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH$ (°)
Model.pdb	OD2, B_ASP_737	NH1, A_ARG_319	HH12, A_ARG_319	2.74	1.78	15.71
Model.pdb	O, A_ILE_692	N, A_GLU_654	H, A_GLU_654	2.86	1.86	5.94
Model.pdb	O, A_THR_696	OG, A_SER_659	HG, A_SER_659	2.72	1.91	26.75
Model.pdb	O, B_PRO_863	N, A_ALA_668	H, A_ALA_668	2.78	1.79	9.02
Model.pdb	O, B_LEU_864	N, A_GLY_669	H, A_GLY_669	2.82	1.95	24.85
Model.pdb	O, A_GLU_654	N, A_ALA_694	H, A_ALA_694	2.80	1.83	12.75
Model.pdb	O, A_GLU_661	OH, A_TYR_695	HH, A_TYR_695	2.62	1.68	9.06
Model.pdb	O, A_GLY_669	N, A_MET_697	H, A_MET_697	2.86	1.87	9.86
Model.pdb	O, A_TYR_660	N, A_SER_698	H, A_SER_698	2.75	1.74	2.22
Model.pdb	O, C_ASP_614	NE, A_ARG_847	HE, A_ARG_847	2.97	2.12	27.07
Model.pdb	O, C_ASP_614	NH1, A_ARG_847	HH11, A_ARG_847	2.71	1.79	19.16
Model.pdb	OD2, C_ASP_614	OG1, A_THR_859	HG1, A_THR_859	2.83	1.95	19.70
Model.pdb	O, C_THR_547	ND2, A_ASN_978	HD21, A_ASN_978	2.99	2.12	24.76
Model.pdb	O, C_LEU_984	NZ, B_LYS_386	HZ3, B_LYS_386	2.72	1.89	28.40
Model.pdb	O, B_ILE_692	N, B_GLU_654	H, B_GLU_654	2.81	1.82	9.58
Model.pdb	O, B_ALA_694	N, B_VAL_656	H, B_VAL_656	2.96	2.02	17.75
Model.pdb	O, B_THR_696	OG, B_SER_659	HG, B_SER_659	2.72	1.93	28.63
Model.pdb	O, C_PRO_863	N, B_ALA_668	H, B_ALA_668	2.82	1.84	9.55
Model.pdb	O, C_LEU_864	N, B_GLY_669	H, B_GLY_669	2.81	1.90	20.41
Model.pdb	OH, B_TYR_695	N, B_SER_673	H, B_SER_673	2.96	2.09	24.82
Model.pdb	O, B_GLU_654	N, B_ALA_694	H, B_ALA_694	2.77	1.83	16.90
Model.pdb	O, B_GLY_669	N, B_MET_697	H, B_MET_697	2.77	1.77	5.38
Model.pdb	O, B_TYR_660	N, B_SER_698	H, B_SER_698	2.65	1.65	3.25
Model.pdb	OD2, A_ASP_614	OG1, B_THR_859	HG1, B_THR_859	2.98	2.16	26.02
Model.pdb	O, A_ARG_983	N, C_SER_383	H, C_SER_383	2.91	1.90	2.08
Model.pdb	OD2, A_ASP_745	OG1, C_THR_549	HG1, C_THR_549	2.94	2.12	25.69
Model.pdb	OD2, A_ASP_848	NH2, C_ARG_646	HH22, C_ARG_646	2.63	1.79	27.55
Model.pdb	O, C_ILE_692	N, C_GLU_654	H, C_GLU_654	2.81	1.89	19.55
Model.pdb	O, C_THR_696	OG, C_SER_659	HG, C_SER_659	2.86	2.07	28.50
Model.pdb	O, A_PRO_863	N, C_ALA_668	H, C_ALA_668	2.82	1.86	14.51
Model.pdb	O, A_LEU_864	N, C_GLY_669	H, C_GLY_669	2.73	1.84	22.00
Model.pdb	O, C_TYR_695	N, C_CYS_671	H, C_CYS_671	2.88	1.91	13.00
Model.pdb	OG, C_SER_689	OG1, C_THR_676	HG1, C_THR_676	2.79	1.86	11.33
Model.pdb	O, C_CYS_671	N, C_TYR_695	H, C_TYR_695	2.78	1.80	9.89
Model.pdb	O, C_GLU_661	OH, C_TYR_695	HH, C_TYR_695	2.57	1.63	8.69
Model.pdb	O, C_GLY_669	N, C_MET_697	H, C_MET_697	2.91	1.93	11.90
Model.pdb	OD1, B_ASP_614	NZ, C_LYS_854	HZ2, C_LYS_854	2.69	1.71	10.87
Model.pdb	O, B_THR_547	ND2, C_ASN_978	HD21, C_ASN_978	2.84	1.92	19.82

**Table 7**

A summary of side chain hydrogen bonding analysis between the two structural fragments generated by the hypothesized cleavage at the polybasic FCS of the S protein of SARS-CoV-2. In this table, the residue naming scheme is Chain ID\_residue name\_residue number,  $\angle ADH$  represents the angle formed by acceptor (A), donor (D) and hydrogen (H) ( $\angle ADH$ ). In this table, for residues A and B, one of them is to be located upstream (residue ID smaller than 681) of the polybasic FCS, while the other is to be located downstream (residue ID larger than 685) of the polybasic FCS.

PDB file name	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH$ (°)
Model.pdb	OD2, B_ASP_737	NH1, A_ARG_319	HH12, A_ARG_319	2.74	1.78	15.71
Model.pdb	OD2, C_ASP_614	OG1, A_THR_859	HG1, A_THR_859	2.83	1.95	19.70
Model.pdb	OD2, A_ASP_614	OG1, B_THR_859	HG1, B_THR_859	2.98	2.16	26.02
Model.pdb	OD2, A_ASP_745	OG1, C_THR_549	HG1, C_THR_549	2.94	2.12	25.69
Model.pdb	OD2, A_ASP_848	NH2, C_ARG_646	HH22, C_ARG_646	2.63	1.79	27.55
Model.pdb	OG, C_SER_689	OG1, C_THR_676	HG1, C_THR_676	2.79	1.86	11.33
Model.pdb	OD1, B_ASP_614	NZ, C_LYS_854	HZ2, C_LYS_854	2.69	1.71	10.87

## Appendix A. Supplementary data

## References

- [1] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, et al., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367 (6483) (2020 feb) 1260–1263.
- [2] J. Shang, G. Ye, K. Shi, Y.S. Wan, H. Aihara, F. Li, Structure of 2019-nCoV chimeric receptor-binding domain complexed with its receptor human ACE2, *Worldwide Protein Data Bank* (2020), <https://doi.org/10.2210/pdb6VW1/pdb>.
- [3] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (7798) (2020 feb) 270–273.
- [4] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, et al., A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020 feb) 265–269.
- [5] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, et al., Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group, (2020 feb).
- [6] S. Jiang, Z. Shi, Y. Shu, J. Song, G.F. Gao, W. Tan, et al., A distinct name is needed for the new coronavirus, *Lancet* 395 (10228) (2020 mar) 949.
- [7] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, *Lancet Infect. Dis.* 20 (5) (2020 feb) 533–534.
- [8] S. Tian, Q. Huang, Y. Fang, J. Wu, FurinDB: a database of 20-residue Furin cleavage site motifs, substrates and their associated drugs, *Int. J. Mol. Sci.* 12 (2) (2011 Feb) 1060–1065.
- [9] K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of SARS-CoV-2, *Nat. Med.* 26 (4) (2020 mar) 450–452.
- [10] X. Han, L.D. Bertzbach, M. Veit, Mimicking the passage of avian influenza viruses through the gastrointestinal tract of chickens, *Vet. Microbiol.* 239 (2019 dec) 108462.
- [11] L.V. Tse, A.M. Hamilton, T. Friling, G.R. Whittaker, A novel activation Mechanism of avian influenza virus H9N2 by Furin, *J. Virol.* 88 (3) (2013 nov) 1673–1683.
- [12] Y. Wan, J. Shang, R. Graham, R.S. Baric, F. Li, Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus, *J. Virol.* 94 (7) (2020 jan).
- [13] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Velesler, Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein, (2020 feb).
- [14] B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N.G. Seidah, E. Decroly, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade, *Antivir. Res.* 176 (2020 Apr) 104742.
- [15] H. Kleine-Weber, M.T. Elzayat, M. Hoffmann, S. Pöhlmann, Functional analysis of potential cleavage sites in the MERS-coronavirus spike protein, *Sci. Rep.* 8 (1) (2018 Nov).
- [16] M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, A multibasic cleavage site in the Spike protein of SARS-CoV-2 is essential for infection of human lung Cells, *Mol. Cell* 78 (4) (2020 May) 779–784.
- [17] M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, *Nat. Microbiol.* 5 (4) (2020 feb) 562–569.
- [18] V.D. Menachery, K.H. Dinnon, B.L. Yount, E.T. McAnarney, L.E. Gralinski, A. Hale, et al., Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection, *J. Virol.* 94 (5) (2019 dec).
- [19] T. Zhang, Q. Wu, Z. Zhang, Pangolin homology associated with 2019-nCoV, (2020 feb).
- [20] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020 feb) 497–506.
- [21] M.C. Wong, S.J.J. Cregeen, N.J. Ajami, J.F. Petrosino, Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019, (2020 feb).
- [22] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, *Nat. Struct. Mol. Biol.* 10 (12) (2003 dec) 980–980.
- [23] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, et al., SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.* 46 (W1) (2018 May) W296–W303.
- [24] W. Li, Visualising the Experimentally Uncharted Territories of Membrane Protein Structures inside Protein Data Bank, (2020 Jan).
- [25] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., UCSF chimera: a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (13) (2004) 1605–1612.
- [26] W. Li, How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLoS One* 12 (6) (2017 jun) e0178519.
- [27] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers.* 22 (1983) 2577–2637.
- [28] W. Kabsch, A solution for the best rotation to relate two sets of vectors, *Acta Crystallogr. A.* 32 (1976) 922–923.
- [29] S.J. Hubbard, J.M. Thornton, Naccess. Computer Program, 2 Department of Biochemistry and Molecular Biology, University College London, 1993.
- [30] W.L. DeLano, Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography*, 40 (2002), pp. 82–92.
- [31] C.B. Anfinsen, Principles that govern the folding of protein chains, *Science.* 181 (4096) (1973 Jul) 223–230.
- [32] E.M. Myers, Compounding Health Risks and Increased Vulnerability to SARS-CoV-2 for Racial and Ethnic Minorities and Low Socioeconomic Status Individuals in the United States, (2020 Apr).