

Article

A Furin Cleavage Site Inserted into the Spike Protein of SARS-CoV-2: A Structural Implication?

Wei Li *

¹ Institute of Special Environment Medicine, Nantong University, No. 9, Seyuan Road, Nantong City, Jiangsu Province, P. R. China

* Correspondence: wli148@aucklanduni.ac.nz

Received: date; Accepted: date; Published: date

Abstract: One notable features of the SARS-CoV-2 genome is that the spike protein of SARS-CoV-2 has a functional polybasic (furin) cleavage site (RRAR) at the S1–S2 boundary through the insertion of 12 nucleotides encoding PRRA. To date, the furin cleavage site (FCS) remains an experimentally uncharted territory both structurally and functionally. For instance, whether or not FCS is actually cleaved, before or after viral cell entry or exit, still remains to be experimentally investigated. With currently available structural data, this article presents a computational structural characterization of the FCS inserted into SARS-CoV-2 spike glycoprotein, and puts forward a set of structural hypothesis against the hypothesis of SARS-CoV-2 from purposeful manipulation: (1), the inserted FCS does not alter, neither stabilize nor de-stabilize, the overall structure of SARS-CoV-2 spike glycoprotein; (2), the net structural consequence of FCS is the insertion of a furin cleavage site into SARS-CoV-2 spike glycoprotein, whose S1 and S2 subunits will still be bonded together even if the FCS is actually cleaved by furin protease.

Keywords: COVID-19; SARS-CoV-2; Furin cleavage site (FCS); Salt bridging analysis

1. Introduction

It has been more than three months since the emergence of the COVID-19 coronavirus disease [1–4]. The past three months saw a considerable amount of studies of SARS-CoV-2, the causative virus of COVID-19 [5–9]. The membrane of SARS-CoV-2 harbours a trimeric ([10]) transmembrane spike (S) glycoprotein, which is essential for the entry of virus particles into the cell. The S protein contains two functional domains: a receptor binding domain (RBD) [11], and a second domain which contains sequences that mediate fusion of the viral and cell membranes [12–16].

Recently, it was reported that the spike glycoprotein of SARS-CoV-2 contains a potential cleavage site for furin protease, including four residues (Pro681, Arg682, Arg683 and Ala684) [17–21]. Specifically, Arg682, Arg683, Ala684 and Arg685 constitute the minimal polybasic furin cleavage site, i.e., RXYR, where X or Y should be Arg or Lys. In the midst of this COVID-19 public health emergency, this polybasic furin cleavage site (FCS) is reportedly linked to a natural selection, instead of purposeful manipulation, -based hypothesis of the origin of the global COVID-19 pandemic [17,22–26]. Irrespective of this hypothesis [17], it still remains elusive what the actual structural consequence is for this polybasic FCS to be inserted into the spike protein of SARS-CoV-2 at the boundary of its two subunits. Therefore, with FCS being the focus here, this article incorporates currently available structural data [27] and aims to uncover the structural impact of the FCS’s insertion into the spike glycoprotein of SARS-CoV-2 at the boundary of its two subunits.

2. Materials and Methods

With the GenBank identification number (QHD43416.1) of surface glycoprotein [Severe acute respiratory syndrome coronavirus 2], its amino acid sequence of was retrieved and listed below,

MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWF
HAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIRGWIFGTTLDSTQSLIVNNATNVVIKVFCE
QFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFNIDG
YFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTGDDSSSGWTAGAAAYYVG
YLQPRTFLLKYNENGTITDAVDCALDPLSEKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPF
GEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGD
EVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGNGYNYLYRLFRKSNLKPFERDISTEY
QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSEFLLHAPATVCGPKKSTNLVKNCV
NENFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQ
VAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSECDIPIGAGICASY
QTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYIC
GDSTECSNLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSK
PSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLTDEMQYTSALLAG
TITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKL
QDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAA
EIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSEFPQSAPHGVVFLHVTYVPAQEKNFTTAPAIC
HDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFK
EELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIW
LGFIAGLIAIVMVTIMLCCMTSCCCLKGCCSCGSCCKFDEDDSEPVLGKVKLHYT

First, the sequence above was plugged into the SwissModel [28] server to search for an experimental structure of the spike protein of SARS-CoV-2. The structure search led to an experimental structure (PDB ID: 6VSB), i.e., prefusion 2019-nCoV spike glycoprotein with a single receptor-binding domain up [10]. The sequence similarity between PDB ID 6VSB and GenBank ID QHD43416.1 is as high as 99.26%. Therefore, PDB ID 6VSB was used by the SwissModel [28] server subsequently to build a homology structural model of the spike protein of SARS-CoV-2. Afterwards, the homology structural model of the spike protein of SARS-CoV-2 was subject to a comprehensive set of electrostatic interaction analysis as described in [29] previously, and also a set of solvent accessible surface area

analysis by DSSP [30,31] to delve into the structural implication of the FCS inserted into the spike protein of SARS-CoV-2.

According to the solvent accessible surface area values, amino acid residue of a protein can be classified as buried or exposed according to a comparison with its standard solvent accessible surface area value contained in the "standard.data" file (supplementary file **standard.data**) available with the Naccess software distribution [32].

3. Results

3.1. Electrostatic analysis of the structural model of the spike protein of SARS-CoV-2

With the electrostatic analysis as described previously in [29], this article puts forward a comprehensive set of electrostatic interaction analysis for the structural model (supplementary file **model.pdb**) of the spike protein of SARS-CoV-2. The result of the electrostatic analysis is included in supplementary file **supplementary.pdf**.

Specifically,

1. no hydrogen bond was structurally identified from the whole structure of spike protein of SARS-CoV-2 (supplementary file **model.pdb**), neither main chain nor side chain hydrogen bond.
2. no salt bridge was structurally identified for the three basic residues at FCS (Arg682, Arg683 or Arg685), according to a close inspection of all tables in the supplementary file **supplementary.pdf**.

In brief, the inserted FCS is not involved in any electrostatic interaction within the spike protein of SARS-CoV-2.

3.2. Solvent accessible surface area analysis of the spike protein of SARS-CoV-2

To actually delve into the structural implication of the FCS inserted into the spike protein of SARS-CoV-2, it is necessary to first take a close look at the spatial location of FCS in the overall structure of the spike protein of SARS-CoV-2.

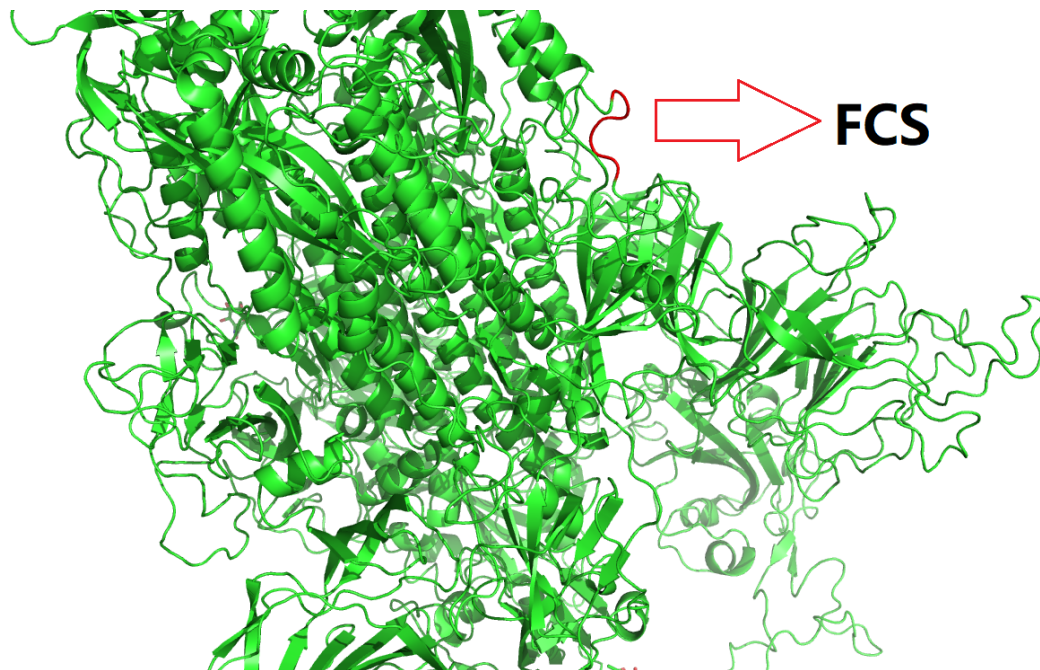


Figure 1. An overall structure of the spike protein of SARS-CoV-2 in green cartoon. In this figure, the inserted FCS fragment from Chain B is coloured red. This figure is prepared using PyMol [33] with supplementary file **model.pdb** as the input.

As shown in Figure 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 spike protein of SARS-CoV-2. In light of the electrostatic analysis that the inserted FCS is not involved in any electrostatic interaction within the spike protein of SARS-CoV-2, this observation suggests that the inserted FCS does not alter, neither stabilize nor de-stabilize, the overall structure of SARS-CoV-2 spike glycoprotein.

To further examine the structural implication of the FCS inserted into the spike protein of SARS-CoV-2, a set of quantitative solvent accessible surface area analysis by DSSP [30,31] was conducted for all three chains in the structure of the spike protein of SARS-CoV-2. The details of the solvent accessible surface area analysis is included in supplementary file **model.dssp**.

Specifically, the details of the solvent accessible surface area analysis is included in Table 1 for FCS as below,

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

Table 1. A quantitative analysis by DSSP [30] of the solvent accessible surface areas and relative accessibilities of the FCS. The relative accessibilities 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 package [32]. 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.

In addition to Table 1, 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 trimeric [10] structure of the spike protein of SARS-CoV-2, respectively. By and large, this result classifies the FCS as exposed, instead of buried, in all three chains of the homology structural model (supplementary file **model.pdb**) of the spike protein of SARS-CoV-2.

This quantitative relative accessibility analysis, along with the visual inspection of Figure 1, suggests that the net structural consequence of FCS is the insertion of a furin cleavage site into SARS-CoV-2 spike glycoprotein, leading to a further question: what if FCS is actually cleaved by furin protease?

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

To chart out what happens if the inserted polybasic FCS is actually cleaved by furin protease, a similar set of salt bridging analysis was conducted for the trimeric [10] structure of the spike protein of SARS-CoV-2, with the difference being that the whole 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 purpose of this specific set of salt bridging analysis is to identify all salt bridges that connect the two parts together even if the inserted polybasic FCS is actually cleaved by furin protease.

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

Table 2. Inter-part salt bridging analysis of the spike 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**.

From Table 2, it is quite clear that a series of inter-chain salt bridges still exist at the interface of the two parts generated from the hypothesized cleavage of the polybasic FCS by furin protease. For instance, a total of four inter-chain salt bridges still exist at the interface of chains B and C of the trimeric spike protein of SARS-CoV-2, as shown in Figures 2, 3 and 4. This set of four inter-chain salt bridges constitute a strong electrostatic force towards the structural stabilization of the interface of the trimeric [10] structure of the spike protein of SARS-CoV-2, even if the inserted polybasic FCS is actually cleaved by furin protease.

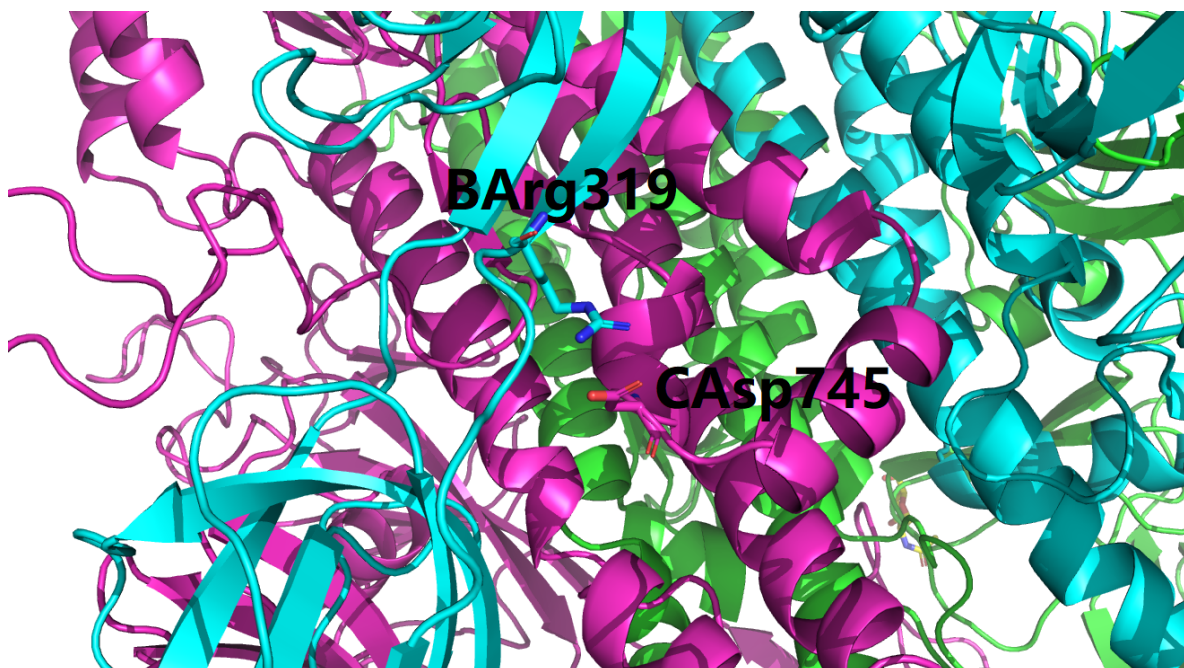


Figure 2. An overall view of the salt bridged residue pair B_ARG_319 and C_ASP_745 (Table 2). This figure is prepared using PyMol [33] with supplementary file **model.pdb** as the input. In this figure, the three (A, B and C) chains of the spike protein of SARS-CoV-2 is shown as green, cyan and purple cartoons, the residue naming scheme is **Chain ID, residue name, residue number**.

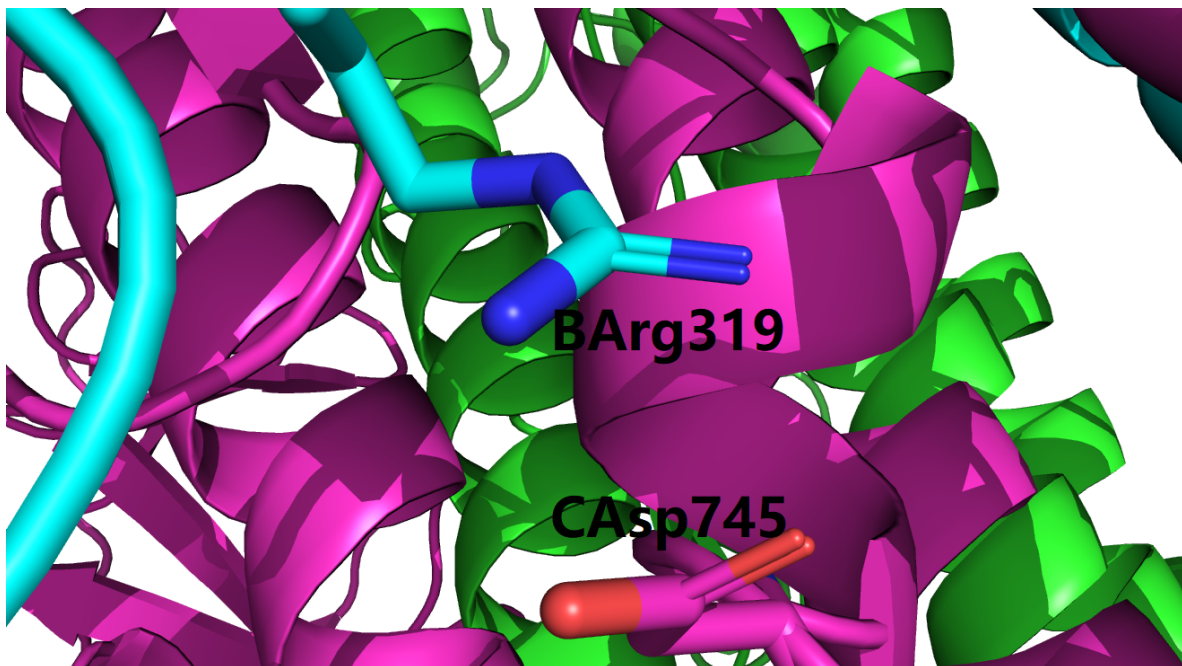


Figure 3. A detailed view of the salt bridged residue pair B_ARG_319 and C_ASP_745 (Table 2). This figure is prepared using PyMol [33] with supplementary file **model.pdb** as the input. In this figure, the three (A, B and C) chains of the spike protein of SARS-CoV-2 is shown as green, cyan and purple cartoons, the residue naming scheme is **Chain ID, residue name, residue number**.

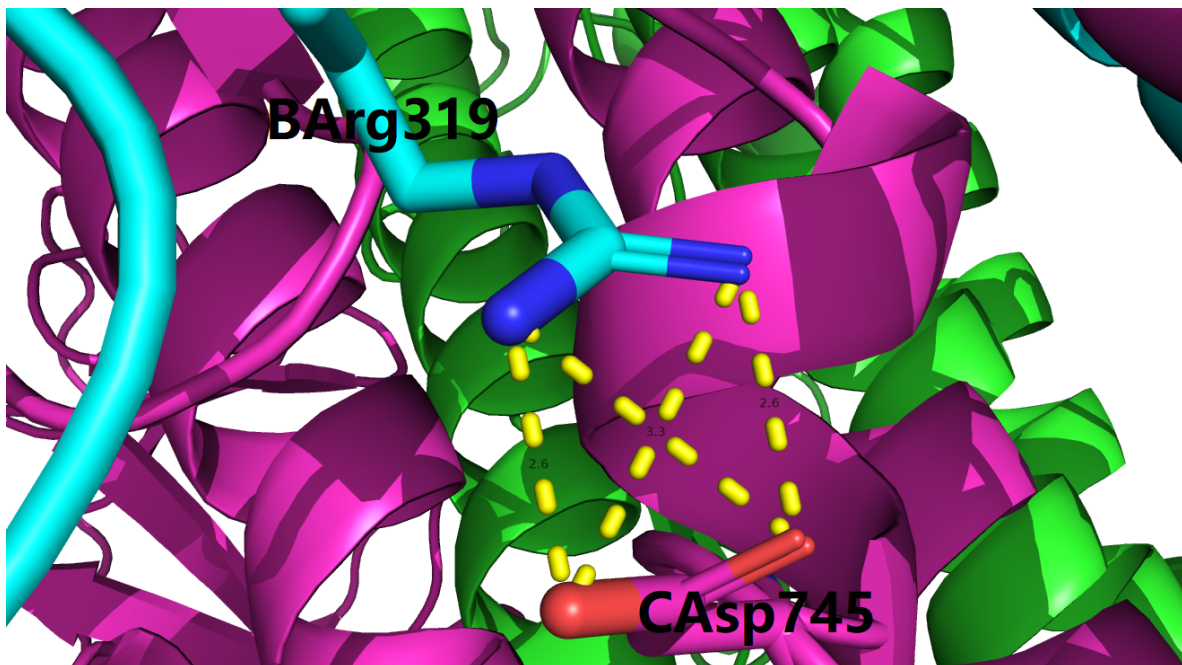


Figure 4. A detailed view of the salt bridged residue pair B_ARG_319 and C_ASP_745 (Table 2). This figure is prepared using PyMol [33] with supplementary file **model.pdb** as the input. In this figure, the three (A, B and C) chains of the spike protein of SARS-CoV-2 is shown as green, cyan and purple cartoons, the residue naming scheme is **Chain ID, residue name, residue number**, the four salt bridges are shown here with a series of dotted yellow lines, with their lengths being 2.6, 2.6, 3.3 and 3.5 Å, respectively.

Overall, this structural analysis supports the hypothesis that the S1 and S2 subunits of the spike protein of SARS-CoV-2 will still be structurally connected with each other, at least electrostatically [34], even if the FCS is actually cleaved by furin protease.

4. Conclusion and Discussion

With currently available structural data [27], this article presents a computational structural characterization of the FCS inserted into SARS-CoV-2 spike glycoprotein, and puts forward a set of structural analysis against the hypothesis of SARS-CoV-2 from purposeful manipulation:

1. the inserted FCS does not alter, neither stabilize nor de-stabilize, the overall structure of SARS-CoV-2 spike glycoprotein.
2. the net structural consequence of FCS is the insertion of a furin cleavage site into SARS-CoV-2 spike glycoprotein.
3. the S1 and S2 subunits of the spike glycoprotein of SARS-CoV-2 will still be bonded together, at least electrostatically [34], even if the FCS is actually cleaved by furin protease.

Finally, of the inserted FCS in the spike protein of SARS-CoV-2, the structural analysis here constitutes a preliminary starting point pointing towards a comprehensive structural and functional understanding of the spike protein of the causative virus of COVID-19 [5–9].

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, W.L.; validation, W.L.; formal analysis, W.L.; investigation, W.L.; resources, W.L.; data duration, W.L.; writing—original draft preparation, W.L.; writing—review and editing, W.L.; visualization, W.L.; supervision, W.L.; project administration, W.L.; funding acquisition, not applicable.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

1. Chan, J.F.W.; Kok, K.H.; Zhu, Z.; Chu, H.; To, K.K.W.; Yuan, S.; Yuen, K.Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections* **2020**, *9*, 221–236.
2. Chan, J.F.W.; Yuan, S.; Kok, K.H.; To, K.K.W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C.C.Y.; Poon, R.W.S.; Tsoi, H.W.; Lo, S.K.F.; Chan, K.H.; Poon, V.K.M.; Chan, W.M.; Ip, J.D.; Cai, J.P.; Cheng, V.C.C.; Chen, H.; Hui, C.K.M.; Yuen, K.Y. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* **2020**, *395*, 514–523.
3. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; Xia, J.; Yu, T.; Zhang, X.; Zhang, L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* **2020**, *395*, 507–513.
4. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; Yuan, M.L.; Zhang, Y.L.; Dai, F.H.; Liu, Y.; Wang, Q.M.; Zheng, J.J.; Xu, L.; Holmes, E.C.; Zhang, Y.Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**.
5. Angeletti, S.; Benvenuto, D.; Bianchi, M.; Giovanetti, M.; Pascarella, S.; Ciccozzi, M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *Journal of Medical Virology* **2020**.
6. Xu, X.W.; Wu, X.X.; Jiang, X.G.; Xu, K.J.; Ying, L.J.; Ma, C.L.; Li, S.B.; Wang, H.Y.; Zhang, S.; Gao, H.N.; Sheng, J.F.; Cai, H.L.; Qiu, Y.Q.; Li, L.J. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ* **2020**, p. m606.
7. Lippi, G.; Plebani, M. Laboratory abnormalities in patients with COVID-2019 infection. *Clinical Chemistry and Laboratory Medicine (CCLM)* **2020**, *0*.
8. Hoffmann, M.; Kleine-Weber, H.; Krüger, N.; Müller, M.; Drosten, C.; Pöhlmann, S. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells **2020**.
9. Li, W. Structurally Observed Electrostatic Features of the COVID-19 Coronavirus-Related Experimental Structures inside Protein Data Bank: A Brief Update **2020**.
10. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263.
11. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Aihara, H.; Li, F. Structure of 2019-nCoV chimeric receptor-binding domain complexed with its receptor human ACE2, 2020.
12. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; Chen, H.D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.D.; Liu, M.Q.; Chen, Y.; Shen, X.R.; Wang, X.; Zheng, X.S.; Zhao, K.; Chen, Q.J.; Deng, F.; Liu, L.L.; Yan, B.; Zhan, F.X.; Wang, Y.Y.; Xiao, G.F.; Shi, Z.L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273.
13. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; Yuan, M.L.; Zhang, Y.L.; Dai, F.H.; Liu, Y.; Wang, Q.M.; Zheng, J.J.; Xu, L.; Holmes, E.C.; Zhang, Y.Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269.
14. Gorbalenya, A.E.; Baker, S.C.; Baric, R.S.; de Groot, R.J.; Drosten, C.; Gulyaeva, A.A.; Haagmans, B.L.; Lauber, C.; Leontovich, A.M.; Neuman, B.W.; Penzar, D.; Perlman, S.; Poon, L.L.; Samborskiy, D.; Sidorov, I.A.; Sola, I.; Ziebuhr, J. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group **2020**.
15. Jiang, S.; Shi, Z.; Shu, Y.; Song, J.; Gao, G.F.; Tan, W.; Guo, D. A distinct name is needed for the new coronavirus. *The Lancet* **2020**, *395*, 949.
16. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* **2020**.

17. Andersen, K.G.; Rambaut, A.; Lipkin, W.I.; Holmes, E.C.; Garry, R.F. The proximal origin of SARS-CoV-2. *Nature Medicine* **2020**.
18. Han, X.; Bertzbach, L.D.; Veit, M. Mimicking the passage of avian influenza viruses through the gastrointestinal tract of chickens. *Veterinary Microbiology* **2019**, *239*, 108462.
19. Tse, L.V.; Hamilton, A.M.; Friling, T.; Whittaker, G.R. A Novel Activation Mechanism of Avian Influenza Virus H9N2 by Furin. *Journal of Virology* **2013**, *88*, 1673–1683.
20. Wan, Y.; Shang, J.; Graham, R.; Baric, R.S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology* **2020**, *94*.
21. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein **2020**.
22. Letko, M.; Marzi, A.; Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology* **2020**.
23. Menachery, V.D.; Dinnon, K.H.; Yount, B.L.; McAnarney, E.T.; Gralinski, L.E.; Hale, A.; Graham, R.L.; Scobey, T.; Anthony, S.J.; Wang, L.; Graham, B.; Randell, S.H.; Lipkin, W.I.; Baric, R.S. Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection. *Journal of Virology* **2019**, *94*.
24. Zhang, T.; Wu, Q.; Zhang, Z. Pangolin homology associated with 2019-nCoV **2020**.
25. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **2020**, *395*, 497–506.
26. Wong, M.C.; Cregeen, S.J.J.; Ajami, N.J.; Petrosino, J.F. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019 **2020**.
27. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980.
28. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **2018**, *46*, W296–W303.
29. Li, W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? *PLOS ONE* **2017**, *12*, e0178519.
30. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogenbonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
31. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **1976**, *32*, 922–923.
32. Hubbard, S.J.; Thornton, J.M. Naccess. *Computer Program, Department of Biochemistry and Molecular Biology, University College London* **1993**, *2*.
33. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **2002**, *40*, 82–92.
34. Anfinsen, C.B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.