

An ancient motif unique for human STING, RGS12 and SARS-CoV-2 spike proteins

Suleyman Aydin (✉ saydin@anadolu.edu.tr)

Anadolu University <https://orcid.org/0000-0002-4052-8470>

Ayca Cakmak Aydin

Yozgat Bozok University, Faculty of Medicine Dept. Medical Pharmacology <https://orcid.org/0000-0003-1767-782X>

Biological Sciences - Article

Keywords: T cell exhaustion, Cytokine Storm, Coagulation, Protein Motifs, Interferon Genes, Pathogenesis, Molecular Mimicry

Posted Date: January 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-132824/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An ancient motif unique for human STING, RGS12 and SARS-CoV-2 spike proteins.

2

The coronavirus named as SARS-CoV-2 is the cause of the COVID-19 pandemic and spreading rapidly¹. It is a pneumonia outbreak² with T cell exhaustion, cytokine storm and coagulation^{3,4}. Short motifs on proteins play important roles on protein-protein interactions^{5,6}.

6 We hypothesized role of molecular mimicry of small-xxx-small motifs for the spike protein of SARS-CoV-2. Here we show that a unique and evolutionary conserved motif is found only on the spike protein of SARS-CoV-2 and stimulator of interferon genes (STING) proteins.

8 Surprisingly we could not find this motif on any other protein of any living form. We found a similar, but not identical motif mimicry for the spike and regulator of G protein signaling 12 (RGS12), C1QT4 and also for proteins of Archaea and beta-lactamase enzymes of bacteria including *Mycobacterium tuberculosis*. STING proteins have roles on coagulation, T cell exhaustion, cytokine release¹⁻³ and RGS12 on inflammation⁷. In contrast to cGAS-STING pathway⁸, the motif mimicry indicated a direct interaction between spike and STING proteins suggesting the importance of STING, RGS12 and C1QT4 on the pathogenesis of COVID-19. To our surprise, the molecular mimicry showed that beta-lactamase inhibitors may be effective against SARS-CoV-2. The motif is unique, as found on Archaea and Cnidaria it is evolutionary old but a new target and mechanism for the COVID-19.

20

Main

COVID-19 pandemic is caused by the beta-coronavirus named as severe acute respiratory syndrome coronavirus (SARS-CoV-2)¹⁻³. The number of cases are over seventy millions and the deaths are more than one and a half million⁹. The COVID-19 is described as unknown pneumonia with gastrointestinal, cardiovascular, immunological and neurological complications. The most serious complication of COVID-19 is hypoxemia due to the respiratory failure and many patients die from acute respiratory distress syndrome (ARDS)¹⁻³. Venous and arterial thrombosis is very common in COVID-19 playing role on multisystem organ dysfunction. Thrombotic abnormalities and cardiovascular complications lead to ischemic stroke, myocardial infarction and venous thromboembolism playing role on multisystem organ dysfunction in COVID-19¹⁰. The pathophysiology is not fully defined but the spike protein of the virus plays role for entering the host cells^{1,3,11}. There is no proven treatment for the COVID-19¹². The angiotensin converting enzyme type 2 (ACE2) was found as the main receptor for the spike protein^{1,11}. ACE2 is absent on T cells¹³ and the interaction between ACE2 and the spike protein is not sufficient for explaining the mechanism(s) of coagulation and cytokine storms of the COVID-19.

GxxxG motifs are one of the small-xxx-small short linear motifs which have important role on the protein-protein interactions^{5,6} including virus proteins¹⁴. These motifs play role on the molecular mimicry and evolutionary arms race¹⁵. There are controversial results on the role of the GxxxG motif for the spike protein of the coronavirus^{16,17}. The aim of our study was to investigate the GxxxG (GG4) motifs on the spike protein of the SARS-CoV-2 using bioinformatical methods.

#1. Unique and evolutionary conserved motif

There were more than one GG4 motifs on the spike protein but one of them was rich in aromatic amino acids. The first amino acid of the motif was alanine but not glycine. We named this small-xxx-small motif as “semi-GG4 motif” which showed motif similarity with the stimulator of interferon genes (STING) proteins (Fig. 1). A molecular mimicry is present for the STING proteins and spike protein of SARS-CoV-2 which can be shown as a single formula: [AS]YY[FIV]GYL

(Fig. 1). The presence of this motif on the SPIKE proteins of many species including *Nematostella*
48 *vectensis* showed us the motif is evolutionary conserved on STING proteins since Cnidarians (Figs.
1,2 Extended Data Fig. 1). We were surprised to see that this motif mimicry is unique for the
50 SARS-CoV-2 and STING proteins (Fig. 1) because our search for this motif on the UniProt
surprisingly showed that it is not found on any other protein.

52 **#2. The motif and other coronaviruses**

A similar aromatic amino acid-rich semi-GG4 was found on the spike proteins of all coronaviruses
54 with a common motif: [ASL][YGKN][FYNSR][FYSVIL]G[FY][LC] (Extended Data Fig. 1).

#3. The spike group of the 21st century

56 Cluster analysis of the spike proteins of coronaviruses isolated from humans showed us that they
can be classified as 3 subgroups. We named them as first (1s), second (2s) and third spike group
58 (3s). There are only three members of the third spike group (3s): MERS, SARS (SARS-CoV) and
SARS2 (SARS-CoV-2) which are the causes of the serious infections of the 21st century (21st
60 century group) (Fig. 2). The Venn diagram showed us that these pathogens of the 21st century make
up a separate group of the human beta-coronavirus spike proteins (Fig. 3).

62 The clinical differences between MERS, SARS-CoV and SARS-CoV-2 are known¹⁸ and a similar
difference between the small-xxx-small motifs of these three spike proteins can be described in
64 terms of their aromatic amino acids. There is only one aromatic amino acid which is phenylalanine
(Phe) on the motif of MERS and three aromatic amino acids (2 Tyr + 1 Phe) on SARS-CoV but the
66 three aromatic amino acids on SARS-CoV-2 are all tyrosine (Tyr) (Fig. 2).

#4. Tyrosine

68 Tyrosine is important on the structure and function of proteins^{19,20}. The amount of aromatic amino
acids and Tyr content of the semi-GG4 motif of the spike proteins of the 3s subgroup is parallel
70 with their pathogenic potentials: Number of aromatic amino acids as Tyr are highest on motif of the
SARS-CoV-2 (Fig. 2) which is more contagious than others³.

72 **#5. Location of the motif**

The amino acid numbers of the unique semi-GG4 motif are 264-270 on the N-terminal domain
(NTD) of the spike protein (accession number: P0DTC2) (Figs. 4a,b). NTD was reported as another
binding region of SARS-CoV-2²¹⁻²³. The unique motif is also the binding site of the endogenous
ligand, cyclic-di-GMP of the STING protein²⁴.

#6. GYL triplet

The GYL triplet of the unique motif is found on the spike proteins of SARS-CoV and SARS-CoV-2
but not on MERS (Fig. 2) suggesting a possible role of the GYL triplet on the different pathogenic
properties of MERS, SARS-CoV and SARS-CoV-2. IGY motif was also reported on the secreted
toxic proteins of fungi²⁵ which is found inside the unique motif indicating the role of evolutionary
mechanisms (Fig. 3).

#7. Aromatic cage

Tyrosine as the 266th amino acid (Y²⁶⁶) is found only on the SARS-CoV-2 but not on SARS-CoV
and MERS (Figs. 2, 4b). There is hydrophobic contact between Y²⁶⁶ and W⁶⁴ and they make an
aromatic cage. Y²⁶⁶ is attached to the R²¹⁴ and A⁹³ of the neighbouring beta-sheets (Fig. 4c)
contributing to a more stabilized structure as reported for some other proteins^{26,27}. There is another
aromatic cage just nearby the Y²⁶⁶ (Fig. 4d) showing that the unique motif is found in an aromatic
cage-rich area. This structure is found only on the spike of SARS-CoV-2 but not on other members
of the 3s group because they (SARS-CoV and MERS) do not have Y²⁶⁶ and W⁶⁴ amino acids (Figs.
2, 5c). The unique motif is rich in aromatic amino acids and aromatic cages (Figs. 4c,d). Aromatic
cages usually capture positively charged molecules and amino acids like lysine²⁶ but there is no data
on this aromatic cage of the spike protein and we do not know the kind of role on the virus-host
relationships.

#8. STING protein

Free DNA in the cytoplasm is abnormal and it starts the STING signaling. Intracellular genomic
structures including viruses are sensed by the cyclic GMP synthase (cGAS) producing cyclic
dinucleotides like c-di-GMP which activate STING proteins²⁸. Activated STING is important in

autophagy²⁹, cytokine release¹, coagulation³⁰, obesity^{31,32} and old age³³. These are among the
100 symptoms of COVID-19, which are all effected by the STING proteins. The unique motif is the c-
di-GMP binding site on the STING protein²⁴ showing the presence of a molecular mimicry enabling
102 a direct interaction between the STING and the spike proteins of SARS-CoV-2 (Figs. 1,2). The c-
di-GMP binding site plays role on the direct interaction between STING and the spike proteins, as a
104 different mechanism from the cyclic guanosine monophosphate-adenosine monophosphate
synthase-stimulator of interferon genes (cGAS-STING) pathway. This direct interaction, in addition
106 to the cGAS-STING pathway, will result with hyperactivation of the STING proteins. STING
activation plays role on vascular and pulmonary pathologies³⁰ and it is a major player for the
108 induction of neutrophil extracellular traps³⁴ contributing to the immunothrombosis³⁵. Activated
STING proteins also have interferon-independent actions leading to T cell death³⁶, but STING null
110 cells and organisms are highly susceptible to infections of viruses, bacteria and intracellular
parasites like Plasmodium³⁷ showing importance of the balance of the functions of STING
112 proteins³⁸. The unique motif shows us one of the mechanisms of hyperstimulation of STING
proteins by the spike protein of SARS-CoV-2 leading to hyperinflammation, coagulation, T cell
114 exhaustion and high levels of neutrophil extracellular traps of COVID-19 and also responsible for
the enhanced actions of COVID-19 on obese³² and the old age¹.

116 #9. RGS12

We found another small-xxx-small (semi-GG4) motif for the spike protein of SARS-CoV-2 and
118 regulator of G protein signaling 12 (RGS12) proteins. It is similar, but not identical to the unique
motif for the STING proteins which can be written as: A[MY][VIY]VGYL (Figs. 5a,b). We
120 searched for this motif and found that it is also unique and found only on the RGS12 and spike
protein of SARS-CoV-2. RGS12 was recently reported to play a key role on inflammatory
122 reactions⁷ suggesting a significant contribution to the pathogenesis of COVID-19. Our results do
not indicate any role for the STING and RGS12 proteins on pain, anosmia, ageusia, sex differences
124 or the impact of air pollution on the COVID-19.

#10. TRPM ion channels

126 A surprising motif similarity between the spike protein of SARS-CoV-2 and a group of TRPM ion
channels (TRPM1-TRPM4) is another example of molecular mimicry which we did not investigate
128 further because it was very different than the unique motif (Extended Data Fig. 2).

#11. *Mycobacterium tuberculosis*

130 RGS12 and STING proteins are not specific for the respiratory system but the main pathogenic
actions of *Mycobacterium tuberculosis*, *Mycoplasma pneumonia* and COVID-19 are on the
132 respiratory system. Some of the proteins of the well known pathogenic bacteria of the pulmonary
infections including *M. pneumonia*, *M. tuberculosis*, *Klebsiella* and *Yersinia* species exhibit a motif
134 similarity to the unique motif. Their motifs are similar, but not identical to the unique motif and also
they are poor in aromatic amino acids (Figs. 1-5, Extended Data Figs. 1,3). Tuberculosis/COVID-19
136 co-infections are reported which may converge in a "perfect storm"³⁹. This motif similarity and the
molecular mimicry may help us understand the interaction between tuberculosis and COVID-19.

#12. C1QT4

It was also very surprising for us to find a motif very similar (but not identical) to the unique motif
140 for the beta-lactamase enzymes of *M. tuberculosis* and the STING proteins (Extended Data Fig. 4)
and also for C1q tumor necrosis factor-related protein 4 (FASTA name is C1QT4) (Extended Data
142 Fig. 4a). The motif similarity between the *M. tuberculosis* beta-lactamase and C1QT4 was high
compared to the STING proteins (Extended Data Fig. 4b). High levels of IL-6 is one of the severity
144 predictors in COVID-19⁴⁰. C1QT4 is one the major IL-6 elevating mechanisms and plays role on
viral infections⁴¹ indicating the role of C1QT4 on COVID-19 and supporting our results which was
146 not reported for COVID-19.

#13. Archaea and evolution

148 It was interesting to find the same unique motif similarity between the spike proteins, ribosomal
protein of *Methanosprillum hungatei* and membrane protein of *Methanococcus maripaludis*
150 (Extended Data Fig. 4c). These prokaryotes are members of anaerobic methanogen Archaea

(Extended Data Fig. 4c)⁴²⁻⁴³ showing evolutionary relationships.

152 **#14. Hub motif**

A small motif and a molecular mimicry enabling interactions of so many proteins, most if not all,
154 are involved in inflammatory reactions shows that the motif is short (only 7 amino acids) but not
functionally so simple. If the sequence of the motif is [AS]YY[FIV]GYL, it is unique for STINGS
156 and the spike protein of the SARS-CoV-2 but the motif [AS]xxxGYL is found on many proteins
including beta-lactamase, C1QT4, RGS12 and on the proteins of Archaea suggesting that the motif
158 is a member of "hub motifs"⁴⁴ with many other features awaiting to be discovered.

The motif is an evolutionary conserved "hub motif" and possibly the STING protein is a "hub
160 protein"⁴⁴.

#15. Beta-lactamase

162 There was a second motif for the beta-lactamase and the spike of SARS-CoV-2. This second motif
similarity between beta-lactamase and the spike was adjacent to the unique motif (Extended Data
164 Fig. 4d) suggesting an unusual interaction for beta-lactamase and the spike proteins. Based on this
surprising molecular mimicry, we suggest that the classical beta-lactamase inhibitors are expected
166 to inhibit some of the pathological effects of COVID-19. There is no proven drug for the COVID-
19⁴⁵ and based on our results (Extended Data Fig. 4), beta-lactamase inhibitors are expected to be
168 effective which may at least reduce IL-6 levels. Beta-lactamase inhibitors can be applied to the
patients without any delay.

170 **#17. Molecular mimicry, evolution, unique motif and beta-lactamase**

Importance of the GxxxG motif was reported on the SARS-CoV-2 proteins⁴⁶. Mimicry and
172 molecular mimicry are among the methods of the evolutionary arms race⁴⁷⁻⁴⁹ and mimicry was
proposed as a mechanism to explain multi-organ damage in COVID-19⁵⁰.

174 The aim of our study was not to investigate the interactions and roles of STING, RGS12, C1QT4
proteins or beta-lactamase enzymes on COVID-19, but the unique motif led us to these proteins and
176 to our surprise, to the beta-lactamase inhibitors. Our results show the importance and presence of

the evolutionary conserved motif mimicry.

178 There may be additional unique motifs on the proteins of SARS-CoV-2 which can help us to
explain the unknowns of the COVID-19 and find effective medicines. The evolutionary conserved
180 molecular mimicry of the unique motif shows us that beta-lactamase inhibitors may be used against
COVID-19. Role of the STING, RGS12, C1QT4 proteins and the unique motif on the COVID-19
182 are new for us but they are ancient which are found on proteins of Anthozoa and Archaea.

184 **References**

01. Hu, B., Guo, H., Zhou, P. & Shi, Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat.*
186 *Rev. Microbiol.* **6**, 1-14 (2020).
02. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin.
188 *Nature* **579**, 270-273 (2020).
03. Tang, D., Comish, P. & Kang, R. The hallmarks of COVID-19 disease. *PLoS Pathog* **16**,
190 e1008536 (2020).
04. Moon, C. Fighting COVID-19 exhausts T cells. *Nat Rev Immunol.* **20**, 277 (2020).
- 192 05. Lock, A. et al. One motif to bind them: A small-XXX-small motif affects transmembrane
domain 1 oligomerization, function, localization and cross-talk between two yeast GPCRs. *Biochim*
194 *Biophys Acta* **1838**, 3036-3051 (2014).
06. Teese, M.G. & Langosch, D. Role of GxxxG motifs in transmembrane domain interactions.
196 *Biochemistry* **54**, 5125-5135 (2015).
07. Yuan, G. et al. RGS12 is a novel critical NF- κ B activator in inflammatory arthritis. *iScience* **23**,
198 101072 (2020).
08. Wan, D., Jiang, W. & Hao, J. Research advances in how the cGAS-STING pathway controls the
200 cellular inflammatory response. *Front Immunol.* **11**, 615 (2020).
09. World Health Organization WHO coronavirus disease (COVID-19) Dashboard.
202 <https://covid19.who.int>.

10. Piazza, G & Morrow D.A. Diagnosis, management, and pathophysiology of arterial and venous
204 thrombosis in COVID-19. *JAMA* doi:10.1001/jama.2020.23422 (2020).
11. Liu, L. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike.
206 *Nature* **584**, 450-456 (2020).
12. Kim, P.S., Read, S.W. & Fauci, A.S. Therapy for early COVID-19: A critical need. *JAMA*
208 doi:10.1001/jama.2020.22813 (2020).
13. Hamming, I. et al. Tissue distribution of ACE2 protein, the functional receptor for SARS
210 coronavirus. A first step in understanding SARS pathogenesis. *J Pathol.* **203**, 631-637 (2004).
14. Bronnimann M.P., Chapman, J.A., Park, C.K. & Campos, S.K. A transmembrane domain and
212 GxxxG motifs within L2 are essential for Papillomavirus infection. *J Virol.* **87**, 464-473 (2013).
15. Guo, H. et al. Evolutionary arms race between virus and host drives genetic diversity in bat
214 severe acute respiratory syndrome-related coronavirus spike genes. *J Virol.* **94**, 00902-20 (2020).
16. Arbely, E., Granot, Z., Kass, I., Orly, J. & Arkin, I.T. A trimerizing GxxxG motif is uniquely
216 inserted in the severe acute respiratory syndrome (SARS) coronavirus spike protein transmembrane
domain. *Biochemistry* **45**, 11349 (2006).
17. Corver, J., Broer, R., van Kasteren, P. & Spaan, W. GxxxG motif of severe acute respiratory
218 syndrome coronavirus spike glycoprotein transmembrane domain is not involved in trimerization
and is not important for entry. *J. Virol.* **81**, 8352-8355 (2007).
18. Zhang, Y-Y., Li, B-R. & Ning, B-T. The comparative immunological characteristics of SARS-
222 CoV, MERS-CoV, and SARS-CoV-2 coronavirus infections. *Front Immunol.* **11**, 2033 (2020).
19. Stenberg, G., Abdalla, A-M. & Mannervik, B. Tyrosine 50 is at the subunit interface of dimeric
224 human glutathione transferase P1-1 is a structural key residue for modulating protein stability and
catalytic function. *Biochim Biophys Res Commun.* **271**, 59-63 (2000).
20. Yao, J. & Gillam, S. A single-amino-acid substitution of a tyrosine residue in the Rubella virus
226 E1 cytoplasmic domain blocks virus release. *J Virol.* **74**, 3029-3036 (2000).
21. Hulswit, R.J.G. et al. Human coronaviruses OC43 and HKU1 bind to 9- O-acetylated sialic
228

acids via a conserved receptor-binding site in spike protein domain A. *Proc Natl Acad Sci U S A*
230 **116**, 2681-2690 (2019).

22. Behloul, N., Baha, S., Shi, R. & Meng, J. Role of the GTNGTKR motif in the N-terminal
232 receptor-binding domain of the SARS-CoV-2 spike protein. *Virus Res.* **286**, 198058 (2020).

23. Seyran, M. et al. The structural basis of accelerated host cell entry by SARS-CoV-2. *FEBS J.*
234 doi: 10.1111/febs.15651 (2020).

24. Huang, Y-H., Liu, X-Y., Du, X-X., Jiang, Z-F. & Su, X-D. The structural basis for the sensing
236 and binding of cyclic di-GMP by STING. *Nat Struct Mol Biol.* **19**, 728-730 (2012).

25. Cheng, Q. et al. Discovery of a novel small secreted protein family with conserved N-terminal
238 IGY motif in Dikarya fungi. *BMC Genomics* **15**, 1151 (2014).

26. Cecchini, M. & Changeux, J.P. The nicotinic acetylcholine receptor and its prokaryotic
240 homologues: Structure, conformational transitions and allosteric modulation. *Neuropharmacology*
96, 137-149 (2015).

242 27. Nirthanan, S. Snake three-finger alpha-neurotoxins and nicotinic acetylcholine receptors:
molecules, mechanisms and medicine. *Biochem. Pharmacol.* **181**, 114168 (2020).

244 28. Li, Z. et al. When STING meets viruses: Sensing, trafficking and response. *Front Immunol.* **11**,
2064 (2020).

246 29. Gui, X. et al. Autophagy induction via STING trafficking is a primordial function of the cGAS
pathway. *Nature*, **567**, 262-266 (2019).

248 30. Liu, Y. et al. Activated STING in a vascular and pulmonary syndrome. *N. Eng. J. Med.* **371**,
507-518 (2014).

250 31. Mao, Y. et al. STING-IRF3 triggers endothelial inflammation in response to free fatty acid-
induced mitochondrial damage in diet-induced obesity. *Arterioscler Thromb Vasc Biol.* **37**, 920-929
252 (2017).

32. Sattar, N., McInnes, I.B & McMurray, J.J.V. Obesity is a risk factor for severe COVID-19
254 infection: Multiple potential mechanisms. *Circulation* **142**, 4-6 (2020).

33. Zhong, W. et al. Aging aggravated liver ischemia and reperfusion injury by promoting STING-mediated NLRP3 activation in macrophages. *Aging Cell* **19**, e13186 (2020).
34. Veras, F.P. et al. SARS-CoV-2-triggered neutrophil extracellular traps mediate COVID-19 pathology *J Exp Med.* **217**, e20201129 (2020).
35. Middleton, E.A. et al. Neutrophil extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory distress syndrome. *Blood* **136**, 1169-1179 (2020).
36. Wu, J., Dobbs, N., Yang, K. & Yan, N. Interferon-independent activities of mammalian STING mediate antiviral response and tumor immune evasion. *Immunity* **53**, 115-126 (2020).
37. Bryant, C.E. et al. International Union of Basic and Clinical Pharmacology. XCVI. Pattern Recognition Receptors in Health and Disease. *Pharmacol Rev.* **67**, 462-504 (2015).
38. Landman, S.L., Rensing, M.E. & van der Veen, A.G. Balancing STING in antimicrobial defense and autoinflammation. *Cytokine Growth Factor Rev.* **55**, 1-14 (2020).
39. Mousquer, G.T., Peres, A. & Fiegenbaum, M. Pathology of TB/COVID-19 co-infection: The phantom menace. *Tuberculosis* **126**, 102020 (2020).
40. Han, H. et al. Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. *Emerg. Microbes Infect.* **9**, 1123-1130 (2020).
41. Xu, W. et al. C1Q/TNF-related protein 4 expression correlates with herpes simplex encephalitis progression. *Ann Transl Med.* **7**, 235 (2019).
42. Hendrickson, E.L. et al. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J Bacteriol* **186**, 6956-6969 (2004).
43. Gunsalus, R.P. et al. Complete genome sequence of *Methanospirillum hungatei* type strain JF1. *Stand Genomic Sci.* **11**, 2 (2016).
44. Johnson, M.E. & Hummer, G. Evolutionary Pressure on the Topology of Protein Interface Interaction Networks. *J. Phys. Chem. B* **117**, 10.1021/jp402944e (2013).
45. Vijayvargiya, P. et al. Treatment considerations for COVID-19: A critical review of the evidence (or lack thereof). *Mayo Clin Proc.* **95**, 1454-1466 (2020).

46. Ionescu, J. An overview of the crystallized structures of the SARS-CoV-2. *Protein J.* doi:
10.1007/s10930-020-09933-w (2020)
47. Chemes, L.B., de Prat-Gay, G.& Sanches, I.E. Convergent evolution and mimicry of protein
linear motifs in host-pathogen interactions. *Curr Opin Struct Biol.* **32**, 91-101 (2015).
48. Elde, N.C. & Malik, H.S. The Evolutionary conundrum of pathogen mimicry. *Nat Rev
Microbiol.* **7**, 787-797 (2009).
49. Guo, H. et al. Evolutionary arms race between virus and host drives genetic diversity in bat
severe acute respiratory syndrome-related coronavirus spike genes. *J. Virol.* **94**, 00902-20 (2020).
50. Angileri, F. et al. Molecular mimicry may explain multi-organ damage in COVID-19.
Autoimmun Rev. **19**, 102591 (2020).
51. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
52. R Core Team (2019). R: A language and environment for statistical computing. R Foundation
for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
53. Rose, A.S. et al. NGL viewer: web-based molecular graphics for large complexes.
Bioinformatics **34**, 3755-3758 (2018).

Legends of the Figures

298 Fig. 1: Alignment of the spike protein of SARS-CoV-2 and the STING proteins. The motif
similarity ([AS]xxxGYL) is shaded. Black square represents the amino acids Ala and Ser [AS].

300

Fig. 2: Dendrogram of the phylogenetic relationships of the STING and the spike proteins of human
302 beta-coronaviruses based on alignment of amino acid sequences generated by Clustal Omega and
hierarchical cluster analysis using R. 'x' denotes any amino acid. The first small amino acid in the
304 small-xxx-small motifs are Ala(A), Leu(L) or Ser(S) and the other small amino acid is Gly(G)
making the semi-GG4 motif. 1s= First group, 2s= Second group, 3s= Third group. Members of the
306 3s are causes of the viral outbreaks of the 21st century and the ongoing pandemic of COVID-19.

308 Fig. 3: Venn diagram of the STING and the spike proteins of human beta-coronaviruses showing
the 3s as a distinct group as if it is evolving towards a new group of beta-coronavirus. The same
310 STING and the spike proteins shown in Fig. 2 were used for this Venn diagram.

312 Fig. 4: The unique motif is, (A) on the NTD of the spike protein marked with stars, (B) located on
one of the beta-sheets with a finger like loop extending outside, (C) there is a hydrogen bond
314 between the R²¹⁴ of the neighbour beta-sheet and Y²⁶⁶ which is found only on the SARS-CoV-2
member of the 3s group. Y²⁶⁶ makes an aromatic cage with the W⁶⁴ which is found only on SARS-
316 CoV-2, (D) another aromatic cage around A⁹³, the unique motif is surrounded with aromatic cages.

318 Fig. 5: Motif similarity (A) for the STING, RGS12 and the spike proteins of SARS-CoV and
SARS-CoV-2, (B) for the RGS12 and the spike protein of SARS-CoV-2 and (C) the presence of
320 W⁶⁴ only on the spike protein of SARS-CoV-2. The black square denotes the amino acids A and S
at the first small amino acid of the semi-GG4 motif.

322

Legends of the Extended Data Figures

324 Extended Data Fig. 1

Dendrogram showing the phylogenetic relationships of the STING and the spike proteins of all
326 coronaviruses based on alignment of amino acid sequences generated by Clustal Omega and
hierarchical cluster analysis using R.

328

Extended Data Fig. 2

330 A motif similarity between the spike protein of SARS-Cov-2 and the TRPM ion channels
(TRPM1, TRPM2, TRPM3 and TRPM4) which is very different from the unique motif.

332

Extended Data Fig. 3

334 Motif similarity (A) for the protein of *Mycoplasma pneumonia* and the STING proteins, (B) for the
proteins of *Mycobacterium tuberculosis* and the STING proteins and (C) for the proteins of STING
336 proteins and the *Klebsiella* and *Yersinia* species.

338 Extended Data Fig. 4

Motif similarity (A) for the C1QT4, beta-lactamase enzymes, STING and the spike protein of
340 SARS-CoV-2, (B) for the beta-lactamase enzymes and C1QT4, spike proteins of SARS-CoV,
SARS-CoV-2, ribosomal protein of *Methanosprillum hungatei* and membrane protein of
342 *Methanococcus maripaludis* of Archaea and (D) for the spike proteins of SARS-CoV, SARS-CoV-
2 and beta-lactamase enzymes. There is a new motif similarity different from the unique motif
344 located adjacent to the unique motif. The new motif was shaded and the unique motif adjacent to it
was shown by a red square.

346

348 **Methods**

350 The protein data as a compressed single file, named as uniprot_sprot.fasta.gz, was downloaded from the ftp servers of European Bioinformatics Institute (ftp://ftp.ebi.ac.uk). Wget (ver. 1.20.3), gunzip (vers.1.8) and grep (vers.2.25) were used for downloading and extracting protein sequences, Clustal Omega (ver. 1.2.4) for the alignment of proteins⁵¹, the R programming language and packages were used for hierarchical cluster analysis, dendrogram and Venn diagrams⁵². Sed (ver. 4.2.2), less (ver. 481) and vi improved (vim) (ver.7.4) were used for editing, imagemagic (ver. 6.9.4_9) and enscrip (ver.1.6.6) for image editing and producing vector images (pdf). The software listed above were operated under the Slackware GNU/Linux (kernel 4.4.157). The unshaded and unedited results of Clustal Omega were merged into a single pdf file using pdftk (ver. 2.02) and given as supplement. 358 The spike protein of SARS-CoV-2 (PDB ID 6XEY)¹¹ for Fig. 4 was created with NGL⁵³. The motif similarities in the Figures were shaded. Names of the proteins in the figures are in FASTA style and 360 amino acids are shown as single letter code.

362

Data Reporting

364 Cluster analysis were performed on the proteins for dendrogram and Venn diagrams. No other
statistical methods were used in the study. All the protein sequences are deposited on the UniProt
366 servers and the PDB ID of the spike protein of the SARS-CoV-2 used in the study is 6XEY found
on the PDB servers.

368

Author Contributions

370 Both authors equally contributed to the study.

Declaration of interests

We declare no competing interests.

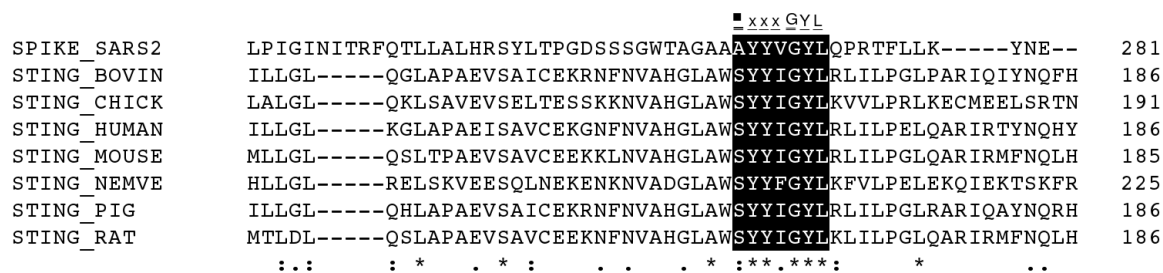
374

Additional information

376 Supplemantary information was given for the unshaded multiple comparison for all figures, as a
single pdf file.

378 Correspondence and requests for materials should be addressed to S.A.

384



388 Fig. 1:

Alignment of the spike protein of SARS-CoV-2 and the STING proteins. The motif similarity
 390 ([AS]xxxGYL) is shaded. Black square represents the amino acids Ala and Ser [AS].

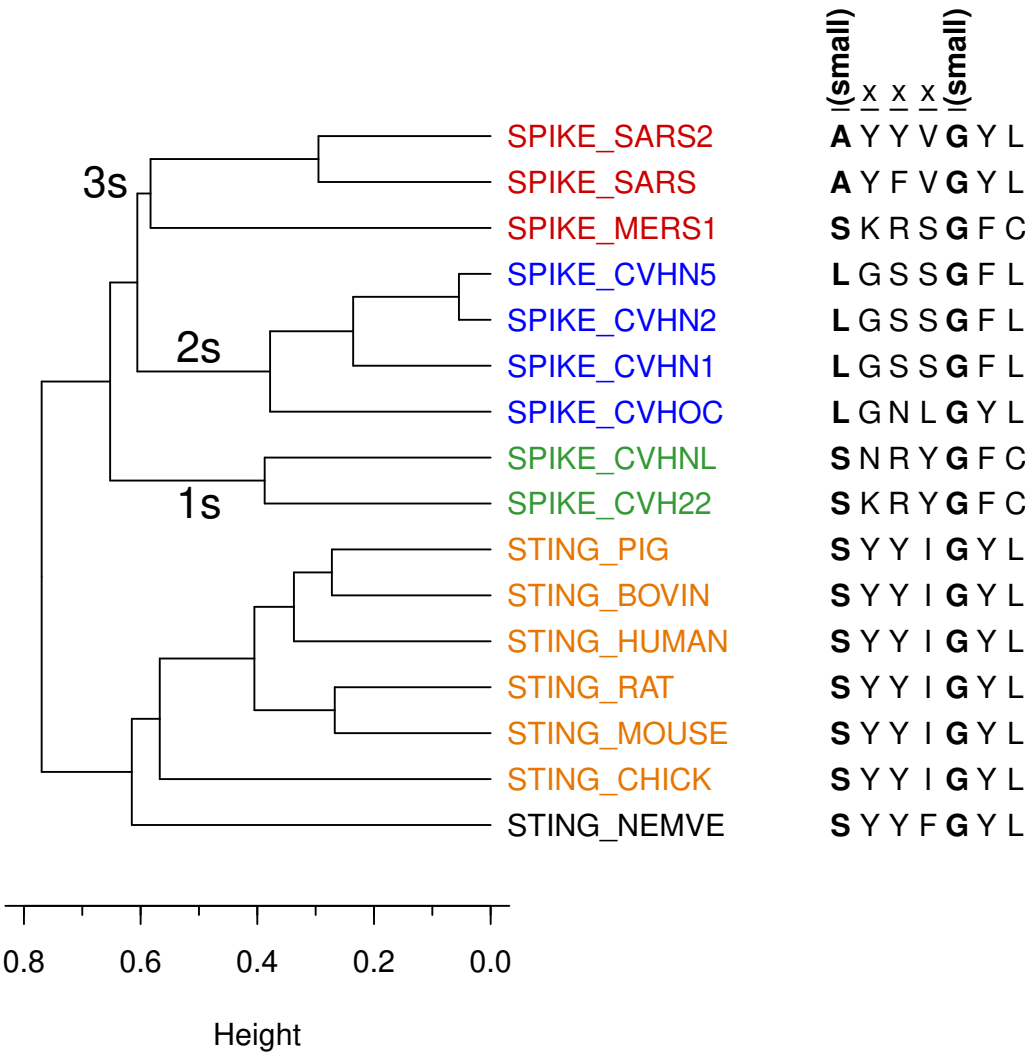


Fig. 2: Dendrogram of the phylogenetic relationships of the STING and the spike proteins of human
394 beta-coronaviruses based on alignment of amino acid sequences generated by Clustal Omega and
hierarchical cluster analysis using R. 'x' denotes any amino acid. The first small amino acid in the
396 small-xxx-small motifs are A, L or S and the other small amino acid is G making the semi-GG4
motif. 1s= First group, 2s= Second group, 3s= Third group. Members of the 3s are the cause of viral
398 outbreaks of the 21st century and the ongoing pandemic of COVID-19.

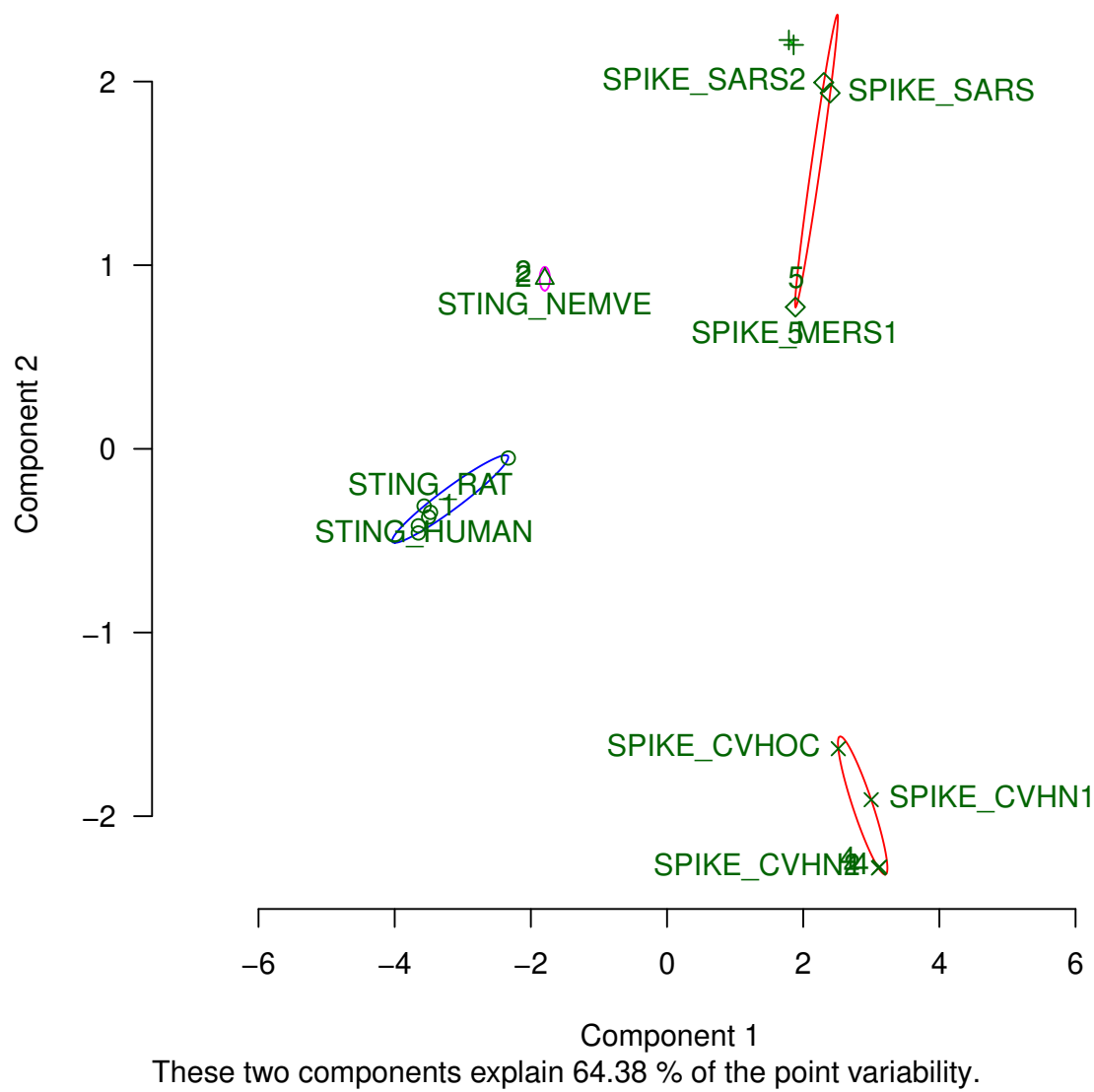


Fig. 3: Venn diagram of the STING and the spike proteins of human beta-coronaviruses showing the 3s as a distinct group as if it is evolving towards a new group of beta-coronavirus. The same STING and the spike proteins shown in Fig. 2 were used for this Venn diagram.

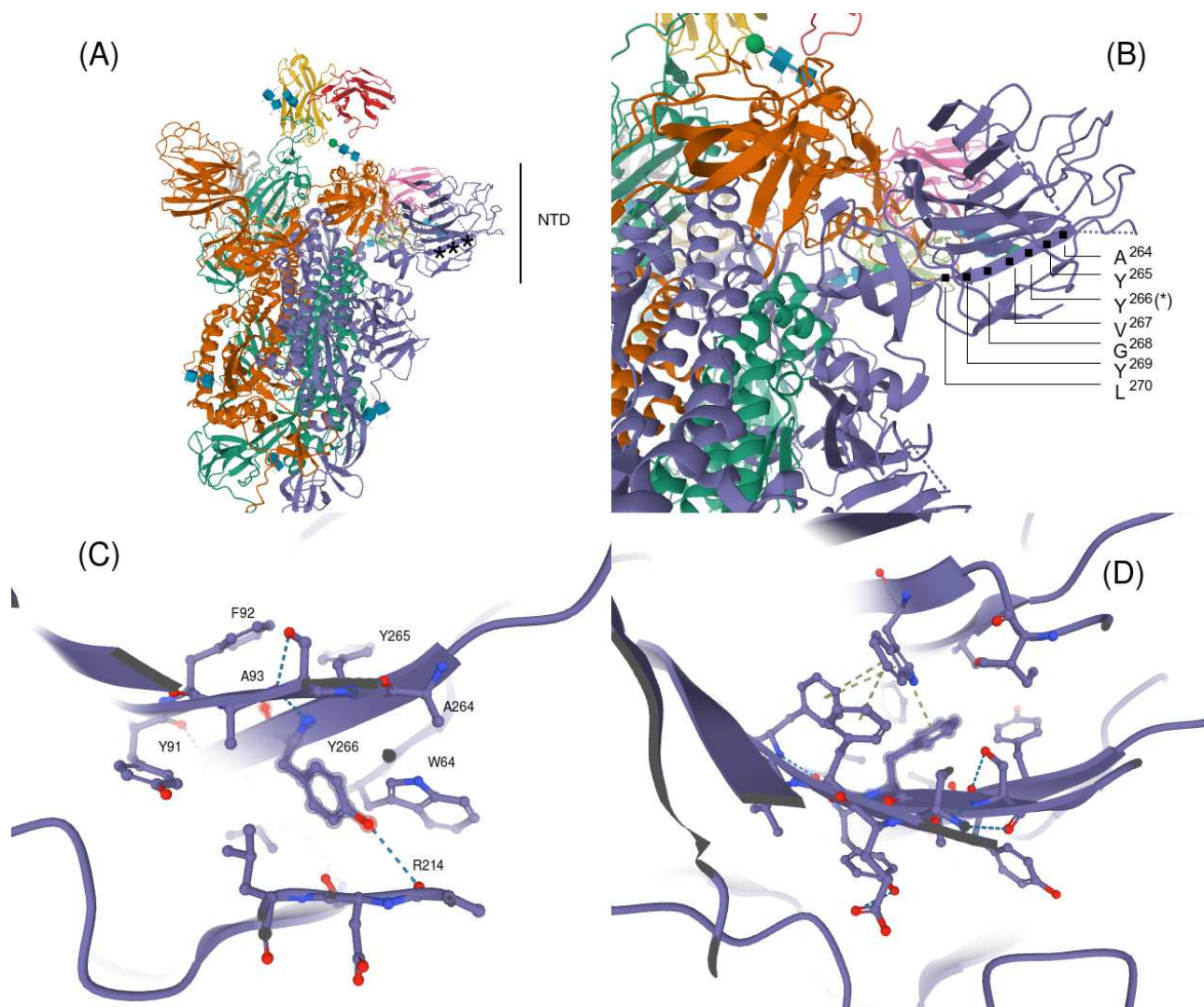


Fig. 4: The unique motif is, (A) on the NTD of the spike protein marked with stars, (B) located on one of the beta-sheets with a finger like loop extending outside, (C) there is a hydrogen bond between the R²¹⁴ of the neighbour beta-sheet and Y²⁶⁶ which is found only on the SARS-CoV-2 member of the 3s group. Y²⁶⁶ makes an aromatic cage with the W⁶⁴ which is found only on SARS-CoV-2, (D) another aromatic cage around A⁹³, the unique motif is surrounded with aromatic cages.

(C)

SPIKE_MERS1	QGR T YSNITITYQGLF-PYQGDHGD M YVYSAGHATGTT P QKLFVANYSQDVKQFANGFVV	118
SPIKE_SARS	DEIFRSDTLYLTQDLFLPFYSNVT---GFHTIN-----HT---FGNPVIPFKDGIYF	89
SPIKE_SARS2	DKVFRSSVLHSTQDLFLPFFSNVT--- W FHAIHVS G TNGTKR---FDNPVLPFNDGVYF	92
	: * : * * * : : : : : : : : : : : * * * :	

Fig. 5: Motif similarity (A) for the STING, RGS12 and the spike proteins of SARS-CoV and SARS-CoV-2, (B) for the RGS12 and the spike protein of SARS-CoV-2 and (C) the presence of W⁶⁴ only on the spike protein of SARS-CoV-2. The black square denotes the amino acids A and S at the first small amino acid of the semi-GG4 motif.

Figures

		■ xxx GYL	
SPIKE_SARS2	LPIGINITRFQTLALHRSYLTTPGDSSSGWTAGAA	AYYVGYLQPRTFLLK-----YNE--	281
STING_BOVIN	ILLGL-----QGLAPAEVSAICEKRNFNVAHGLAW	SYIIGYLRILPGLPARIQIYNQFH	186
STING_CHICK	LALGL-----QKLSAVEVSELTESSKKNVAHGLAW	SYIIGYLRKVLPRLKECMEELSRTN	191
STING_HUMAN	ILLGL-----KGLAPAEISAVCEKGNFNVAHGLAW	SYIIGYLRILPELQARIRTYNQHY	186
STING_MOUSE	MLLGL-----QSLTPAEVSACEEKKLNVAHGLAW	SYIIGYLRILPGLQARIRMFNQIH	185
STING_NEMVE	HLLGL-----RELSKVEESQLNEKENKNVADGLAW	SYIFGYLKFVLPELEKQIEKTSKFR	225
STING_PIG	ILLGL-----QHLAPAEVSAICEKRNFNVAHGLAW	SYIIGYLRILPGLRARIQAYNQRH	186
STING_RAT	MTLDL-----QSLAPAEVSACEEKNFNVAHGLAW	SYIIGYLRILPGLQARIRMFNQIH	186
	:.: : * . * : . . . * : ** . *** :	* ..	

Figure 1

Alignment of the spike protein of SARS-CoV-2 and the STING proteins. The motif similarity ([AS]xxxGYL) is shaded. Black square represents the amino acids Ala and Ser [AS].

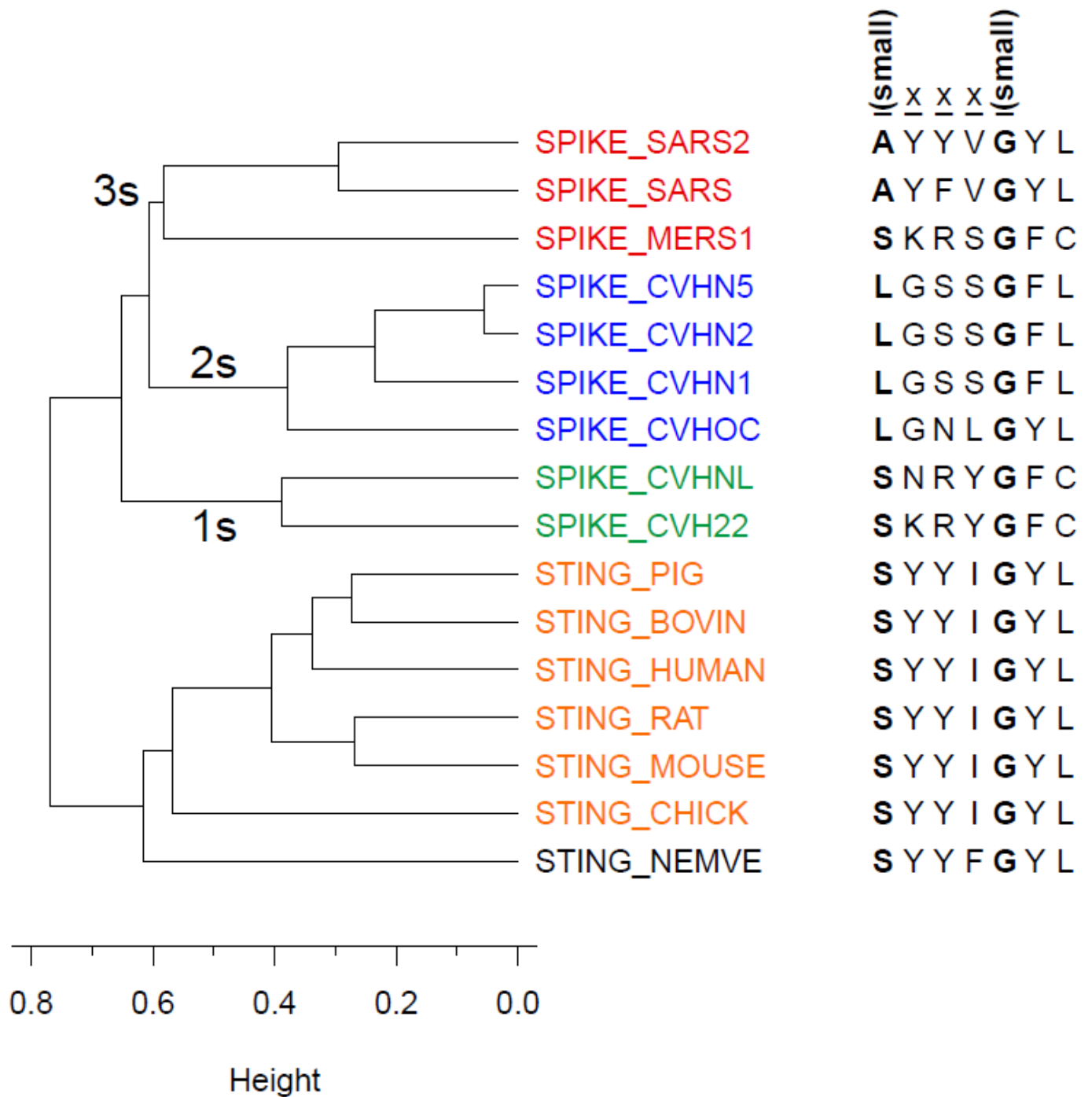


Figure 2

Dendrogram of the phylogenetic relationships of the STING and the spike proteins of human beta-coronaviruses based on alignment of amino acid sequences generated by Clustal Omega and hierarchical cluster analysis using R. 'x' denotes any amino acid. The first small amino acid in the small-xxx-small motifs are A, L or S and the other small amino acid is G making the semi-GG4 motif. 1s= First group, 2s= Second group, 3s= Third group. Members of the 3s are the cause of viral outbreaks of the 21st century and the ongoing pandemic of COVID-19.

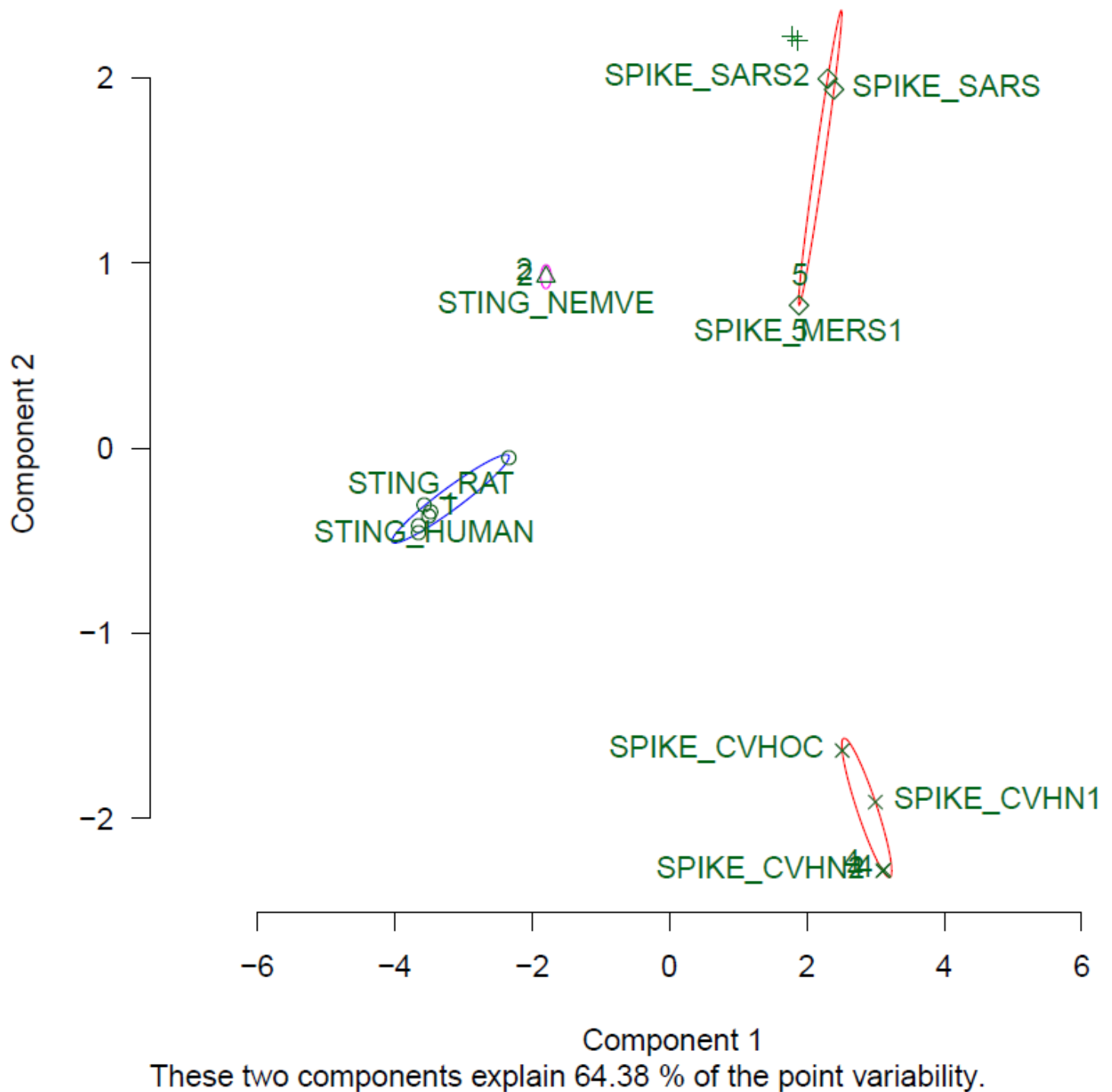


Figure 3

Venn diagram of the STING and the spike proteins of human beta-coronaviruses showing the 3s as a distinct group as if it is evolving towards a new group of beta-coronavirus. The same STING and the spike proteins shown in Fig. 2 were used for this Venn diagram.

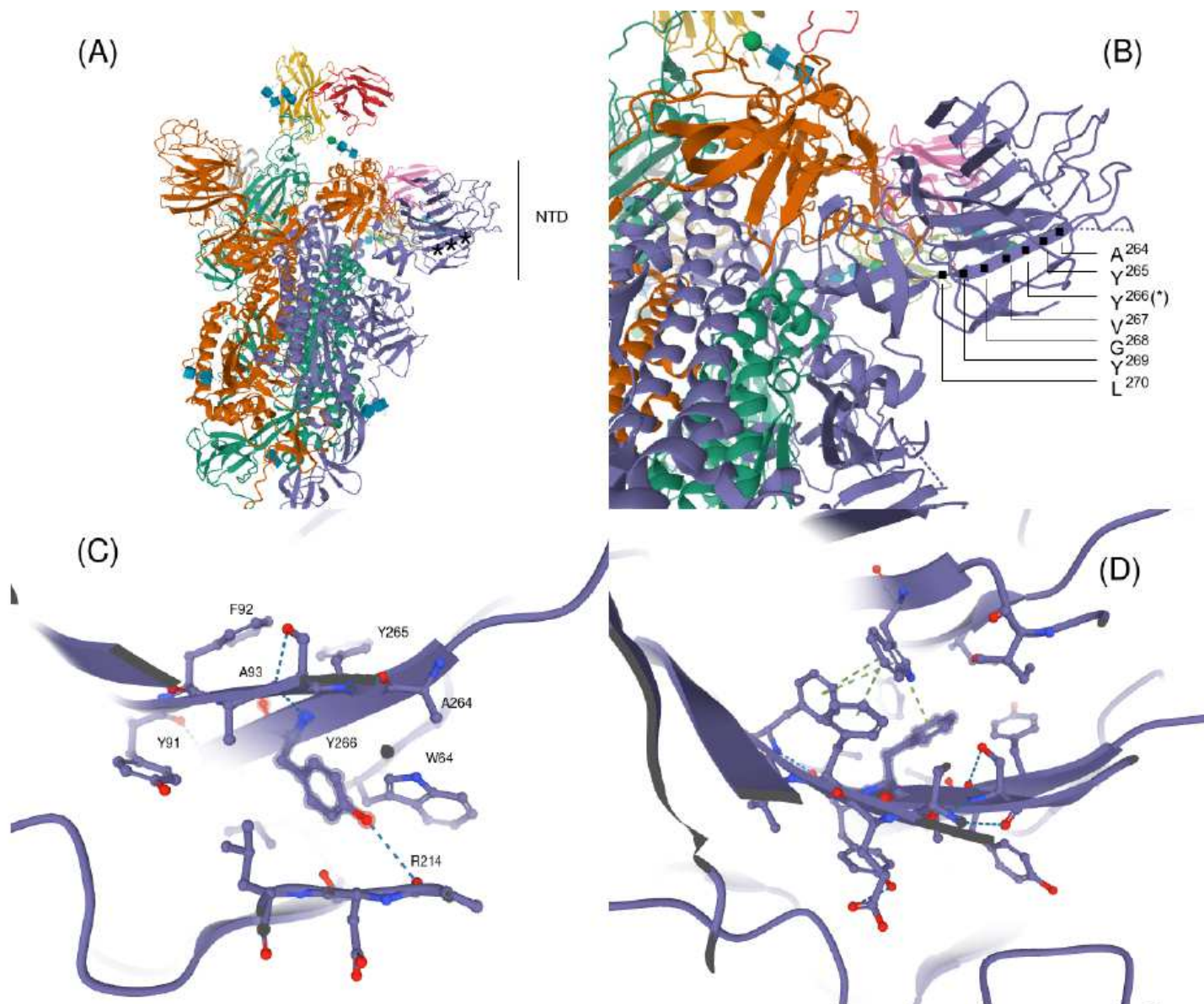


Figure 4

The unique motif is, (A) on the NTD of the spike protein marked with stars, (B) located on one of the beta-sheets with a finger like loop extending outside, (C) there is a hydrogen bond between the R214 of the neighbour beta-sheet and Y266 which is found only on the SARS-CoV-2 member of the 3s group. Y266 makes an aromatic cage with the W64 which is found only on SARSCoV-2, (D) another aromatic cage around A93, the unique motif is surrounded with aromatic cages.



Figure 5

Motif similarity (A) for the STING, RGS12 and the spike proteins of SARS-CoV and SARS-CoV-2, (B) for the RGS12 and the spike protein of SARS-CoV-2 and (C) the presence of W64 only on the spike protein of SARS-CoV-2. The black square denotes the amino acids A and S at the first small amino acid of the semi-GG4 motif.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [extendeddatafig01.pdf](#)
- [extendeddatafig02.pdf](#)
- [extendeddatafig03.pdf](#)
- [extendeddatafig04.pdf](#)
- [supplementforfigures.pdf](#)