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

The coronavirus named as SARS-CoV-2 is the cause of the COVID-19 pandemic and spreading rapidly[1]. It is a pneumonia outbreak[2] with T cell exhaustion, cytokine storm and coagulation[3,4]. Short motifs on proteins play important roles on protein-protein interactions[5,6]. We hypothetized 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. 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 release[1-3] and RGS12 on inflammation[7]. In contrast to cGAS-STING pathway[8], 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.

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 million[9]. 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-19[10]. 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-19[12]. The angiotensin converting enzyme type 2 (ACE2) was found as the main receptor for the spike protein[1,11]. ACE2 is absent on T cells[13] 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 proteins[14]. These motifs play role on the molecular mimicry and evolutionary arms race[15]. 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 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 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.

#2. The motif and other coronaviruses

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

#3. The spike group of the 21st century

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 (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 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).

The clinical differences between MERS, SARS-CoV and SARS-CoV-2 are known[18] and a similar difference between the small-xxx-small motifs of these three spike proteins can be described in 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 three aromatic amino acids on SARS-CoV-2 are all tyrosine (Tyr) (Fig. 2).

#4. Tyrosine

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 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 others[3].

#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-2[21-23]. The unique motif is also the binding site of the endogenous ligand, cyclic-di-GMP of the STING protein[24].

#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 fungi[25] which is found inside the unique motif indicating the role of evolutionary mechanisms (Fig. 3).

#7. Aromatic cage

Tyrosine as the 266th amino acid (Y^{266}) is found only on the SARS-CoV-2 but not on SARS-CoV and MERS (Figs. 2, 4b). There is hydrophobic contact between Y^{266} and W^{64} and they make an aromatic cage. Y^{266} is attached to the R^{214} and A^{93} 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 Y^{266} (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 Y^{266} and W^{64} 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 lysine[26] 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 proteins[28]. Activated STING is important in

autophagy[29], cytokine release[1], coagulation[30], obesity[31,32] and old age[33]. These are among the 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 protein[24] showing the presence of a molecular mimicry enabling 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 different mechanism from the cyclic guanosine monophosphate-adenosine monophosphate synthase-stimulator of interferon genes (cGAS-STING) pathway. This direct interaction, in addition to the cGAS-STING pathway, will result with hyperactivation of the STING proteins. STING activation plays role on vascular and pulmonary pathologies[30] and it is a major player for the induction of neutrophil extracellular traps[34] contributing to the immunothrombosis[35]. Activated STING proteins also have interferon-independent actions leading to T cell death[36], but STING null cells and organisms are highly susceptible to infections of viruses, bacteria and intracellular parasites like Plasmodium[37] showing importance of the balance of the functions of STING proteins[38]. 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 exhaustion and high levels of neutrophil extracellular traps of COVID-19 and also responsible for the enhanced actions of COVID-19 on obese[32] and the old age[1].

#9. RGS12

We found another small-xxx-small (semi-GG4) motif for the spike protein of SARS-CoV-2 and 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 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 reactions[7] 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 or the impact of air pollution on the COVID-19.

#10. TRPM ion channels

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 further because it was very different than the unique motif (Extended Data Fig. 2).

#11. *Mycobacterium tuberculosis*

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 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 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 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 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 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 predictors in COVID-19[40]. C1QT4 is one the major IL-6 elevating mechanisms and plays role on viral infections[41] indicating the role of C1QT4 on COVID-19 and supporting our results which was not reported for COVID-19.

#13. Archaea and evolution

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* (Extended Data Fig. 4c). These prokaryotes are members of anaerobic methanogen Archaea

(Extended Data Fig. 4c)[42-43] showing evolutionary relationships.

#14. Hub motif

A small motif and a molecular mimicry enabling interactions of so many proteins, most if not all, 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 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 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 protein"[44].

#15. Beta-lactamase

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 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 to inhibit some of the pathological effects of COVID-19. There is no proven drug for the COVID-19[45] and based on our results (Extended Data Fig. 4), beta-lactamase inhibitors are expected to be effective which may at least reduce IL-6 levels. Beta-lactamase inhibitors can be applied to the patients without any delay.

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

Importance of the GxxxG motif was reported on the SARS-CoV-2 proteins[46]. Mimicry and molecular mimicry are among the methods of the evolutionary arms race[47-49] and mimicry was proposed as a mechanism to explain multi-organ damage in COVID-19[50].

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 to our surprise, to the beta-lactamase inhibitors. Our results show the importance and presence of

the evolutionary conserved motif mimicry.

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 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 are new for us but they are ancient which are found on proteins of Anthozoa and Archaea.

References

01. Hu, B., Guo, H., Zhou, P. & Shi, Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol*. **6**, 1-14 (2020).

02. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).

03. Tang, D., Comish, P. & Kang, R. The hallmarks of COVID-19 disease. *PLoS Pathog* **16**, e1008536 (2020).

04. Moon, C. Fighting COVID-19 exhausts T cells. *Nat Rev Immunol.* **20**, 277 (2020).

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 Biophys Acta* **1838**, 3036-3051 (2014).

06. Teese, M.G. & Langosch, D. Role of GxxxG motifs in transmembrane domain interactions. *Biochemistry* **54**, 5125-5135 (2015).

07. Yuan, G. et al. RGS12 is a novel critical NF-kB activator in inflammatory arthritis. *iScience* **23**, 101072 (2020).

08. Wan, D., Jiang, W. & Hao, J. Research advances in how the cGAS-STING pathway controls the cellular inflammatory response. *Front Immunol*. **11**, 615 (2020).

09. World Health Organization WHO coronavirus disease (COVID-19) Dashboard. 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).

218 17. Corver, J., Broer, R., van Kasteren, P. & Spaan, W. GxxxG motif of severe acute respiratory syndrome coronavirus spike glycoprotein transmembrane domain is not involved in trimerization
220 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 mudulating protein stability and catalytic function. *Biochim Biophys Res Commun*. **271**, 59-63 (2000).

226 20. Yao, J. & Gillam, S. A single-amino-acid substitution of a tyrosine residue in the Rubella virus E1 cytoplasmic domain blocks virus release. *J Virol*. **74**, 3029-3036 (2000).

228 21. Hulswit, R.J.G. et al. Human coronaviruses OC43 and HKU1 bind to 9- O-acetylated sialic

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

22. Behloul, N., Baha, S., Shi, R. & Meng, J. Role of the GTNGTKR motif in the N-terminal 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*. 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 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 IGY motif in Dikarya fungi. *BMC Genomics* **15**, 1151 (2014).

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

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

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

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

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

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 (2017).

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

33. Zhong, W. et. al. Aging aggravated liver ischemia and reperfusion injury by promoting STING-
256 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
258 pathology *J Exp Med*. **217**, e20201129 (2020).

35. Middleton, E.A. et al. Neutrophil extracellular traps contribute to immunothrombosis in
260 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
262 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
264 Recognition Receptors in Health and Disease. *Pharmacol Rev*. **67**, 462-504 (2015).

38. Landman, S.L., Ressing, M.E. & van der Veen, A.G. Balancing STING in antimicrobial defense
266 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
268 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
270 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
272 progression. *Ann Transl Med*. **7**, 235 (2019).

42. Hendrickson, E.L. et al. Complete genome sequence of the genetically tractable
274 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.
276 *Stand Genomic Sci*. **11**, 2 (2016).

44. Johnson, M.E. & Hummer, G. Evolutionary Pressure on the Topology of Protein Interface
278 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
280 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

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].

Fig. 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 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 3s are causes of the viral 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 R^{214} of the neighbour beta-sheet and Y^{266} which is found only on the SARS-CoV-2 member of the 3s group. Y^{266} makes an aromatic cage with the W^{64} which is found only on SARS-CoV-2, (D) another aromatic cage around A^{93}, the unique motif is surrounded with aromatic cages.

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 W^{64} 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.

Legends of the Extended Data Figures

Extended Data Fig. 1

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

Extended Data Fig. 2

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.

Extended Data Fig. 3

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 proteins and the *Klebsialla* and *Yersinia* species.

Extended Data Fig. 4

Motif similarity (A) for the C1QT4, beta-lactamase enzymes, STING and the spike protein of 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 *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 located adjacent to the unique motif. The new motif was shaded and the unique motif adjacent to it was shown by a red square.

Methods

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 proteins[51], the R programming language and packages were used for hierarchical cluster analysis, dendrogram and Venn diagrams[52]. 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 enscript (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. The spike protein of SARS-CoV-2 (PDB ID 6XEY)[11] for Fig. 4 was created with NGL[53]. The motif similarities in the Figures were shaded. Names of the proteins in the figures are in FASTA style and amino acids are shown as single letter code.

Data Reporting

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 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.

Author Contributions

Both authors equally contributed to the study.

Declaration of interests

We declare no competing interests.

Additional information

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

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

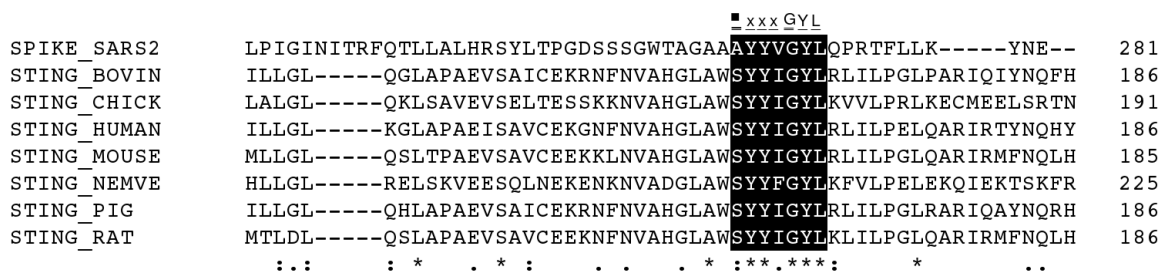

Fig. 1:

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].

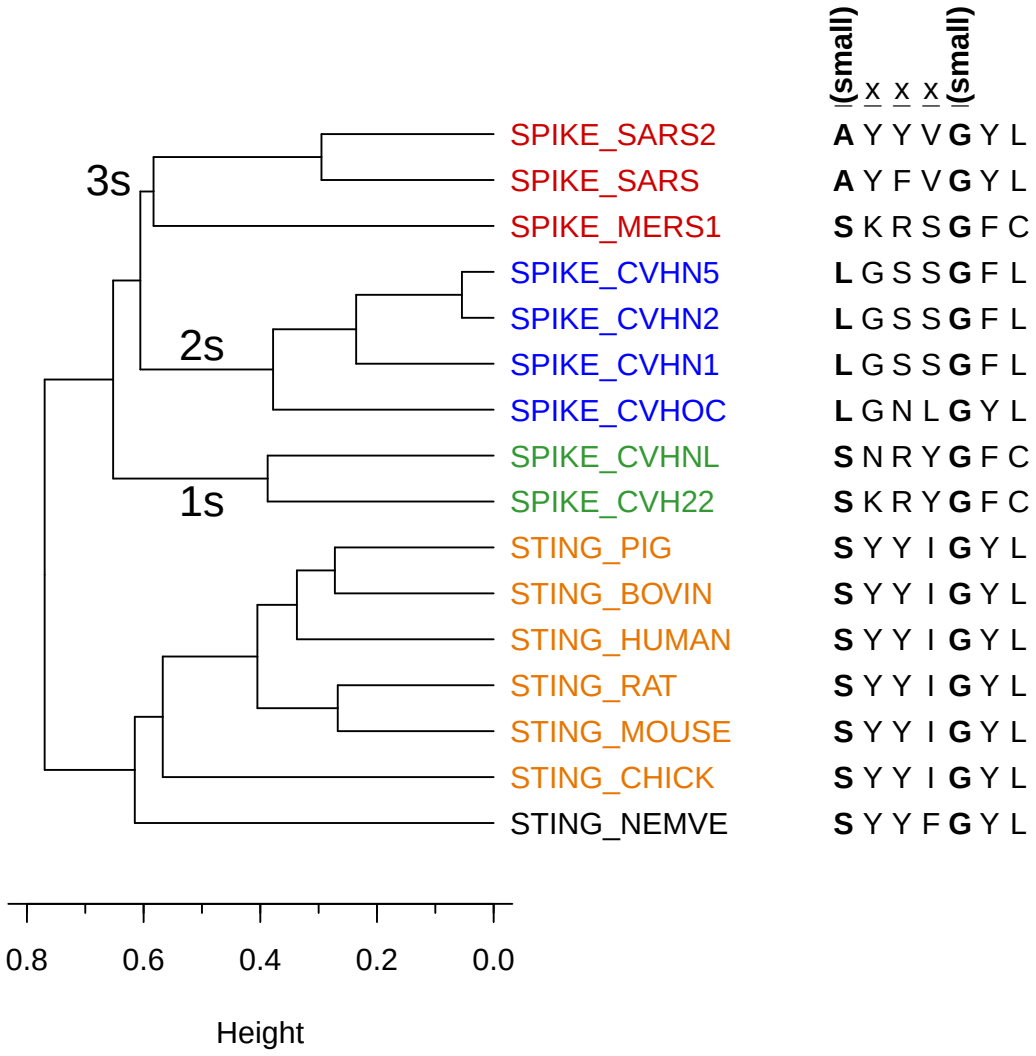

Fig. 2: Dendrogram of the phylogenetic relationships of the STING and the spike proteins of human beta-coronaviruses based on alignment of amino acid sequences generated by Clustal Omega and hierarchical cluster analysis using R. 'x' denotes any amino acid. The first small amino acid in the small-xxx-small 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.

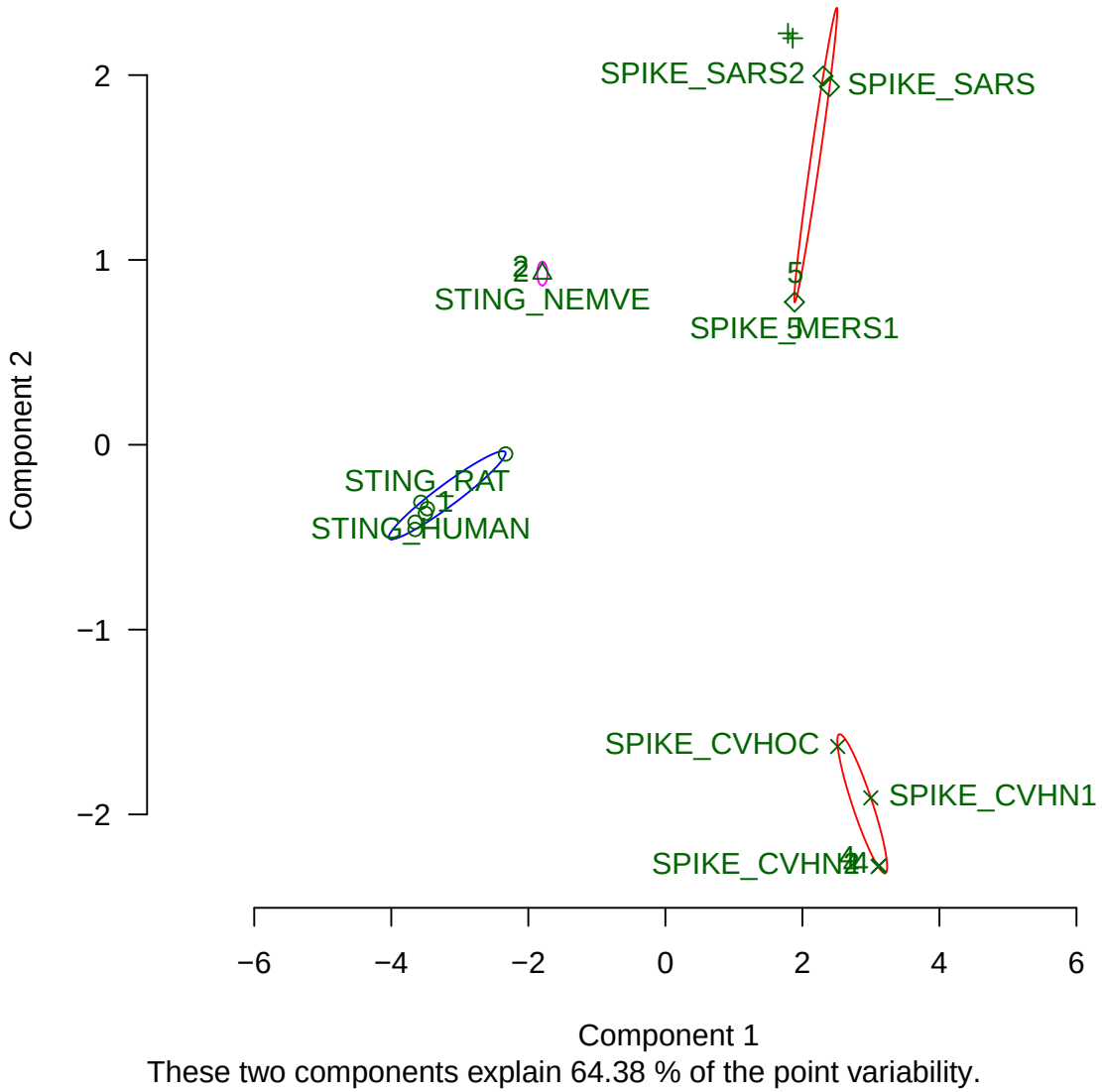

Fig. 3: Venn diagram of the STING and the spike proteins of human beta-coronaviruses showing the 3s as a distinct group as if it is evolving towards a new group of beta-coronavirus. The same STING and the spike proteins shown in Fig. 2 were used for this Venn diagram.

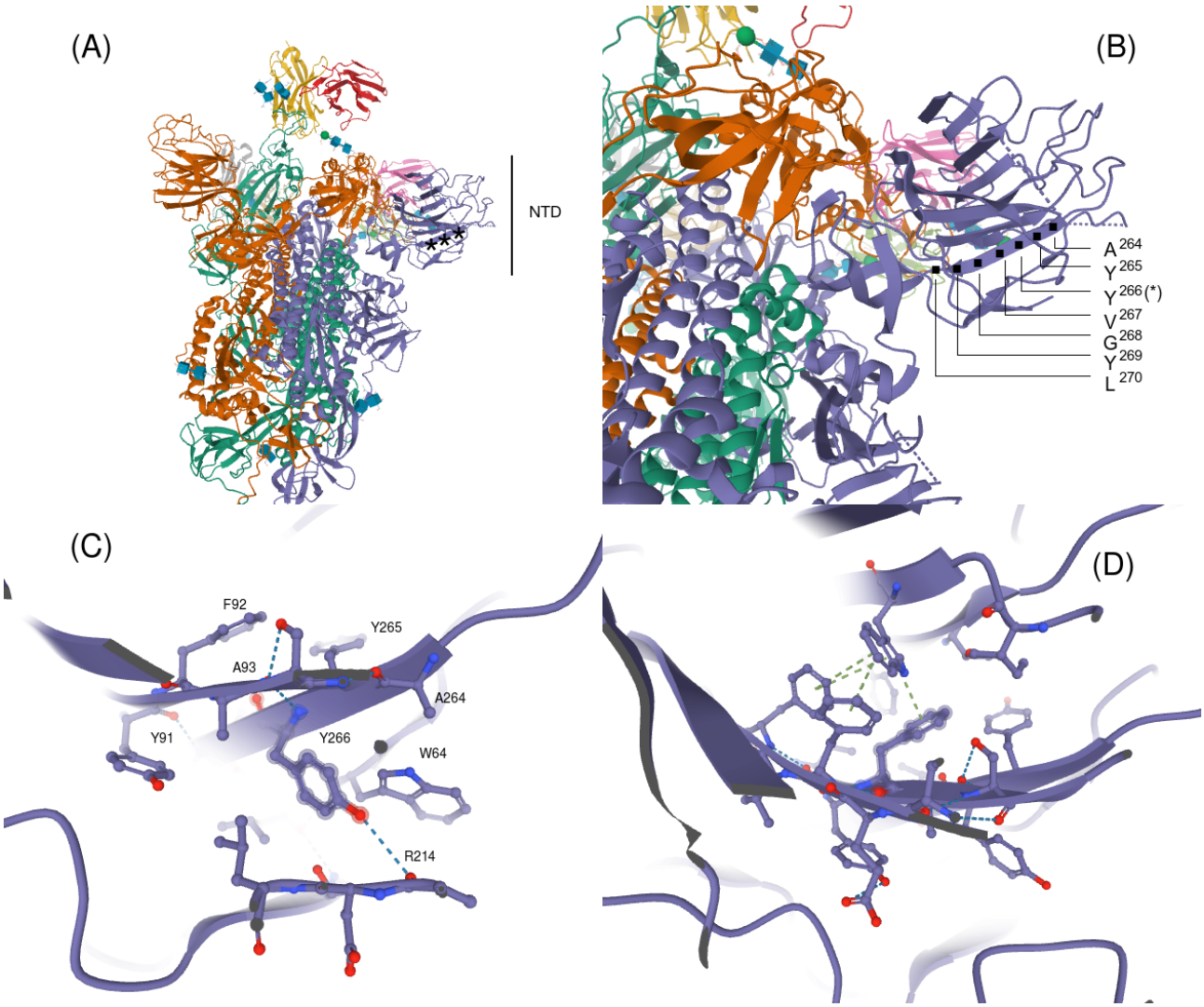

Fig. 4: The unique motif is, (A) on the NTD of the spike protein marked with stars, (B) located on one of the beta-sheets with a finger like loop extending outside, (C) there is a hydrogen bond between the R^{214} of the neighbour beta-sheet and Y^{266} which is found only on the SARS-CoV-2 member of the 3s group.Y^{266} makes an aromatic cage with the W^{64} which is found only on SARS-CoV-2, (D) another aromatic cage around A^{93}, the unique motif is surrounded with aromatic cages.

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 W^{64} 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.