

Biovacc-19:

A Candidate Vaccine for Covid-19 (SARS-CoV-2) Developed from Analysis of its General Method of Action for Infectivity

B. Sørensen(1), A. Susrud(1), A.G.Dalgleish(2)

1) Immunor AS, Oslo, Norway

2) Foundation Professor of Oncology, St. George's, University of London, Institute of Infection and Immunity, London, United Kingdom

ABSTRACT

This study presents the background, rationale and Method of Action of Biovacc-19, a candidate vaccine for Covid-19, now in advanced pre-clinical development, which has already passed the first acute toxicity testing. Unlike conventionally developed vaccines, Biovacc-19's Method of Operation is upon non human-like (NHL) epitopes in 21.6% of the composition of SARS-CoV-2's Spike protein, which displays distinct distributed charge including the presence of a charged furin-like cleavage site. The logic of the design of the vaccine is explained, which starts with empirical analysis of the aetiology of SARS-CoV-2. Mistaken assumptions about SARS-CoV-2's aetiology risk creating ineffective or actively harmful vaccines, including the risk of Antibody-Dependent Enhancement (ADE). Such problems in vaccine design are illustrated from past experience in the HIV domain. We propose that the dual effect general method of action of this chimeric virus's spike, including receptor binding domain, includes membrane components other than the ACE2 receptor, which explains clinical evidence of its infectivity and pathogenicity. We show the non-receptor dependent phagocytic general method of action to be specifically related to cumulative charge from inserted sections placed on the SARS-CoV-2 Spike surface in positions to bind efficiently by salt bridge formations; and from blasting the Spike we display the non human-like epitopes from which Biovacc-19 has been down-selected.

Design methodology and parameters

Although no other Covid-19 vaccine design programme appears to follow this methodology, we believe, from experience, that successful vaccine design logically starts with a thorough understanding of the aetiology of the target virus which appears in this case to be quite singular. In consequence of our researches and therefore unlike conventionally developed vaccines, Biovacc-19's Method of Operation is solely upon non human-like (NHL) epitopes which are 21.6% of the composition of this coronavirus's Spike protein. The Spike displays distinct distributed charge including a charged furin-like cleavage site. Following principles previously employed to design a therapeutic Human Immunodeficiency Viruses (HIV) therapeutic vaccine (Vacc4x), we have therefore first examined and publish here sequences and alignments of the SARS-CoV-2 Spike protein looking for unique properties of this virus that can be exploited for successful epitope presentation. We have also reviewed, and publish here, relevant cryo electron microscopy results, and structure and function relevant Cys-Cys loops, to discover what is - and is not - revealed at amino-acid level analysis of sequences by previous authors and to elicit the general mode of action for infectivity of this virus. This methodology has two parts.

In order to formulate this fact-based account of its general mode of action, first we present an explanatory model for the difference between SARS-CoV and SARS-CoV-2 supported by substantial physical/chemical data from different classes of convergent sources. 3D models and sequence analyses have been used to reveal properties associated with specific amino acids/clusters of amino acids of both Spike and Co-receptors. Second, we propose a theoretical explanation which future experimental model systems will be able to study to elucidate further structural details as well as to provide evidence for detailed mechanistic explanations.

These data reveal the biological structure of SARS-CoV-2 Spike and confirm that accumulated charge from inserts and salt bridges are in surface positions capable of binding with cell membrane components other than the ACE2 receptor. We have also looked at the naked coronavirus spike protein as a concept for the basis of a vaccine, which we have rejected because of high risk of contamination with human-like epitopes.

Analysis of the Spike protein of SARS-CoV-2 shows 78.4% similarity with human-like (HL) epitopes. For the avoidance of confusion, a standard protein blast searches for functionalities and homologies to other proteins.

However, antibodies can only recognize 5-6 amino acids and therefore a 6 amino acid rolling window search for antibody epitopes was performed. A search so tailored to match against all human known proteins will give a 78.4% human similarity to the SARS-CoV-2 Spike protein, i.e if all epitopes on the 1255 amino acid long SARS-CoV-2 Spike protein can be used by antibodies then there will be 983 antibody binding sites which also could bind to epitopes on human proteins. This is what we did and found.

We were in the minority of vaccine designers with regard to Human Immunodeficiency Viruses (HIV) vaccine development, having concluded that a vaccine based on the envelope gp120 would not be effective. We proposed instead using the unique gag proteins as the basis of the Vacc4x vaccine which has been shown to induce robust immune responses and reduce the HIV viral load in several multicentre studies (Pollard *et al.*, 2014; Huang *et al.*, 2016; Huang *et al.*, 2017; Huang *et al.*, 2018). It is thirty six years since the world was promised an HIV vaccine that would be ready in eighteen months. We correctly predicted the failure of all three major HIV/AIDS vaccines over those years, and specifically the danger of poor immune responses to conserved human-like domains and antibody-enhanced infectivity to high mutating domains. Earlier this year, the latest South African trial was terminated due to futility in preventing HIV transmission (UNAIDS, 2020). From our past HIV experience, we therefore observe that in the present context, any vaccine design based on the whole Spike protein of SARS-CoV-2 may not be immunogenic due its high human similarity compared to a vaccine with specifically selected NHL epitopes, such as Biovacc-19 does - and is.

Covid-19 candidate vaccines designed without appreciating these problems may run similar risks to those experienced with HIV vaccines that failed to show protection. The possibility of inducing autoimmune responses or antibody-dependent enhancements, needs to be carefully guarded against because there is published evidence that an HIV candidate vaccine has actually enhanced infectivity (Duerr *et al.*, 2012): "Vaccinations were halted; participants were unblinded. In *post hoc* analyses, more HIV infections occurred in vaccinees vs placebo recipients in men who had Ad5-neutralizing antibodies and/or were uncircumcised. Follow-up was extended to assess relative risk of HIV acquisition in vaccinees vs placebo recipients over time". Such antibody-dependent enhancement (ADE) has been observed for coronaviruses in animal models, allowing them to enter cells expressing FcγR. ADE is not fully understood: however, it is suggested that antibody-dependent enhancements may come as a result of amino acid variability and antigenic drift (Ricke *et al.*, 2020; Negro *et al.*, 2020).

#### Adjuvants are not secondary considerations

Conventional vaccine methodologies tend to deal sequentially with the choice of adjuvant after the primary design work has been achieved. In contrast, we believe that the two aspects of design are indissoluble and that adjuvant choice is *ab initio* an essential aspect of a successful vaccine design. That is because it has been observed that with the right adjuvant there can be a valuable inverse correlation with infectivity, with morbidity and with fatality in published cohorts (Lam *et al.*, 2017).

Most adjuvants have a strong T-Helper 2 (TH-2) bias in order to achieve a good neutralising antibody response. Given the results of our research into the SARS-CoV-2 aetiology, we posit that an adjuvant is required that specifically activates innate and cell mediated immunity which will give the necessary enhancement in TH-2 response to the peptide specific epitopes.

It has been known for several years that Bacillus Calmette–Guérin (BCG) can enhance a TH-1 response and can be used as an adjuvant for cell-based vaccines such as the melanoma cell-based Cancervax pioneered by Donald Morton (Faries *et al.*, 2017). However, it cannot be used repeatedly because it will induce a non-specific humoral response that will enhance cancer progression. In an attempt to overcome this limitation, Standford and Rook developed a programme for improved agents that could replace Bacillus Calmette–Guérin in the control of Tuberculosis and noted that heat killed *Mycobacterium vaccae* had the ability to enhance TH-1 responses and suppress the humoral response, making it a therapeutic agent in its own right (Bourinbaier *et al.*, 2019). Further research showed that *Mycobacterium Obuense* was superior to *Mycobacterium vaccae* and was much easier to manufacture to good manufacturing practice standards. Therefore it was selected by Immodulon as the immune modulator of choice for cancer studies.

Immodulon (IMM-101) is a systemic immune modulator containing a suspension of heat-killed whole cell *Mycobacterium obuense*, a rapidly dividing, environmental, harmless saprophyte. The heat-killing treatment during manufacture safeguards patients from side-effects associated with delivering live or attenuated organisms. IMM-101 has been successfully used in stage 4 melanoma trials as a single agent (Stebbing *et al.*, 2012). Furthermore, it has shown an ability to enhance responses to Check Point Inhibitors (CPIs) (Dalgleish *et al.*, 2018) as well as to increase the effectiveness of Gemcitabine in a randomised study in advanced pancreatic cancer patients

where a significant survival and quality of life benefit was seen on the IMM-101/Gemcitabine arm versus Gemcitabine alone (Dalglish *et al.*, 2016).

Of particular relevance to the development of Biovacc-19 is that the majority of patients who have received IMM-101 for advanced cancer have reported that, subsequently, they have not experienced the usual seasonal colds or influenzas from which they had previously suffered. IMM-101 has been administered to over 300 patients without a single serious side effect and is therefore a safe product to consider as a single priming adjuvant. Unpublished observations by these authors (Dalglish) have suggested that this agent was the best adjuvant for producing effective immune responses against melanoma antigens.

A study to define an optimal antigen/adjuvant combination is in progress using a combination of Immodulon as a separate priming adjuvant together with a formulation of the vaccine peptides and an adjuvant. The study will define the balance between the adjuvant and antigen in such a way that one or a maximum of two doses will be needed to obtain protective immunity.

#### The Biovacc-19 design concept and analysis of the target virus's general method of action for Infectivity

The Covid-19 vaccine Biovacc-19 is a peptide vaccine designed to develop antibodies to those parts of the SARS-CoV-2 Spike protein which are engaged in binding and infecting cells.

The human SARS spike protein consists of two parts (Uniprot – P0DTC2) <https://www.uniprot.org/uniprot/P0DTC2>. The S1 part attaches the virion to the cell membrane by interacting with host receptors like human ACE2 (Uniprot – Q9BYF1) <https://www.uniprot.org/uniprot/Q9BYF1> and with attachment receptors such as C-type Lectin domain family 4 member M (CLEC4M)/ Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin(DC-SIGNR) also known as CD209 (Marzi *et al.*, 2004; Uniprot – Q9H2X3; Uniprot – Q9NNX6) <https://www.uniprot.org/uniprot/Q9H2X3> and <https://www.uniprot.org/uniprot/Q9NNX6> ) thereby introducing the virus into the endosomes of the host cell, where the fusion peptide of the S2 part is unmasked and activated membrane fusion within endosomes occurs to permit virus replication in cytosol.

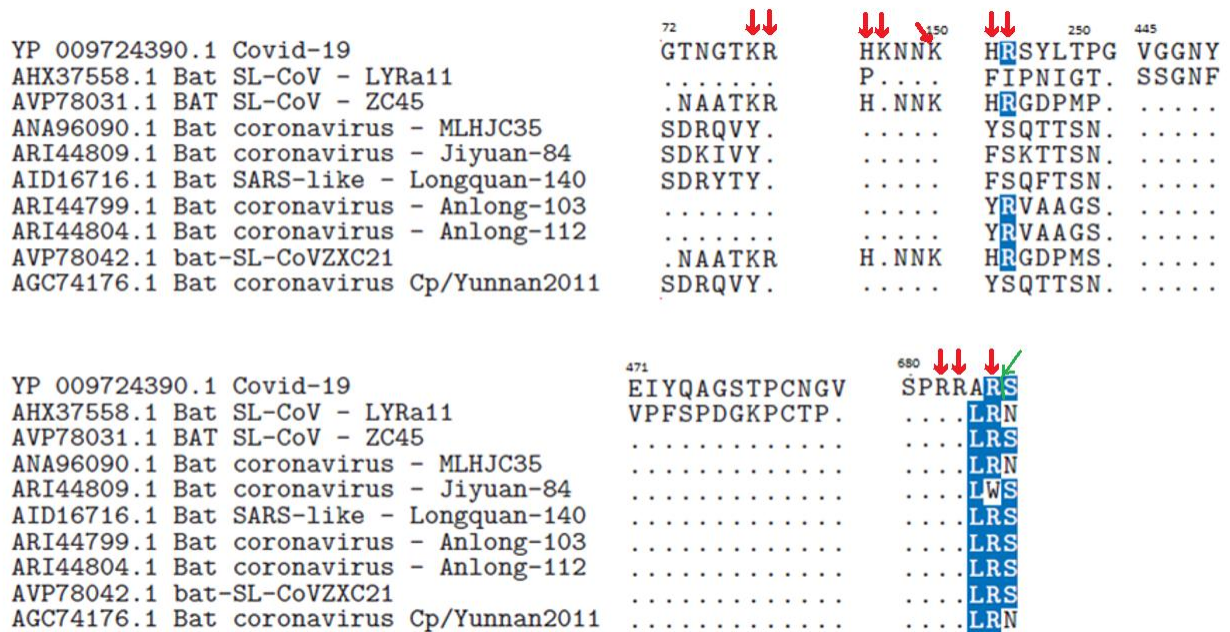
The SARS-CoV-2 spike is significantly different from any other SARS that we have studied (Lu *et al.*, 2020). The additional charge it carries (SARS-CoV-2 S1, isoelectric point (pI) pI=8.24 versus human SARS-CoV S1, pI= 5.67) will strongly improve the interactions with the receptor C-type lectin tail on CLEC4M/DC-SIGNR, which may, by itself, mediate the endocytosis of pathogens by acting as an attachment receptor, as happens for a number of other highly pathogenic viruses such as Ebolavirus, Marburg, HIV-1, Hepatitis C, Measles, human CytoMegalovirus, Influenza and others (Marzi *et al.*, 2004; Uniprot – Q9N2X3; Uniprot – Q2NNX6).

It is well documented that the receptor binding domain of the SARS-CoV-2 spike protein uses the ACE2 receptor. But clinical findings discussed below observed in Covid-19 patients suggest that other receptors for attachment such as CLEC4M/DC-SIGNR may be involved as well. We have investigated and sustained this supposition from amino-acid scale bio-chemical analysis.

Cumulative data suggests that the general method of action of this chimeric virus includes membrane components other than the ACE2 receptor, which may explain clinical evidence of its infectivity and pathogenicity. Data shows the non-spike receptor binding domain dependent phagocytic general method of action to be specifically related to cumulative charge from inserted sections on the SARS-CoV-2 Spike (see Fig 1) poised to form salt bridges with attachment receptors. This suggests that attachment to such previously reported membrane proteins has been enhanced directly due to the basic and positive charged inserts in the Spike protein together with other basic and positive charged amino acid substitutions enabling formation of salt bridges with the receptor CLEC4M/DC-SIGNR or, indirectly, by the additional salt bridges formed between the positive charged amino acids and negative charged phospholipids on the cell membrane.

Positive charged amino acids are inserted into peptides and proteins to enhance cell affinity and can also be used for transport of peptides and proteins through the cell wall (Richard *et al.*, 2003; Thorén *et al.*, 2000; Åmand *et al.*, 2011). In addition, these positive charges may be used for co-receptor binding where the opposite negative charge is available.

It is a matter of fact that there are unique inserts in the SARS-CoV-2 spike protein when they are aligned with other SARS-CoV sequences as shown in (Zhou *et al.*, 2020).



**Figure 1.** Alignments of Corona virus Spike protein inserts .

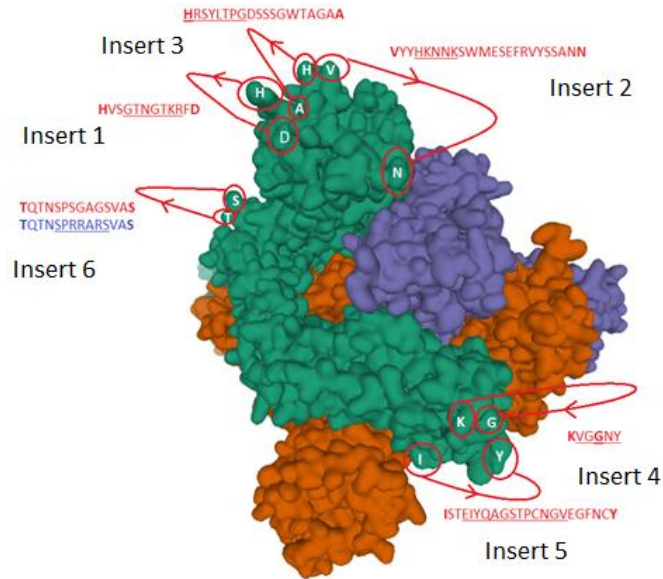
Figure 1 shows 6 alignments with inserts. The first 5 inserts are pointed out by (Zhou *et al.*, 2020) and located near/around position 72, 150, 250, 445, 471 while the insert around 680 is pointed out by (Coutard *et al.*, 2020) as a furin-like cleavage site with cleavage between R and S. Apart from insert 4 and 5, these inserts are all basic inserts. The red arrows point out the basic amino acids. The green arrow and line point out the furin-like cleavage site.

It has been recently suggested that new X-Ray crystallography can assist in one of our key investigations: the docking of spike with receptor. (Jian *et al.* 2020) Unfortunately we cannot agree. The *Nature* paper entitled “Structural Basis of Receptor Recognition by SARS-CoV-2” does not, in fact, represent a true structure of the Spike SARS-CoV-2 trimer apart from a modified part of the receptor binding motif (RBM). It uses the structure and the sequence for SARS-CoV deposited on 1 August 2005 as the backbone and then creates a chimera with the RBM (437-508) of SARS-CoV-2 modified and inserted. This is a confusing structural determination, representing neither virus. The authors were keen to focus on the importance of the ACE-2 receptor: “we improved the ACE2-binding affinity of the chimeric RBD by keeping a short loop from the SARS-CoV RBM.” But then four of the six new charged inserts (1,2,3 &6) outside the RBM were also excluded from their chimera, and the Cov-2 specific Cys538-Cys590 bridge which brings in additional charge from 526-560 (with pI=10.03) via the Cys391-Cys525 right next to the RBM.

By these excisions the very essence of the novel structure and functionalities of SARS-Cov-2, and hence of its general mode of action for infectivity, is obscured. Therefore using this structure from this paper would be quite misleading.

The SARS-CoV-2 3D structures have been usefully determined for SARS-CoV-2 trimer Spike protein by (Walls *et al.*, 2020).

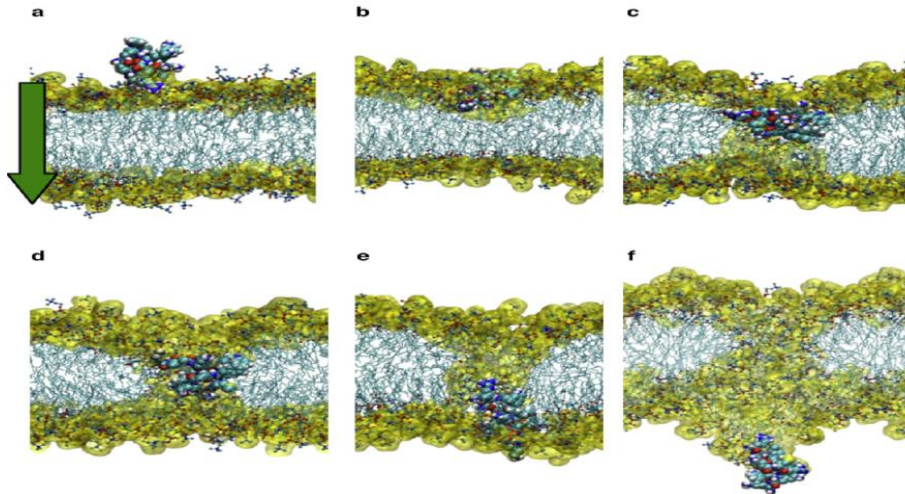




**Fig. 2.** The identified inserts examined in the PDB 6VXX electron microscopy structure (Walls *et al.*, 2020) The sequences highlighted in red could not be found in the cryo-electron microscopy structure data. The 6 aligned sequences in Fig. 1 are underlined in the missing sequences. Bold amino acids indicate first and last amino acids used to build the structure where the missing part is in between. Insert 6 did not have the same sequence in 6VXX as in the reference Sars-CoV-2 sequence. The authors stated that a designed mutated strain lacking the furin cleavage site residues was used.

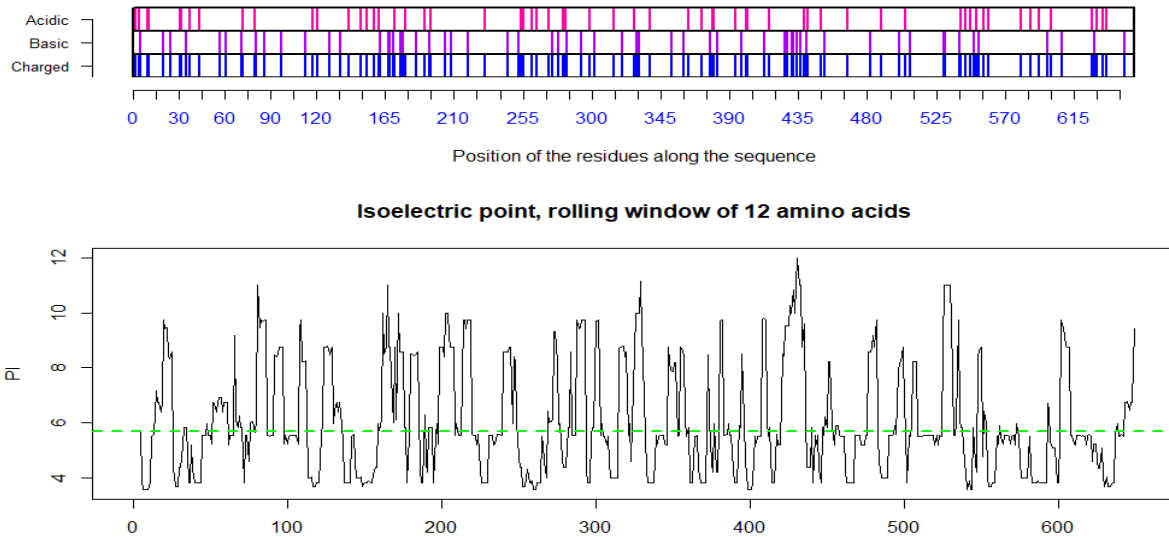
Our findings confirm (Coutard *et al.*, 2020) that the SARS-CoV-2 contains a furin-like cleavage site absent in CoV of the same clade. Also in Fig 1, Coutard highlight that enriched basic charge associated with this cleavage site are found in a number of viruses such as Human Immunodeficiency Viruses, Influenza, human CytoMegal Virus (Herpes) and Respiratory Syncytial Virus, Yellow fever, Zika and Ebola. Coutard *et al* furthermore state that, "conversely, the highly pathogenic forms of influenza have a furin-like cleavage site cleaved by different cellular proteases, including furin, which are expressed in a wide variety of cell types allowing a widening of the cell tropism of the virus." Furthermore the insertion of a multibasic motif RERRRKKR↓GL at the H5N1 hemagglutinin HA cleavage site was likely associated with the hyper-virulence of the virus during the Hong Kong 1997 outbreak. Extensive clinical evidence in this pandemic suggests that SARS-CoV-2 poses such widened cell tropism.

The mechanism of action linked to such basic Arginine rich domains is known as the binding of cell-penetrating peptides (Thorén *et al.*, 2000). The important point to grasp is that such positively charged amino acids need to be located in such a way that they span four amino acids (or more) in length to act as an initial membrane anchor. Use of such positive charged vaccine peptides allows for attachment and/or direct cell uptake depending on the net charge present in the peptide (Åmand *et al.*, 2011; US patent – US9950811B2n). The present authors have used such basic properties in uploading vaccine peptides to cells (typically macrophages and dendritic cells). We have found that more than 3 Arginines are required for uptake. In addition this charge needs to be distributed over the peptide. (Yesylevskyy *et al.*, 2009). Fig. 3 (a-f) illustrate the process in dynamic sequence from attachment to cell membrane until complete cell penetration.

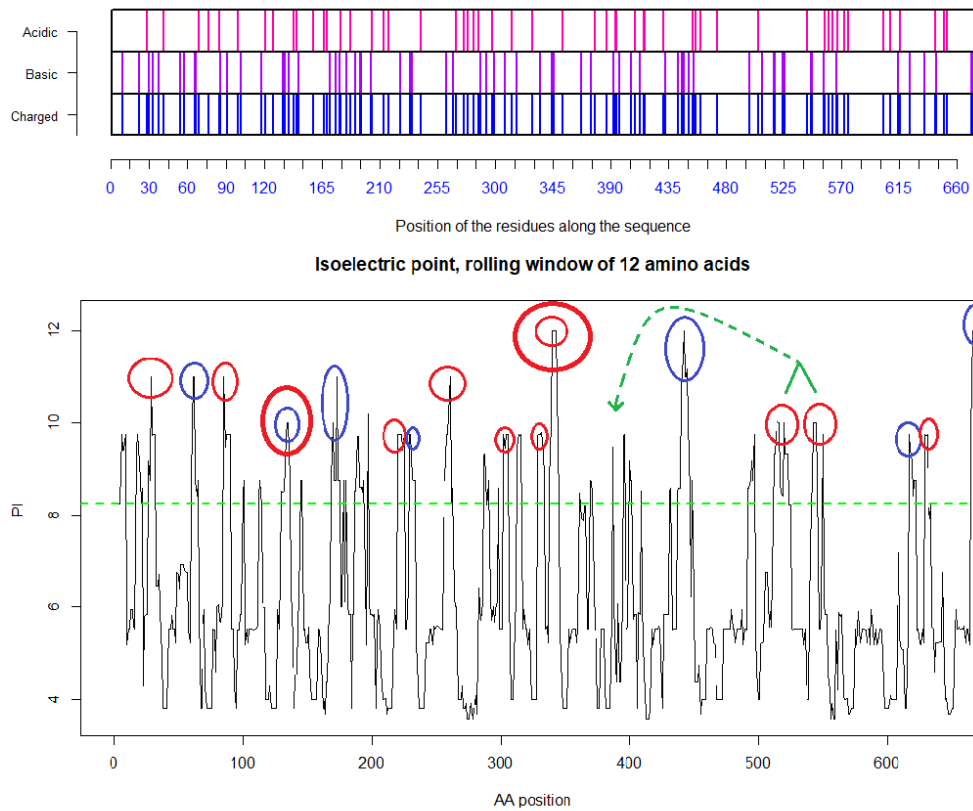


**Fig. 3.** from (Yesylevskyy *et al.*,2009)

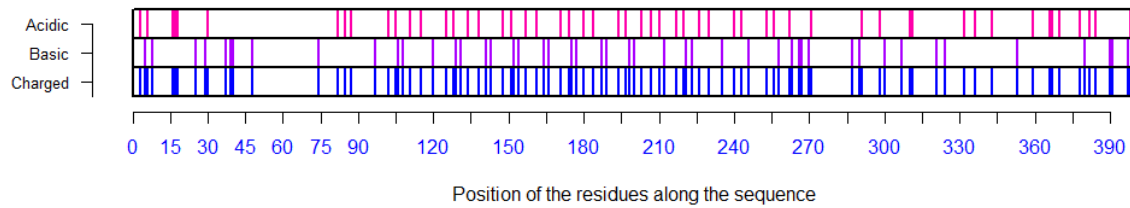
For SARS-CoV-2 the sequence (SPRRAR|S) is longer and more basic than the SARS-CoV (TVSLLR|S) and hence is more potent (Coutard *et al.*, 2020). The mode of action of a furin-like cleavage site is, following endosomal encapsulation, to facilitate attachment and penetration of the inside wall of the endosome to release the uncoated virus into cytosol where it can start replication and release its full pathogenic potential. In the context of SARS-CoV-2 having a general elevated pI with additional charge located in the receptor binding domain as shown below, will make it fit for membrane penetration.



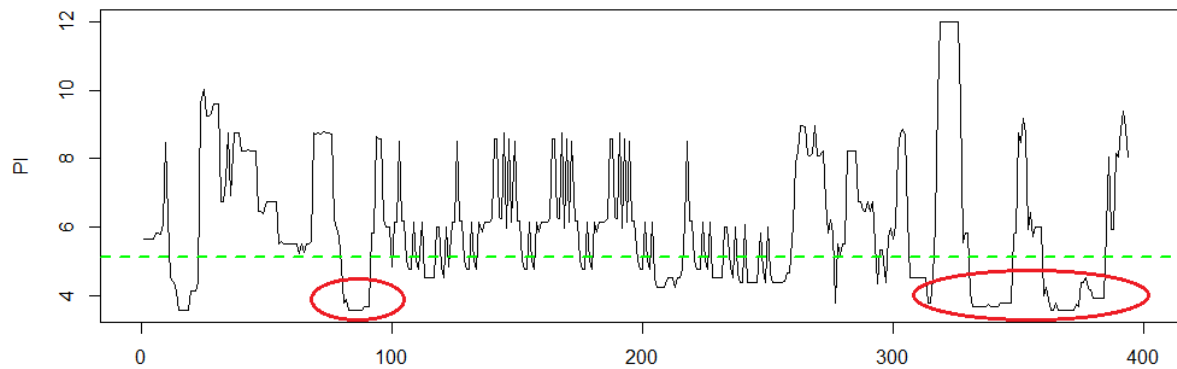
**Fig. 4.** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on human SARS-CoV shows a dominance of acidic amino acids giving a pI= 5.67 (green dotted line). Repulsion will be observed in the presence of a protein/peptide domain with a similarly low pI=5.12 as found for ELEC4M/DC-SIGN.  
 Ref.: <https://www.uniprot.org/uniprot/P59594> sp|P59594|14-667 [SARS-CoV Urbani]



**Fig.5** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on SARS-CoV-2 shows a dominance of basic amino acids. The red circles are verified to be surface exposed while the blue circles are missing in PDB: 6VXX. Non-labelled peaks are not surface exposed. Above the green dotted line (Isoelectric point for the SARS-CoV-2 Spike protein S1  $pI=8.24$ ) the peaks contain basic residues which can form salt bridges in the presence of acidic amino acids such as in the receptor CLEC4M. The Cys131-Cys166 and Cys336-Cys361 loops are highlighted by as bold red/blue and red/red circles. Furthermore the Cys391-Cys525 bridge: a highly basic domain (526-560) with  $pI=10.03$  is moved forward and sits next to the Receptor Binding Motif (408-508) as part of a Cys538-Cys590 loop, as indicated by the green arrow. Ref.: <https://www.uniprot.org/uniprot/P0DTC2> sp|P0DTC2| 13-685 [SARS-CoV-2]



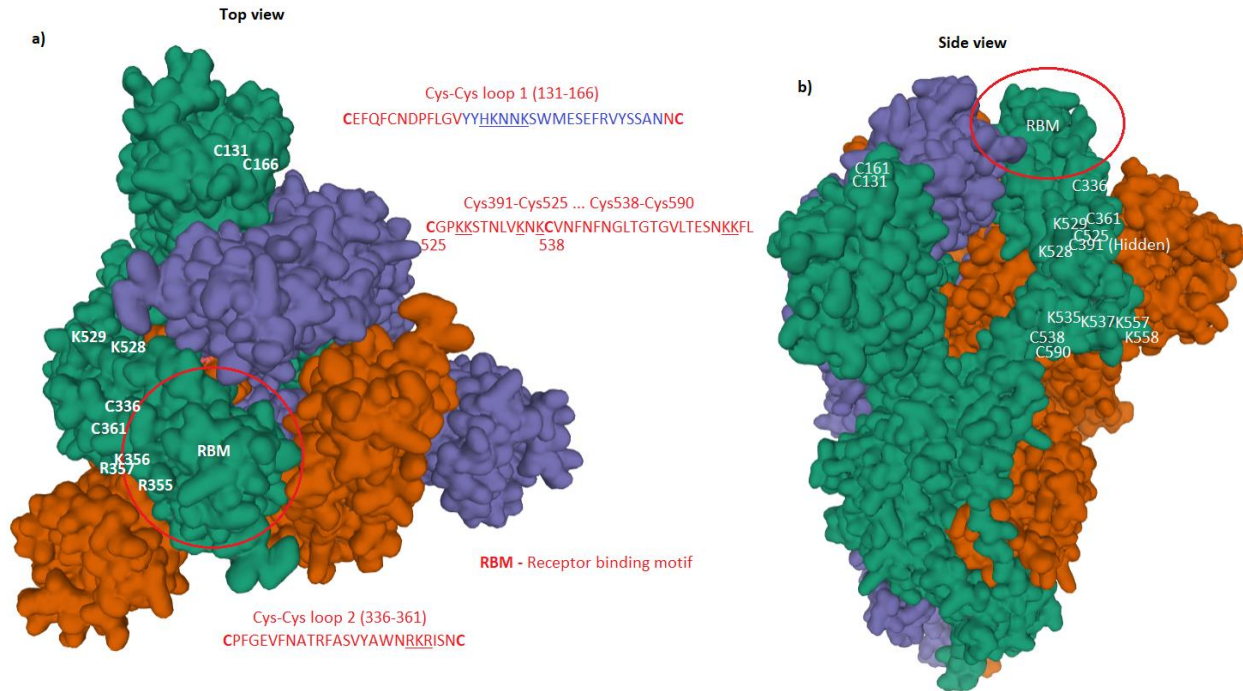
**Isoelectric point, rolling window of 12 amino acids**



**Fig.6** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on the receptor CLEC4M having  $pI=5.12$ . The ovals identify the most likely contributors to the salt bridges between the acidic domain, here on CLEC4M and the highlighted basic domains on the SARS-CoV-2 S1 protein in Fig.3 above. The two disulfide bonds Cys296-Cys389 and Cys368-Cys381 present in the C-type lectin C-terminal part of CLEC4M pull back the Cys368-Cys381 acidic domain to position 296 and make that domain a highly condensed with acidic amino acids ready for formation of salt-bridges with the basic amino acids in S1 of SARS-CoV-2 (Marzi *et al.*, 2004). Further investigations might possibly show that other amino acids on the S1 SARS-CoV-2 are involved in such attachment receptor binding. Ref.: CLEC4m/DC-SIGN - also referred to as CD209 <https://www.uniprot.org/uniprot/Q9H2X3>

The co-receptor dependent phagocytic general method of action of SARS-CoV-2 appears to be specifically related to cumulative charge: please refer to SARS-CoV-2 peaks above  $pI=8.24$  (Fig.5) compared to human SARS-CoV (Fig 4). These basic domains - partly inserted and partly substituted amino acids - explain the salt bridges formed between the SARS-CoV-2 Spike and its co-receptors on the cell membrane. Indeed, these data suggest that the infectivity of SARS-CoV-2 is best explained by this cumulative charge associated with these basic charged domains, enabling extra salt bridges to attach to membrane components as well as to the membrane itself.





**Fig. 7** Spike trimer (a) top view and (b) side view. The specific receptor binding motif (RBM) is located on the sequence (437-508), while the receptor binding domain (RBD) has a broader location (319-541) ref. <https://www.uniprot.org/uniprot/P0DTC2>. The charged cysteine associated domains are Cys131-Cys166, Cys336-Cys361, Cys391-Cys525...Cys538-Cys590. As can be seen, there is a high concentration of positive charged surface exposed amino acids within the receptor domain next to the receptor binding motif.

Cys131-Cys166 loop 1 is partially missing in the electron microscopy structure. The missing section is highlighted in blue in the Figure. No information is given to explain why the SARS-CoV-2 sequence was changed for the electron microscopy. When preparing samples for examination, perhaps it may have been necessary to alter the surface of trimer containing a surplus of hydrophilic and basic/positive amino acids by removing sequences as highlighted in Fig.2 and Fig.5. By making visible the complete repertoire in this paper we are therefore able to display data that, taken together, present the general mechanism of action necessary for understanding how the SARS-CoV-2 virus enter cells and hence where and how to attack it with a vaccine. Absent the complete repertoire, it is significantly more difficult to understand the general method of action and therefore to find a vaccine or a therapy.

Underlined sequences are identified charged amino acids contributing to attachment/co-receptor binding such as CLEC4M/DC-SIGNR(CD209). The exposed and positive charged amino acids (in red capital letters) contributing to salt bridges are located within the Cys131-Cys166 loop 1: (KVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTF) and Cys336-Cys361 loop 2: (NLCFPGEVFNATRFASVYAWNKRKRISNCVA). This second charged domain (fig 7.a) is positioned right next to the receptor binding motif (437-508) via the adjacent Cys379-Cys432 bridge and can therefore facilitate binding to opposite charged attachment receptors. The domain 526-560 (pI=10.03) is brought into the receptor binding domain via Cys391-Cys525 bridges furthermore significantly enhances the overall charge on the receptor binding domain (CGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFL). This additional Lysine (K) driven charge on SARS-CoV-2 coming from the domain 526-560 does not exist on SARS-CoV due to the unique SARS-CoV-2 Cys538-Cys590 bridge. The RBD Cys131-Cys166 loop1 on SARS-CoV-2 has a pI=5.34 while the similar Cys128-Cys159 loop2 on SARS-CoV has a pI=4.36 . For the SARS-CoV-2 Cys336-Cys361 domain we have found a pI of 9.5 whereas for the similar SARS-CoV 323-348 we have a pI of 8.82. As shown here the surface exposed cysteine loops on the RBD have consistently higher pI for SARS-CoV-2 than for SARS-CoV.

This method of action using ACE2 as main receptor and ELEC4M/DC-SIGN as co-receptors is similar to what is observed for HIV and its use of CD4 as main receptor and the V3 Cys-Cys loop docking on the CCR5/CXCR4 co-receptors. A substitution of the amino acid Cys348 with an Ala for SARS-CoV leads to complete loss of human ACE2 binding in vitro (Wong *et al.*, 2004; Uniprot - P59594). The construction of these Cys-Cys bridges in SARS-CoV-2 are similar. So would it be reasonable to assume that a similar substitution of Cys361 with an Ala for SARS-CoV-2 would lead to a similar loss of human ACE2 binding? The answer is probably yes for ACE2 binding. But due to the additional cumulative charge it would still be able to attach to the co-receptor.

This is an important finding for our vaccine design. Furthermore, that SARS-CoV-2 can enter cells without using the ACE2 but also by promiscuous attachment has implications for understanding disease epidemiology, for treatment drug method of action as well as for vaccine development strategies. Similar co-receptor observations have been made before. Wong S.K. *et al.* (2004) report interestingly the first 330 amino acids of the 769-residue S1 subunit of the mouse hepatitis virus (MHV) S protein is sufficient to bind carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), to the cellular receptor for the mouse hepatitis virus (MHV). Furthermore, a different region of the S1 domain of HCoV-229E, between residues 407 and 547, is sufficient to associate with the cellular receptor for this coronavirus, aminopeptidase N (APN, CD13).

In particular the insert in alignment 6 in the SARS-CoV-2 spike has three positive Arginines in combination with a Proline, which together secure the anchoring to the membrane (but not acting in the same way as a typical cell penetration peptide due to there being only four amino acids). Therefore these data show that the molecular structure of the SARS-CoV-2 Spike receptor binding domain, with its accumulated charge from inserts and salt bridges in surface positions, is capable of binding with cell membrane components. This is an essential key to understanding its potency.

Furthermore, the Covid-19 pandemic is revealing neurological, haematological and immunological pathogenicity in the virus which cannot be explained by the ACE2 receptor alone. Profuse clinical observations of loss of taste, smell, sore throat, dry cough and headache and severe stomach /gastrointestinal pain with diarrhoea arising in the pandemic are evidence that early phase Covid-19 is binding to the bitter/sweet receptors which also provides a perfect location for follow on transmission by coughing. How is it doing this? The answer to these clues is also important for our vaccine design because, as explained in the next section, it caused us to select and deploy epitopes accordingly in order to deny the virus these binding options.

Compromising the function of olfaction and bitter/sweet receptors will effect a timely release of products from the innate immune system, thereby enhancing infectivity and transmission. At the onset of infection, there is enhanced mucociliary transport of mucus from the nasopharynx or oropharynx. By swallowing innate immune products at the same time, they are disseminated on the airway surface (Workman A.D *et al.*, 2015). These immune products include directly anti-microbial compounds such as defensins, lactoferrin, cathelicidins, and lysozyme, in addition to reactive oxygen species (ROS) and nitric oxide (NO) that also display potent antimicrobial activity (Roper *et al.*, 2013). Several indirect pathways are activated as well, with the release of cytokines and chemokines that recruit the adaptive immune system and begin inflammatory cascades.

SARS-CoV-2 may use a direct route too. Attached to epithelial cells in the oral cavity, potentially it may also be transported together with food via the stomach to the intestine. As noted, clinicians report severe stomach /gastrointestinal pain with acute diarrhoea in Covid-19 patients. Such clinical indicators support our earlier analysis and also suggest that SARS-CoV-2 can use other attachment/co-receptors than ACE2.

Further collateral support for these hypotheses came in 2018 when (Zhou *et al.*, 2018) isolated a new coronavirus which they named SADS (Swine Acute Diarrhoea Syndrome). They investigated receptor usage in the intestines of infected piglets but could find no evidence of involvement of any of the three receptors known from previous SARS epidemics: angiotensin-converting enzyme 2 (ACE2), aminopeptidase N (APN) and dipeptidyl peptidase 4 (DPP4). As with SADS, the similar pathology associated with Covid-19 points in the direction of a different and more promiscuous attachment/ co-receptor like CLEC4M/DC-SIGN driven by a positively charged Spike trimer surface.

We have earlier explained the enhanced presence of basic amino acids in the inserts such as Lysine (K) and Arginine (R) and their association with enhanced pathogenicity in other pathogens like the 1997 H7N1 Hong Kong Flu (Coutard *et al.*, 2020; Kido *et al.*, 2012). We noted above the critical importance of understanding that cumulative positive charge associated with the inserted short sections has the effect of enabling extra salt bridges to attach to the membrane. Under this general method of action, this combination of basic amino acids in the SARS-CoV-2 Spike

binds to cells in the upper airways. Its high infectivity is associated with olfaction and taste; and systemic release of the virus explains the clinical findings associated with destruction of erythrocytes (Liu *et al.*, 2020), T-cells and cells associated with neuropathological conditions (Henry *et al.*,2020).

From all these convergent data, we therefore posit that the general method of action for SARS-CoV-2 is indeed as a co-receptor dependent phagocytic process: the cell envelops the virus due to its opposite charged binding to co-receptors on the cell membrane such as CLEC4M/DC-SIGNR and possibly to the membrane itself. But furthermore, simultaneously, it is capable of binding to ACE2 receptors in its receptor binding domain: in fact SARS-CoV-2 is possessed of dual action capability.

Once the virus is phagocytosed it will take over cell machinery, replicate and kill the host cells and rapidly increase systemic infection. The virus is thereby killing off erythrocytes which would account for the hypoxia observed in advanced patients, which leads to the shortage of oxygen uptake which may eventually prove fatal. Furthermore, clinically identified involvement of this class of olfaction and bitter/sweet receptors as potential co-receptor(s), or as alternative receptor sites to ACE2, has possible implications for binding, the replication, metabolism and pathology of the SARS-CoV-2 virus.

Although undertaken for the purposes of vaccine design, these virological findings have major positive implications for enhanced treatment options for advanced or relapsed Covid-19 patients, five of which such treatments one of us in his clinician role (Dalglish) has published elsewhere (Dalglish *et al.*, 2020).

#### Specific implications from the target virus's general method of action for design of Biovacc-19

In order to ensure that Biovacc-19 covers all the various cell receptor binding options, combined with our guiding criterion of using non-human like (NHL) epitopes, we systematically blasted (Uniprot – P0DTC2) the Spike protein. These NHL epitopes are displayed in Table 1

Peptide vaccine antigens used in Biovacc-19 are peptide strings with a total length of 30 to 36 amino acids consisting of epitopes placed in scaffolds similar to those described in US patent US9950811B2n (US patent – US9950811B2n). The precise sequences are not disclosed here.

Generation of antibodies requires T-Helper-2 (Th2) responses. By using a set of peptides spanning more than 130 amino acids it will be as if a medium sized protein was used as vaccine antigen resulting in large sequences variation and hence giving a surplus of Th2 epitopes.

The design of such synthetic vaccine peptides offers a further advantage. It will permit discrimination between antibody responses coming from natural infection and those coming as a result of vaccination.

A group of four of these vaccine peptides have been tested for acute toxicity with success.

**Table 1.** Non Human-Like (NHL) sequences found in SARS-CoV-2 spike protein S1. These sequences were obtained by blasting the Spike protein sequence using moving window of 6 amino acids in steps of 1 against the human protein sequence database on Uniprot (Uniprot – P0DTC2).

|          |           |            |             |
|----------|-----------|------------|-------------|
| 1 PLVSSQ | 41 SKHTPI | 81 GKIADY  | 121 GVLTES  |
| 2 SSQCVN | 42 LVRDLP | 82 KIADYN  | 122 TESNKK  |
| 3 VNLTRR | 43 RDLPQG | 83 DDFTGC  | 123 PFQQFG  |
| 4 LTRRTQ | 44 PQGFSA | 84 FTGCVI  | 124 QQFGRD  |
| 5 TRTQLP | 45 GFSALE | 85 GCVIAW  | 125 ADTTDA  |
| 6 PAYTNS | 46 VDLPIG | 86 LDSKVG  | 126 DTTDAV  |
| 7 VFRSSV | 47 PIGINI | 87 DSKVGG  | 127 TTD AVR |
| 8 VLHSTQ | 48 GINITR | 88 KVG GNY | 128 DAVRDP  |
| 9 LFLPFF | 49 LTPGDS | 89 GGN YNY | 129 AVRDPQ  |

|    |        |    |        |     |        |     |        |
|----|--------|----|--------|-----|--------|-----|--------|
| 10 | SNVTWF | 50 | TPGDSS | 90  | KSNLKP | 130 | DITPCS |
| 11 | VSGTNG | 51 | AAYYVG | 91  | PFERDI | 131 | ITPCSF |
| 12 | SGTNGT | 52 | GYLQPR | 92  | ISTEIQ | 132 | TPCSFG |
| 13 | TNGTKR | 53 | ALDPLS | 93  | STEIQY | 133 | FGGVSV |
| 14 | NGTKRF | 54 | LDPLSE | 94  | EIQYAG | 134 | GGVSVI |
| 15 | RFDNPV | 55 | PLSETK | 95  | STPCNG | 135 | SVITPG |
| 16 | VYFAST | 56 | SETKCT | 96  | TPCNGV | 136 | ITPGTN |
| 17 | ASTEKS | 57 | KCTLKS | 97  | PCNGVE | 137 | TPGTNT |
| 18 | STEKSN | 58 | TVEKGI | 98  | GVEGFN | 138 | PGTNTS |
| 19 | IRGWIF | 59 | TSNFRV | 99  | PLQSYG | 139 | TSNQVA |
| 20 | WIFGTT | 60 | FRVQPT | 100 | FQPTNG | 140 | VAVLYQ |
| 21 | FGTTLD | 61 | TESIVR | 101 | TNGVGY | 141 | QLTPTW |
| 22 | TTLDSK | 62 | SIVRFP | 102 | GVGYQP | 142 | STGSNV |
| 23 | LDSKTQ | 63 | PNITNL | 103 | LLHAPA | 143 | GSNVFQ |
| 24 | DSKTQS | 64 | ITNLCP | 104 | LHAPAT | 144 | FQTRAG |
| 25 | SKTQSL | 65 | NLCPFG | 105 | HAPATV | 145 | QTRAGC |
| 26 | KTQSL  | 66 | LCPFGE | 106 | APATVC | 146 | RAGCLI |
| 27 | VNNATN | 67 | ATRFAS | 107 | PATVCG | 147 | AGCLIG |
| 28 | ATNVVI | 68 | TRFASV | 108 | ATVCGP | 148 | AEHVNN |
| 29 | CEFQFC | 69 | SNCVAD | 109 | TVCGPK | 149 | IPIGAG |
| 30 | FCNDPF | 70 | VLYNVA | 110 | VCGPKK | 150 | AGICAS |
| 31 | CNDPFL | 71 | FKCYGV | 111 | GPKKST | 151 | SYQTQT |
| 32 | LGVYYH | 72 | KCYGVS | 112 | PKKSTN | 152 | QTQTNS |
| 33 | GVYYHK | 73 | CYGVSP | 113 | KKSTNL | 153 | TQTNSP |
| 34 | VYSSAN | 74 | YGVSP  | 114 | KSTNLV | 154 | TNSPRR |
| 35 | SANNCT | 75 | GVSPK  | 115 | VKNKCV |     |        |
| 36 | YVSQPF | 76 | VSPTKL | 116 | FNFNGL |     |        |
| 37 | VSQPFL | 77 | ADSFVI | 117 | FNGLTG |     |        |
| 38 | LEGKQG | 78 | QIAPGQ | 118 | NGLTGT |     |        |
| 39 | EGKQGN | 79 | APGQTG | 119 | TGTGVL |     |        |
| 40 | GKQGNF | 80 | TGKIAD | 120 | GTGVL  |     |        |

The Biovacc-19 vaccine is based on the method of action described. The vaccine peptides have therefore been selected from such NHL epitopes located in or close to the charged inserts and to the expected co-receptor binding locations outside the main receptor binding domain in addition to the NHL epitopes available for use within the receptor binding domain.

The benefit of using this strategy compared to conventional virus, Ribonucleic acid (RNA) or other vector-based vaccine systems is that the immune system will be guided *directly* to the epitopes which are relevant for virus neutralization. A further advantage of using NHL epitopes is that the immune system is free to mount robust, broad and long-lasting immune responses without being limited by local or even systemic immune-toxic reactions against our own human protein epitopes.

A comparison of the three most relevant properties/factors used for determining the probability of success in a vaccine design is presented in table 2.

**Table 2.** Probability of success of different vaccine technologies

| Property \ vaccine technology   | Synthetic peptides   | Vector based (Virus or RNA)  |
|---|--|--|
| Epitope targeting receptor binding domains (neutralization)                   | Specifically selected  | No selection.<br>The immune system of each individual will select and present the most dominating epitopes   |
| Epitope presented and likelihood for getting local or systemic toxicity (SAE) | Low to very low.<br>In theory no SAE should be observed since all epitopes are non-human like.             | Difficult to predict<br>The epitopes presented will be a mixture of human like (78,4%) and non-human like (21,6%) epitopes.  |
| Antibody-dependent enhancement (ADE)  | Low<br>Since the antibodies are directed towards the receptor binding domain and other co-receptor domains | Difficult to predict<br>In such vaccine designs, there is no innate guiding of where the antibodies should bind. However, due to continued boosting of these epitopes through life, there is an elevated risk for development of ADE which must be expected due to the fact that if the virus returns at a later date in a mutated form, having modified antigenic composition, partial binding may occur and hence result in ADE (Ricke D, <i>et al.</i> , 2020; Negro <i>et al.</i> , 2020). |

The patented vaccine peptides constituting Biovacc-19 are a combination of various sequences of 5 amino acids found on the Spike protein. The scaffold design of the vaccine peptides will create “in-between” epitopes which will facilitate discrimination between vaccine induced antibodies and antibodies which result from exposure to a Covid-19 infection.

Since successful acute toxicity tests have already been performed and since safe NHL epitopes are used, it is logical for the synthesis of Active Pharmaceutical Ingredients under Good Manufacturing Practice while optimizing the combination of antigen doses and adjuvants and preparing for manufacture at scale to be undertaken in parallel.

It is envisaged that Biovacc-19 will be ready for human clinical trials in the fourth quarter of 2020 although acceleration will be possible, given strategic funding to do so.

## Conclusion

We have offered a rationale for the design methodology and the necessary design parameters of a successful and safe vaccine against SARS-CoV-2. It is not included in any of the eight vaccine design routes identified in a recent *Nature* summary graphic. (Callaway, 2020) We have shown in this paper why a comprehensive analysis of the aetiology of the target virus is prerequisite, not optional. From the HIV experience, we have illustrated the risks of not so doing.

Next, we explained why, unlike in conventional vaccine design procedures, the choice of adjuvant is not to be seen as an afterthought but as integral from the beginning. We have deliberately chosen an adjuvant which has been shown to activate the innate and cell-mediated immune responses which are crucial to the successful presentation of the relevant epitopes. We have shown how Biovacc-19 has employed our understanding of the general method of action for infectivity and pathogenicity of the target virus to optimise action and to minimise risk, especially Antibody Dependent Enhancement; and we have presented the Non Human-Like epitopes in the SARS-CoV-2 Spike from which Biovacc-19 has been down-selected.

Oslo and London

28 May 2020



## Acknowledgements

The authors acknowledge invaluable assistance with analysis and critical comment by Dr J-F Moxnes of the Norwegian Defence Research Establishment, Kjeller, Norway.

## Author Contributions

Birger Sørensen: Writing original draft

Angus Dalglish: Review and editing

Andres Susrud: Investigation and Review and editing

## Financial support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

## Conflicts of interest declarations

Birger Sørensen is an employee and shareholder of Immunor AS.

Andres Susrud is an employee and shareholder of Immunor AS.

Angus Dalglish is on the scientific advisory board of Immodulon, and has stock options in Immunor AS.

## Figure captions

**Figure 1.** Alignments of Corona virus Spike protein inserts .

**Fig. 2.** The identified inserts examined in the PDB 6VXX electron microscopy structure (Walls *et al.*, 2020) The sequences highlighted in red could not be found in the cryo-electron microscopy structure data. The 6 aligned sequences in Fig. 1 are underlined in the missing sequences. Bold amino acids indicate first and last amino acids used to build the structure where the missing part is in between. Insert 6 did not have the same sequence in 6VXX as in the reference Sars-CoV-2 sequence. The authors stated that a designed mutated strain lacking the furin cleavage site residues was used.

**Fig. 3.** from (Yesylevskyy *et al.*,2009)

**Fig. 4.** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on human SARS-CoV shows a dominance of acidic amino acids giving a  $pI= 5.67$  (green dotted line). Repulsion will be observed in the presence of a protein/peptide domain with a similarly low  $pI=5.12$  as found for ELEC4M/DC-SIGN.  
Ref.: <https://www.uniprot.org/uniprot/P59594> sp|P59594|14-667 [SARS-CoV Urbani]

**Fig.5** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on SARS-CoV-2 shows a dominance of basic amino acids. The red circles are verified to be surface exposed while the blue circles are missing in PDB: 6VXX. Non-labelled peaks are not surface exposed. Above the green dotted line (Isoelectric point for the SARS-CoV-2 Spike protein S1  $pI=8.24$ ) the peaks contain basic residues which can form salt bridges in the presence of acidic amino acids such as in the receptor CLEC4M. The Cys131-Cys166 and Cys336-Cys361 loops are highlighted by as bold red/blue and red/red circles. Furthermore the Cys391-Cys525 bridge: a highly basic domain (526-560) with  $pI=10.03$  is moved forward and sits next to the Receptor Binding Motif (408-508) as part of a Cys538-Cys590 loop, as indicated by the green arrow. Ref.: <https://www.uniprot.org/uniprot/P0DTC2> sp|P0DTC2| 13-685 [SARS-CoV-2]

**Fig.6** Distribution of charge using a rolling window of 12 amino acids in steps of 1 on the receptor CLEC4M having  $pI=5.12$ . The ovals identify the most likely contributors to the salt bridges between the acidic domain, here on CLEC4M and the highlighted basic domains on the SARS-CoV-2 S1 protein in Fig.3 above. The two disulfide bonds Cys296-Cys389 and Cys368-Cys381 present in the C-type lectin C-terminal part of CLEC4M pull back the Cys368-Cys381 acidic domain to position 296 and make that domain a highly condensed with acidic amino acids ready for formation of salt-bridges with the basic amino acids in S1 of SARS-CoV-2 (Marzi *et al.*, 2004). Further investigations

might possibly show that other amino acids on the S1 SARS-CoV-2 are involved in such attachment receptor binding. Ref.: <https://www.uniprot.org/uniprot/Q9H2X3>

**Fig. 7** Spike trimer (a) top view and (b) side view. The specific receptor binding motif (RBM) is located on the sequence (437-508), while the receptor binding domain (RBD) has a broader location (319-541) ref. <https://www.uniprot.org/uniprot/P0DTC2>. The charged cysteine associated domains are Cys131-Cys166, Cys336-Cys361, Cys391-Cys525...Cys538-Cys590. As can be seen, there is a high concentration of positive charged surface exposed amino acids within the receptor domain next to the receptor binding motif.

## References

Åmand HL, Boström CL, Lincoln P, Nordén B, Esbjörner EK, (2011) "Binding of cell-penetrating penetratin peptides to plasma membrane vesicles correlates directly with cellular uptake", *Biochimica et Biophysica Acta* (BBA) – Biomembranes, <https://doi.org/10.1016/j.bbamem.2011.03.011>

Bourinbaiar AS, Batbold U, Efremenko Y, Sanjagdorj M, Butov D, Damdinpurev N, Grinishina E, Mijiddorj O, Kovolev M, Baasanjav K, Butova T, Prihoda N, Batbold O, Yurchenko L, Tseveendorj A, Arzhanova O, Chunt E, Stepanenko H, Sokolenko N, Makeeva N, Tarakanovskaya M, Borisova V, Reid A, Kalashnikov V, Nyasulu P, Prabowo SA, Jirathitikal V, Bain AI, Stanford C, Stanford J. (2020) "Phase III, placebo-controlled, randomized, double-blind trial of tableted, therapeutic TB vaccine (V7) containing heat-killed *M. vaccae* administered daily for one month". *J Clin Tuberc Other Mycobact Dis.* 2019 Dec 12;18:100141. doi: 10.1016/j.jctube.2019.100141. eCollection 2020 Feb., *J Clin Tuberc Other Mycobact Dis.* 2019.PMID: 31890902

Callaway E, (2020) "The Race for Coronavirus Vaccines", *Nature*, Vol 580, pp 576-577 (<https://www.nature.com/articles/d41586-020-01221-y>)

Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG & Decroly E, (2020) "The spike glycoprotein of the new coronavirus 2019-nCoV contains a furinlike cleavage site absent in CoV of the same clade", *Antiviral Research* 176, 104742, <https://doi.org/10.1016/j.antiviral.2020.104742>

Dalgleish AG, (2020) "Is Britain's coronavirus response bogged down in bureaucracy?" *The Spectator*, 28 April 2020, <https://www.spectator.co.uk/article/is-britain-s-coronavirus-response-bogged-down-in-bureaucracy#>

Dalgleish AG, Mudan S, Fusi A. (2018) "Enhanced effect of checkpoint inhibitors when given after or together with IMM-101: significant responses in four advanced melanoma patients with no additional major toxicity" *J Transl Med.*, Aug 14;16(1):227. doi: 10.1186/s12967-018-1602-8. *J Transl Med.* 2018.PMID: 30107850

Dalgleish AG, Stebbing J, Adamson DJ, Arif SS, Bidoli P, Chang D, Cheeseman S, Diaz-Beveridge R, Fernandez-Martos C, Glynne-Jones R, Granetto C, Massuti B, McAdam K, McDermott R, Martín AJ, Papamichael D, Pazo-Cid R, Vieitez JM, Zaniboni A, Carroll KJ, Wagle S, Gaya A, Mudan SS (2016) "Randomised, open-label, phase II study of gemcitabine with and without IMM-101 for advanced pancreatic cancer". *Br J Cancer*, 2016 Oct 25;115(9):e16. doi: 10.1038/bjc.2016.342. Epub Oct 11. *Br J Cancer.* 2016.PMID: 27727233

Duerr A, Huang Y, Buchbinder S, Coombs RW, Sanchez J, Del Rio C, Casapia M, Santiago S, Gilbert P, Corey L, Robertson MN, for the Step/HVTN 504 Study Team, (15 July 2012) "Extended Follow-up Confirms Early Vaccine-Enhanced Risk of HIV Acquisition and Demonstrates Waning Effect Over Time Among Participants in a Randomized Trial of Recombinant Adenovirus HIV Vaccine (Step Study)", *The Journal of Infectious Diseases*, Volume 206, Issue 2, Pages 258–266, <https://doi.org/10.1093/infdis/jis342>

Faries MB, Mozzillo N, Kashani-Sabet M, Thompson JF, Kelley MC, DeConti RC, Lee JE, Huth JF, Wagner J, Dalgleish A, Pertschuk D, Nardo C, Stern S, Elashoff R, Gammon G, Morton DL; MMAIT-IV (2017) "Clinical Trial Group. Long-Term Survival after Complete Surgical Resection and Adjuvant Immunotherapy for Distant Melanoma Metastases". *Ann Surg Oncol.* 2017 Dec;24(13):3991-4000. doi: 10.1245/s10434-017-6072-3. Epub 2017 Oct 10., *Ann Surg Oncol.* 2017.PMID: 29019177Clinical Trial.

Henry B. M., Santos de Oliveira M.H., Benoit S., Plebani M. and Lippi G. (2020) « Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in

coronavirus disease 2019 (COVID-19): a meta-analysis» *Clinical Chemistry and Laboratory Medicine (CCLM)* <https://doi.org/10.1515/cclm-2020-0369>

Huang Y, Pantaleo G, Tapia G, Sanchez B, Zhang L, Trondsen M, Hovden AO, Pollard R, Rockstroh J, Ökvist M, Sommerfelt MA, (2017) "Cell-Mediated Immune Predictors of Vaccine Effect on Viral Load and CD4 Count in a Phase 2 Therapeutic HIV-1 Vaccine Clinical Trial." *EBioMedicine*. Oct;24:195-204. doi: 10.1016/j.ebiom.2017.09.028. Epub 2017 Sep 22. <https://www.ncbi.nlm.nih.gov/pubmed/28970080>

Huang, Y, Zhang L, Jolliffe D, Hovden AO., Ökvist M, Pantaleo G, Sommerfelt MA, (2016) "A Case for preART-Adjusted Endpoints in HIV Therapeutic Vaccine Trials. *Vaccine* 34:1282-1288.<http://www.ncbi.nlm.nih.gov/pubmed/26826543>

Huang Y, Zhang L, Jolliffe D, Sanchez B, Stjernholm G, Jelmert Ø, Ökvist M, Sommerfelt MA "Postvaccination C-Reactive Protein and C5/gp41732-744 Antibody Level Fold-Changes Over Baseline Are Independent Predictors of Therapeutic HIV Vaccine Effect in a Phase 2 Clinical Study of Vacc-4x." (2018) *AIDS Res Hum Retroviruses*. Mar;34(3):307-313. doi: 10.1089/AID.2017.0179. Epub 2018 Feb 13. <https://www.ncbi.nlm.nih.gov/pubmed/29297230>

Kido H, Okumura Y, Takahashi E, Pan HY, Wang S, Yao D, Yao M, Chida J, Yano M, (2012) "Role of host cellular proteases in the pathogenesis of influenza and influenza-induced multiple organ failure", *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, <https://doi.org/10.1016/j.bbapap.2011.07.001>

Lam VC and Lanier LL. (2017) "NK cells in host responses to viral infections". *Curr Opin Immunol*. Feb; 44: 43–51.

Liu W, Li H (2020): "COVID-19:Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit Human Heme Metabolism". *ChemRxiv*. Preprint. <https://doi.org/10.26434/chemrxiv.11938173.v8>

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W, (2020) "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding". *Lancet*; 395:565-74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)

Marzi A, Gramberg T, Simmons G, Möller P, Rennekamp AJ, Krumbiegel M, Geier M, Eisemann J, Turza N, Saunier B, Steinkasserer A, Becker S, Bates P, Hofmann H, Pöhlmann S, (2004) "DC-SIGN and DC-SIGNR Interact with the Glycoprotein of Marburg Virus and the S Protein of Severe Acute Respiratory Syndrome Coronavirus", *Journal of Virology*, DOI: 10.1128/JVI.78.21.12090-12095.

Negro F, (2020) "Is antibody-dependent enhancement playing a role in COVID-19 pathogenesis?", *Swiss Medical Weekly*, <https://doi.org/10.4414/smw.2020.20249>

Pollard RB, Rockstroh JK, Pantaleo G, *et al.* (2014) "Safety and efficacy of the peptide-based therapeutic vaccine for HIV-1, Vacc-4x: a phase 2 randomised, double-blind, placebo-controlled trial". *Lancet Infect Dis*. published online Feb 11. [http://dx.doi.org/10.1016/S1473-3099\(13\)70343-8](http://dx.doi.org/10.1016/S1473-3099(13)70343-8)

Richard JP, Melikov K, Vives E, Ramos C, Verbeure B, Gait MJ, Chernomordik LV and Lebleu B, (2003) "Cell-penetrating Peptides: A re-evaluation of the mechanism of cellular uptake" *J.Biol.Chem.*, DOI 10.1074/jbc.M209548200

Ricke D and Malone RW, (2/27/2020) "Medical Countermeasures Analysis of 2019-nCoV and Vaccine Risks for Antibody-Dependent Enhancement (ADE)" Available at SSRN: <https://ssrn.com/abstract=3546070> or <http://dx.doi.org/10.2139/ssrn.3546070>

Roper SD, (2013) "Taste buds as peripheral chemosensory processors", *Semin Cell Dev Biol.*; 24(1): 71–79

Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F, (2020) "Structural Basis of Receptor Recognition by SARS-CoV-2" *May*; 581(7807): 221-224. doi: 10.1038/s41586-020-2179-y. Epub 2020 30 March 2020. <https://pubmed.ncbi.nlm.nih.gov/32225175/>

Stebbing J, Dalglish AG, Gifford-Moore A, Martin A, Gleeson C, Wilson G, Brunet LR, Grange J, Mudan S, Stebbing J, *et al.*, (2012) *Ann Oncol*, May;23(5):1314-9. doi: 10.1093/annonc/mdr363. Epub 2011 Sep 19, *Ann Oncol.*, 2012.PMID: 21930686Clinical Trial.

Thorén PEG, Persson D, Karlsson M, Nordén B, (2000) "The Antennapedia peptide penetratin translocates across lipid bilayers – the first direct observation", *FEBS letters*, [https://doi.org/10.1016/S0014-5793\(00\)02072-X](https://doi.org/10.1016/S0014-5793(00)02072-X)

UNAIDS, (February 2020). "HVTN 702 clinical trial of an HIV vaccine stopped"  
[https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2020/february/20200204\\_vaccine](https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2020/february/20200204_vaccine)

Uniprot, P0DTC2 - Spike glycoprotein, Human SARS coronavirus 2 (SARS-CoV-2)  
<https://www.uniprot.org/uniprot/P0DTC2>

Uniprot, P59594 – Spike glycoprotein, Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)

Uniprot, Q9BYF1 - Angiotensin-converting enzyme 2, <https://www.uniprot.org/uniprot/Q9BYF1>

Uniprot, Q9N2X3 - C-type lectin domain family 4 member M, <https://www.uniprot.org/uniprot/Q9N2X3>

Uniprot, Q9NNX6 – CD209, <https://www.uniprot.org/uniprot/Q9NNX6>

US patent - US9950811B2n

Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D, (2020) "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein", *Cell*, Volume 181, Issue 2, 16 April, Pages 281-292.e6, DOI: <https://doi.org/10.1016/j.cell.2020.02.058>

Wong SK, Li W, Moore MJ, Choe H and Farzan M, (2004) "A 193-Amino Acid Fragment of the SARS Coronavirus S Protein Efficiently Binds Angiotensin-converting Enzyme 2", *Journal of Biological Chemistry*, DOI: 10.1074/jbc.C300520200

Workman AD, Palmer JN, Adappa ND and Cohen NA, (2015) "The Role of Bitter and Sweet Taste Receptors in Upper Airway Immunity " *Curr Allergy Asthma Rep*. December; 15(12): 72. doi:10.1007/s11882-015-0571-8.

Yesylevskyy S, Marrink SJ and Mark AE, (2009) "Alternative Mechanisms for the interaction of the Cell-Penetrating Peptides Penetratin and the TAT Peptide with Lipid Bilayers", *Biophysical Journal*, Volume 97 July 2009, DOI: 10.1016/j.bpj.2009.03.059

Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, Zhu Y, Zhang YW, Xie QM, Mani S, Zheng XS, Li B, Li JM, Guo H, Pei GQ, An XP, Chen JW, Zhou L, Mai KJ, Wu ZX, Li D, Anderson DE, Zhang LB, Li SY, Mi ZQ, He TT, Cong F, Guo PJ, Huang R, Luo Y, Liu XL, Chen J, Huang Y, Sun Q, Zhang XLL, Wang YY, Xing SZ, Chen YS, Sun Y, Li J, Daszak P, Wang LF, Shi ZL, Tong YG & Ma JY, (2018) "Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin", *Nature* 556, 255–258, <https://doi.org/10.1038/s41586-018-0010-9>

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF & Shi ZL, (2020) "A pneumonia outbreak associated with a new coronavirus of probable bat origin", *Nature* 579, 270–273, <https://doi.org/10.1038/s41586-020-2012-7>

# Norwegian scientist Birger Sorensen claims coronavirus was lab-made and 'not natural in origin'



HOME » HEALTH RESEARCH

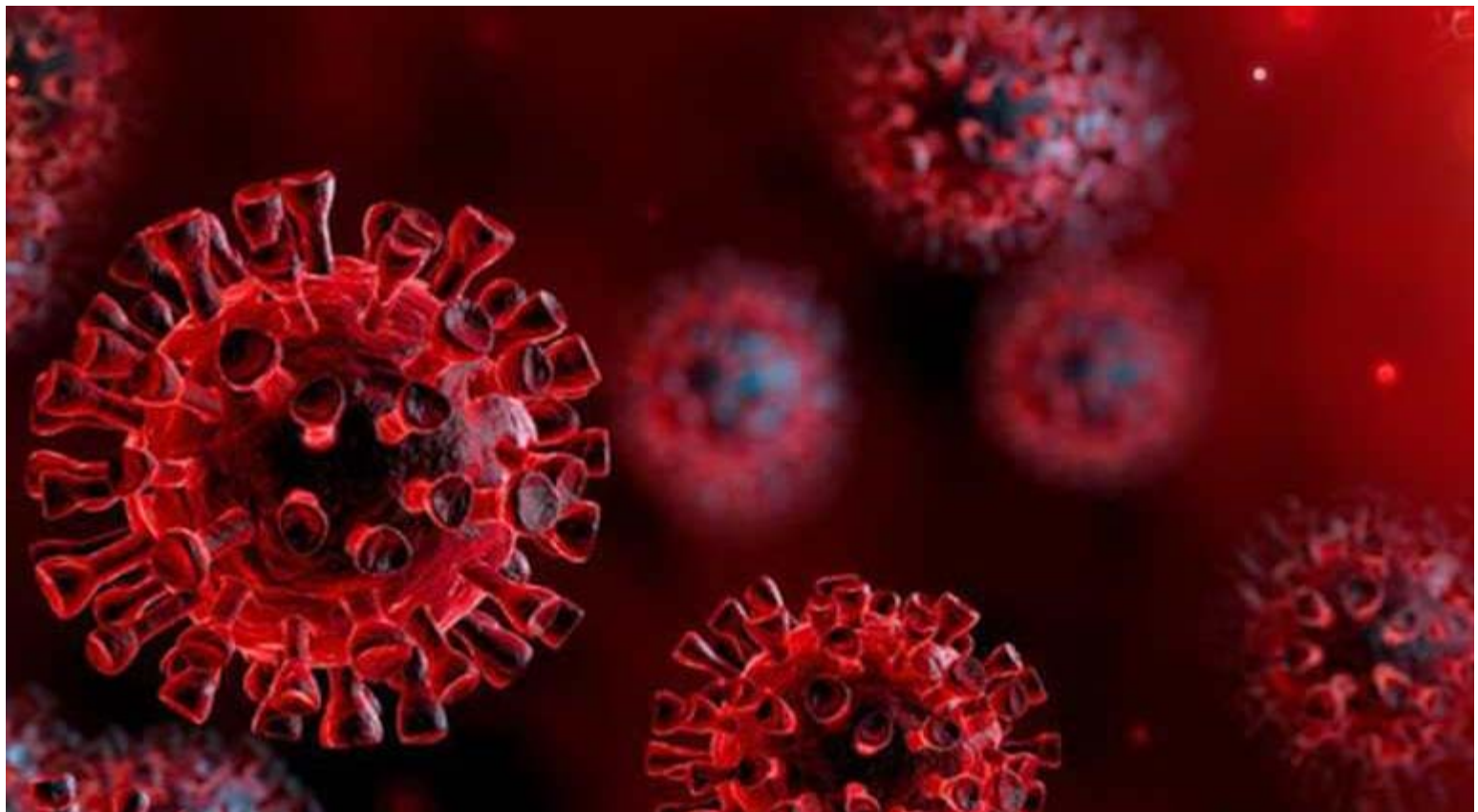
## Norwegian scientist Birger Sorensen claims coronavirus was lab-made and 'not natural in origin'



Nickie Louise POSTED ON JUNE 7, 2020

Share On Facebook

Tweet It



Just after we thought the question about the origin of coronavirus has been settled, a new study



from Norwegian virologist Birger Sørensen is now reigniting the debate about the possible origin of the deadly coronavirus. In a new peer-reviewed paper published together with Professor Angus Dalglish of St George’s Hospital at the University of London, Sorensen claimed the novel coronavirus SARS-CoV-2 is not natural in origin.

According to the study, which was [published](#) in the Quarterly Review of Biophysics, the authors found that the coronavirus’s spike protein contains sequences that appear to be artificially inserted. “The inserted sequences should never have been published. Had it been today, it would never have happened. It was a big mistake the Chinese made. The inserted sequences have a functionality that we describe. We explain why they are essential. But the Chinese pointed to them first,” Sørensen told the [NRK](#).

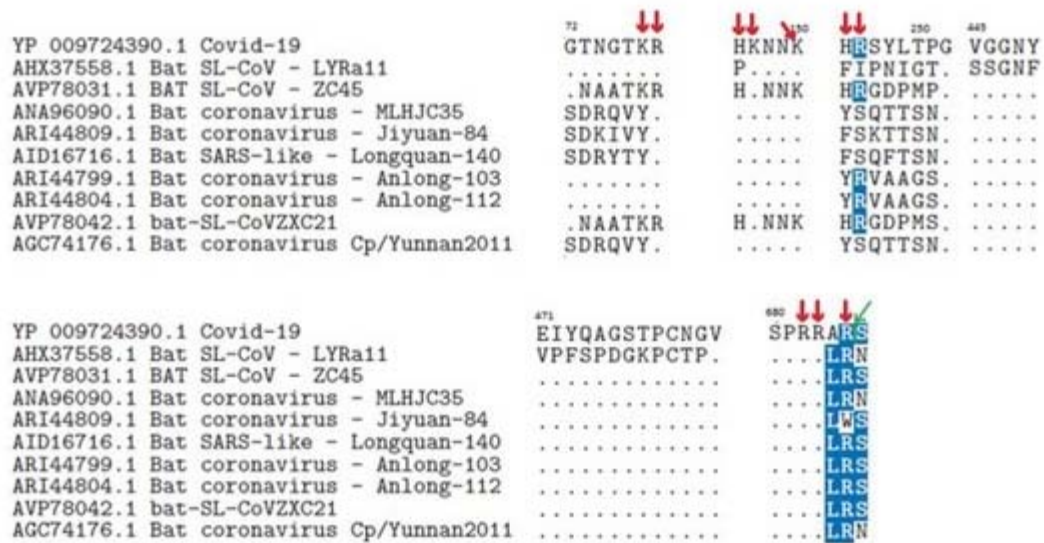
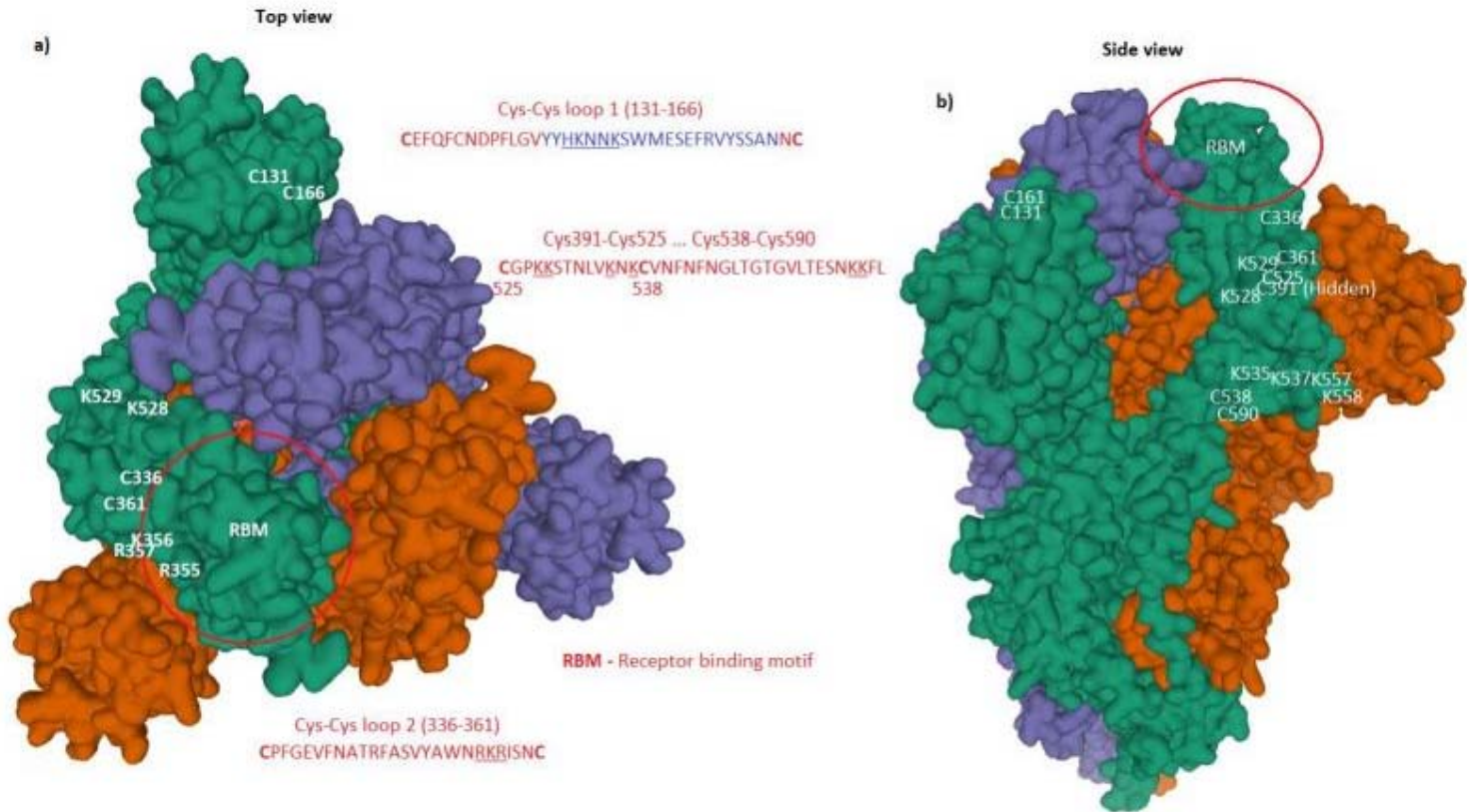


Figure 1. Alignments of Corona virus Spike protein inserts .

INSERTS: These are the sequences Sørensen thinks point to that the virus has no natural origin. The sequences were first described by Chinese scientists.

The eye opening claims also found that the virus had been doctored to bind to humans: “We are aware that these findings could have political significance and raise troubling questions.” The two researchers also pointed out that the virus has hardly mutated since it began to infect humans,

suggesting that it was already fully adapted to humans. According to Sørensen, this is quite unusual for viruses that cross species barriers. According to Sørensen, the virus has properties that differ greatly from SARS, and which have never been detected in nature.



*Spike trimer (a) top view and (b) side view. The specific receptor binding motif (RBM) is located on the sequence (437-508), while the receptor binding domain (RBD) has a broader location (319-541) ref. <https://www.uniprot.org/uniprot/P0DTC2>. The charged cysteine associated domain are Cys131-Cys166, Cys336- Cys361, Cys391-Cys525...Cys538-Cys590. As can be seen, there is a high concentration of positive charged surface exposed amino acids within the receptor domain next to the receptor binding motif.*

Sørensen said, “When we technically describe the virus, we see that it has not come about in a natural development. It’s done by Americans and Chinese, as part of what’s called “gain of function” studies. It is done all over the world. You say you don’t, but it happens all the time in advanced labs.”

“ *These data reveal the biological structure of SARS-CoV-2 Spike and confirm that accumulated charge from inserts and salt bridges are in surface positions capable of binding with cell membrane components other than the ACE2 receptor. We have also looked at the natural coronavirus spike protein as a concept for the basis of a vaccine, which we have rejected because of high risk of contamination with human-like epitopes. Analysis of the Spike protein SARS-CoV-2 shows 78.4% similarity with human-like (HL) epitopes. For the avoidance of confusion, a standard protein blast searches for functionalities and homologies to other proteins. However, antibodies can only recognize 5-6 amino acids and therefore a 6 amino acid window search for antibody epitopes was performed. A search so tailored to match against human known proteins will give a 78.4% human similarity to the SARS-CoV-2 Spike protein. All epitopes on the 1255 amino acid long SARS-CoV-2 Spike protein can be used by antibodies then there will be 983 antibody binding sites which also could bind to epitopes on human proteins. This is what we did and found.*

The authors also highlight the lack of mutation since its discovery, which suggests it was already fully adapted to humans. The study goes on to explain the rationale for the development of Biovacc-19, a candidate vaccine for COVID-19 that is now in advanced pre-clinical development.

The initial paper was rejected by well-known journals such as Nature and the Journal of Virology back in April before the manuscript of the paper was later accepted and published on Cambridge University Press website, according to a report from the Telegraph. The conclusions of the study were also dismissed by Imperial College London and the Francis Crick Institute.

Below is the abstract of the study.

“ *This study presents the background, rationale and Method of Action of Biovacc-19, a candidate vaccine for Covid-19, now in advanced pre-clinical development, which has already passed first acute toxicity testing. Unlike conventionally developed vaccines, Biovacc-19's Method of Action is upon non human-like (NHL) epitopes in 21.6% of the composition of SARS-CoV-2 Spike protein, which displays distinct distributed charge including the presence of a charge furin-like cleavage site. The logic of the design of the vaccine is explained, which starts with an empirical analysis of the aetiology of SARS-CoV-2. Mistaken assumptions about SARS-CoV-2*

*aetiology risk creating ineffective or actively harmful vaccines, including the risk of Antibody Dependent Enhancement (ADE).*

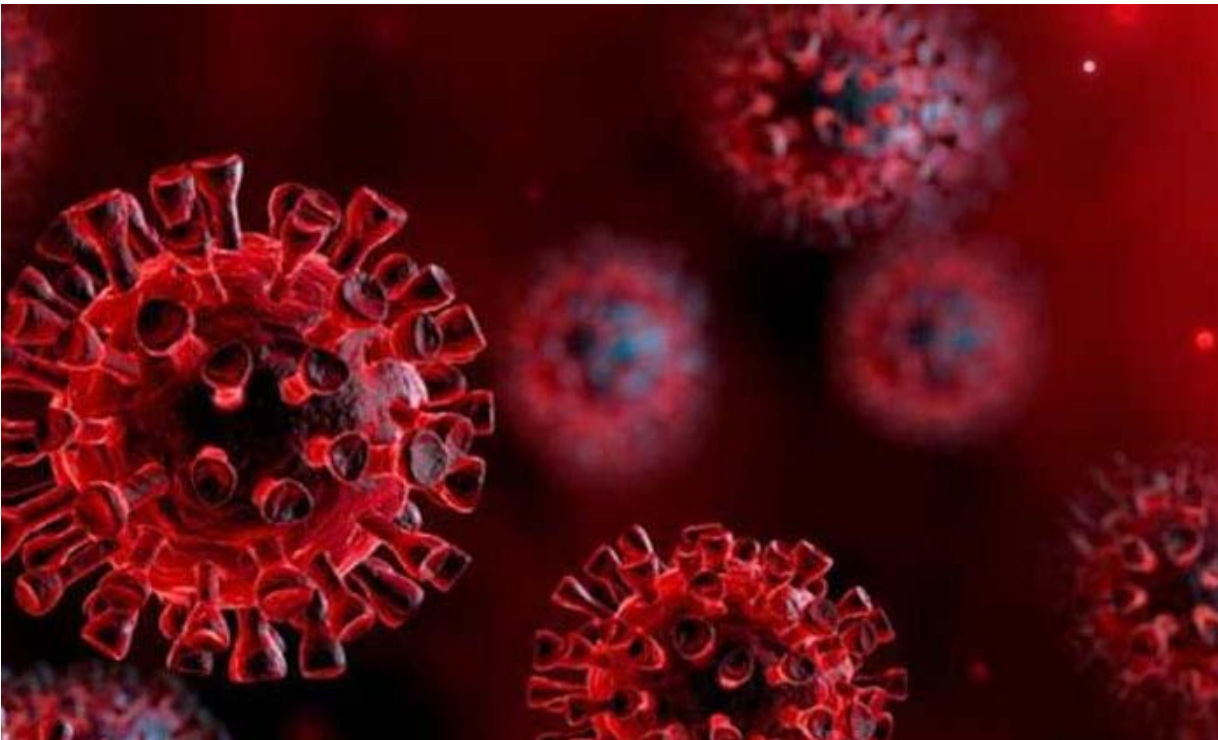
*Such problems in vaccine design are illustrated from past experience in the HIV domain propose that the dual effect general method of action of this chimeric virus's spike, including receptor binding domain, includes membrane components other than the ACE2 receptor, explains clinical evidence of its infectivity and pathogenicity. We show the non-receptor dependent phagocytic general method of action to be specifically related to cumulative charge from inserted sections placed on the SARS-CoV-2 Spike surface in positions to efficiently by salt bridge formations; and from blasting the Spike we display the non human epitopes from which Biovacc-19 has been down-selected.*

[Share On Facebook](#)

[Tweet It](#)



## TRENDING NOW





# Phylogenetic network analysis of SARS-CoV-2 genomes

Peter Forster<sup>a,b,c,1</sup>, Lucy Forster<sup>d</sup>, Colin Renfrew<sup>b,1</sup>, and Michael Forster<sup>c,e</sup> 

<sup>a</sup>Institute of Forensic Genetics, 48161 Münster, Germany; <sup>b</sup>McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom; <sup>c</sup>Fluxus Technology Limited, Colchester CO3 0NU, United Kingdom; <sup>d</sup>Lakeside Healthcare Group at Cedar House Surgery, St Neots PE19 1BQ, United Kingdom; and <sup>e</sup>Institute of Clinical Molecular Biology, Christian-Albrecht-University of Kiel, 24105 Kiel, Germany

Contributed by Colin Renfrew, March 30, 2020 (sent for review March 17, 2020; reviewed by Toomas Kivisild and Carol Stocking)

**In a phylogenetic network analysis of 160 complete human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes, we find three central variants distinguished by amino acid changes, which we have named A, B, and C, with A being the ancestral type according to the bat outgroup coronavirus. The A and C types are found in significant proportions outside East Asia, that is, in Europeans and Americans. In contrast, the B type is the most common type in East Asia, and its ancestral genome appears not to have spread outside East Asia without first mutating into derived B types, pointing to founder effects or immunological or environmental resistance against this type outside Asia. The network faithfully traces routes of infections for documented coronavirus disease 2019 (COVID-19) cases, indicating that phylogenetic networks can likewise be successfully used to help trace undocumented COVID-19 infection sources, which can then be quarantined to prevent recurrent spread of the disease worldwide.**

SARS-CoV-2 evolution | subtype | ancestral type

The search for human origins seemed to take a step forward with the publication of the global human mitochondrial DNA tree (1). It soon turned out, however, that the tree-building method did not facilitate an unambiguous interpretation of the data. This motivated the development, in the early 1990s, of phylogenetic network methods which are capable of enabling the visualization of a multitude of optimal trees (2, 3). This network approach, based on mitochondrial and Y chromosomal data, allowed us to reconstruct the prehistoric population movements which colonized the planet (4, 5). The phylogenetic network approach from 2003 onward then found application in the reconstruction of language prehistory (6). It is now timely to apply the phylogenetic network approach to virological data to explore how this method can contribute to an understanding of coronavirus evolution.

In early March 2020, the GISAID database (<https://www.gisaid.org/>) contained a compilation of 253 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) complete and partial genomes contributed by clinicians and researchers from across the world since December 2019. To understand the evolution of this virus within humans, and to assist in tracing infection pathways and designing preventive strategies, we here present a phylogenetic network of 160 largely complete SARS-CoV-2 genomes (Fig. 1).

Zhou et al. (7) recently reported a closely related bat coronavirus, with 96.2% sequence similarity to the human virus. We use this bat virus as an outgroup, resulting in the root of the network being placed in a cluster of lineages which we have labeled “A.” Overall, the network, as expected in an ongoing outbreak, shows ancestral viral genomes existing alongside their newly mutated daughter genomes.

There are two subclusters of A which are distinguished by the synonymous mutation T29095C. In the T-allele subcluster, four Chinese individuals (from the southern coastal Chinese province of Guangdong) carry the ancestral genome, while three Japanese and two American patients differ from it by a number of mutations. These American patients are reported to have had a history of residence in the presumed source of the outbreak in Wuhan. The C-allele subcluster sports relatively long mutational branches

and includes five individuals from Wuhan, two of which are represented in the ancestral node, and eight other East Asians from China and adjacent countries. It is noteworthy that nearly half (15/33) of the types in this subcluster, however, are found outside East Asia, mainly in the United States and Australia.

Two derived network nodes are striking in terms of the number of individuals included in the nodal type and in mutational branches radiating from these nodes. We have labeled these phylogenetic clusters B and C.

For type B, all but 19 of the 93 type B genomes were sampled in Wuhan ( $n = 22$ ), in other parts of eastern China ( $n = 31$ ), and, sporadically, in adjacent Asian countries ( $n = 21$ ). Outside of East Asia, 10 B-types were found in viral genomes from the United States and Canada, one in Mexico, four in France, two in Germany, and one each in Italy and Australia. Node B is derived from A by two mutations: the synonymous mutation T8782C and the nonsynonymous mutation C28144T changing a leucine to a serine. Cluster B is striking with regard to mutational branch lengths: While the ancestral B type is monopolized (26/26 genomes) by East Asians, every single (19/19) B-type genome outside of Asia has evolved mutations. This phenomenon does not appear to be due to the month-long time lag and concomitant mutation rate acting on the viral genome before it spread outside of China (Dataset S1, Supplementary Table 2). A complex founder scenario is one possibility, and a different explanation worth considering is that the ancestral Wuhan B-type virus is immunologically or environmentally adapted to a large section of the East Asian population, and may need to mutate to overcome resistance outside East Asia.

## Significance

This is a phylogenetic network of SARS-CoV-2 genomes sampled from across the world. These genomes are closely related and under evolutionary selection in their human hosts, sometimes with parallel evolution events, that is, the same virus mutation emerges in two different human hosts. This makes character-based phylogenetic networks the method of choice for reconstructing their evolutionary paths and their ancestral genome in the human host. The network method has been used in around 10,000 phylogenetic studies of diverse organisms, and is mostly known for reconstructing the prehistoric population movements of humans and for ecological studies, but is less commonly employed in the field of virology.

Author contributions: P.F. and M.F. performed research; P.F., L.F., and M.F. analyzed data; P.F. and M.F. performed statistical analyses; P.F., C.R., and M.F. wrote the paper; and C.R. wrote the Introduction.

Reviewers: T.K., Katholieke Universiteit Leuven; and C.S., University Medical Center Hamburg-Eppendorf.

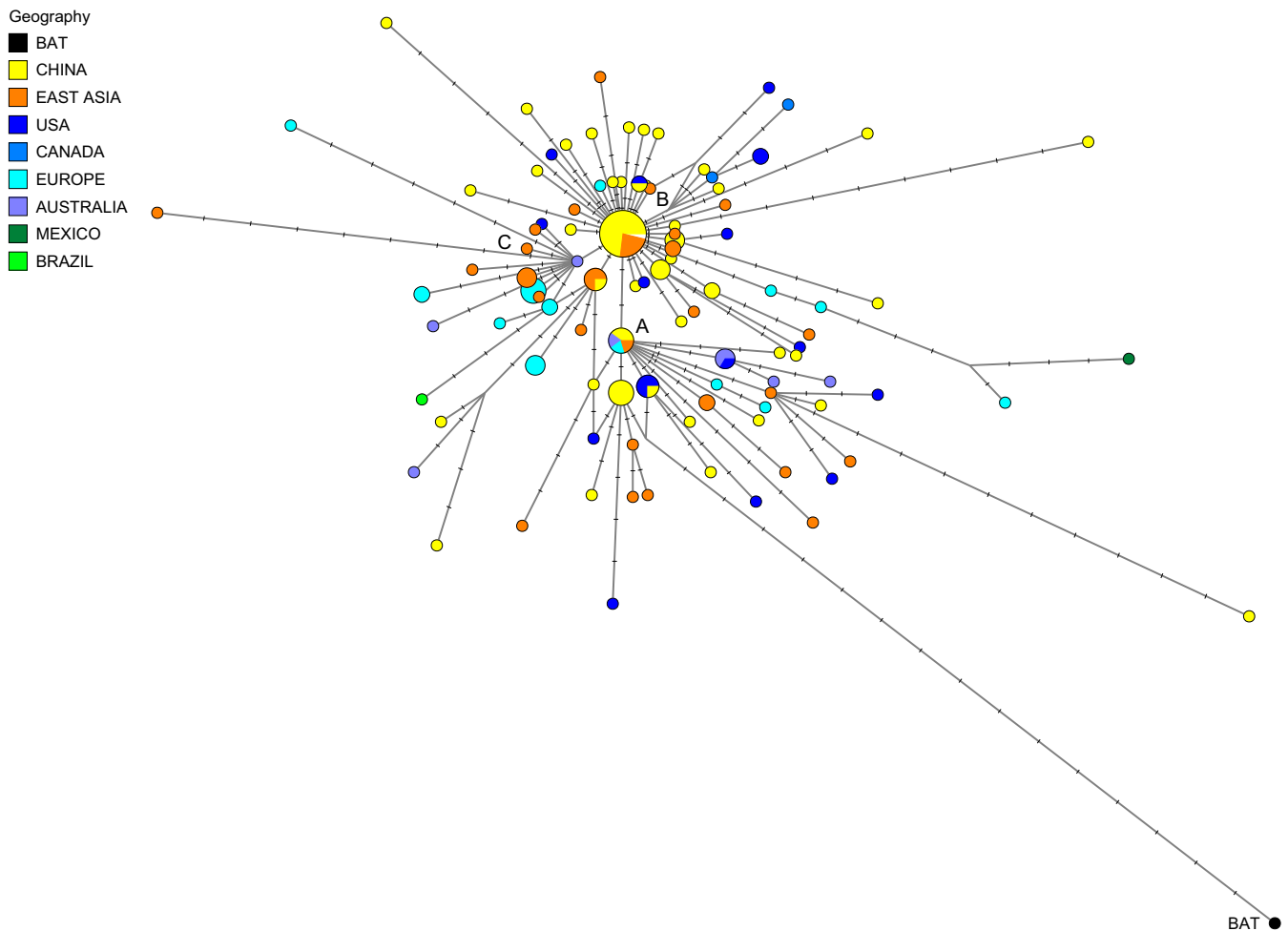
The authors declare no competing interest.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: pf223@cam.ac.uk or acr10@cam.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2004999117/-DCSupplemental>.





**Fig. 1.** Phylogenetic network of 160 SARS-CoV-2 genomes. Node A is the root cluster obtained with the bat (*R. affinis*) coronavirus isolate BatCoVraTG13 from Yunnan Province. Circle areas are proportional to the number of taxa, and each notch on the links represents a mutated nucleotide position. The sequence range under consideration is 56 to 29,797, with nucleotide position (np) numbering according to the Wuhan 1 reference sequence (8). The median-joining network algorithm (2) and the Steiner algorithm (9) were used, both implemented in the software package Network5011CS (<https://www.fluxus-engineering.com/>), with the parameter epsilon set to zero, generating this network containing 288 most-parsimonious trees of length 229 mutations. The reticulations are mainly caused by recurrent mutations at np11083. The 161 taxa (160 human viruses and one bat virus) yield 101 distinct genomic sequences. The phylogenetic diagram is available for detailed scrutiny in A0 poster format (*SI Appendix, Fig. S5*) and in the free Network download files.

Type C differs from its parent type B by the nonsynonymous mutation G26144T which changes a glycine to a valine. In the dataset, this is the major European type ( $n = 11$ ), with representatives in France, Italy, Sweden, and England, and in California and Brazil. It is absent in the mainland Chinese sample, but evident in Singapore ( $n = 5$ ) and also found in Hong Kong, Taiwan, and South Korea.

One practical application of the phylogenetic network is to reconstruct infection paths where they are unknown and pose a public health risk. The following cases where the infection history is well documented may serve as illustrations (*SI Appendix*). On 25 February 2020, the first Brazilian was reported to have been infected following a visit to Italy, and the network algorithm reflects this with a mutational link between an Italian and his Brazilian viral genome in cluster C (*SI Appendix, Fig. S1*). In another case, a man from Ontario had traveled from Wuhan in central China to Guangdong in southern China and then returned to Canada, where he fell ill and was conclusively diagnosed with coronavirus disease 2019 (COVID-19) on 27 January 2020. In the phylogenetic network (*SI Appendix, Fig. S2*), his virus genome branches from a reconstructed ancestral node, with derived virus variants in Foshan and Shenzhen (both in Guangdong

province), in agreement with his travel history. His virus genome now coexists with those of other infected North Americans (one Canadian and two Californians) who evidently share a common viral genealogy. The case of the single Mexican viral genome in the network is a documented infection diagnosed on 28 February 2020 in a Mexican traveler to Italy. Not only does the network confirm the Italian origin of the Mexican virus (*SI Appendix, Fig. S3*), but it also implies that this Italian virus derives from the first documented German infection on 27 January 2020 in an employee working for the Webasto company in Munich, who, in turn, had contracted the infection from a Chinese colleague in Shanghai who had received a visit by her parents from Wuhan. This viral journey from Wuhan to Mexico, lasting a month, is documented by 10 mutations in the phylogenetic network.

This viral network is a snapshot of the early stages of an epidemic before the phylogeny becomes obscured by subsequent migration and mutation. The question may be asked whether the rooting of the viral evolution can be achieved at this early stage by using the oldest available sampled genome as a root. As *SI Appendix, Fig. S4* shows, however, the first virus genome that was sampled on 24 December 2019 already is distant from the root type according to the bat coronavirus outgroup rooting.

The described core mutations have been confirmed by a variety of contributing laboratories and sequencing platforms and can be considered reliable. The phylogeographic patterns in the network are potentially affected by distinctive migratory histories, founder events, and sample size. Nevertheless, it would be prudent to consider the possibility that mutational variants might modulate the clinical presentation and spread of the disease. The phylogenetic classification provided here may be used to rule out or confirm such effects when evaluating clinical and epidemiological outcomes of SARS-CoV-2 infection, and when designing treatment and, eventually, vaccines.

## Materials and Methods

The Global Initiative on Sharing Avian Influenza Data (GISAID) was founded in 2006, and, since 2010, has been hosted by the German Federal Ministry of Food, Agriculture and Consumer Protection. GISAID has also become a coronavirus repository since December 2019. As of 4 March 2020, the cutoff point for our phylogenetic analysis, the GISAID database (<https://www.gisaid.org/>) had compiled 254 coronavirus genomes, isolated from 244 humans, nine Chinese pangolins, and one bat *Rhinolophus affinis* (BatCoVraTG13 from Yunnan Province, China). The sequences have been deposited by 82 laboratories listed in [Dataset S1, Supplementary Table 1](#). Although SARS-CoV-2 is an RNA virus, the deposited sequences, by convention, are in DNA format. Our initial alignment confirmed an earlier report by Zhou et al. (7) that the pangolin coronavirus sequences are poorly conserved with respect to the human SARS-CoV-2 virus, while the bat coronavirus yielded a sequence similarity of 96.2% in our analysis, in agreement with the 96.2% published by Zhou et al. We discarded partial sequences, and used only the most complete genomes that we aligned to the full reference genome by Wu et al. (8) comprising 29,903 nucleotides. Finally, to ensure comparability, we truncated the flanks of all sequences to the consensus range 56 to 29,797, with nucleotide position numbering according to the Wuhan 1 reference sequence (8). The laboratory

codes of the resulting 160 sequences and the bat coronavirus sequences are listed in [Dataset S1, Supplementary Table 2](#) (Coronavirus Isolate Labels).

The 160 human coronavirus sequences comprised exactly 100 different types. We added to the data the bat coronavirus as an outgroup to determine the root within the phylogeny. Phylogenetic network analyses were performed with the Network 5.011CS package, which includes, among other algorithms, the median joining network algorithm (3) and a Steiner tree algorithm to identify most-parsimonious trees within complex networks (9). We coded gaps of adjacent nucleotides as single deletion events (these deletions being rare, up to 24 nucleotides long, and mostly in the amino acid reading frame) and ran the data with the epsilon parameter set to zero, and performed an exploratory run by setting the epsilon parameter to 10. Both settings yielded a low-complexity network. The Steiner tree algorithm was then run on both networks and provided the identical result that the most-parsimonious trees within the network were of length 229 mutations. The structures of both networks were very similar, with the epsilon 10 setting providing an additional rectangle between the A and B clusters. The network output was annotated using the Network Publisher option to indicate geographic regions, sample collection times, and cluster nomenclature.

**Data Availability.** The nucleotide sequences of the SARS-CoV-2 genomes used in this analysis are available, upon free registration, from the GISAID database (<https://www.gisaid.org/>). The Network5011 software package and coronavirus network files are available as shareware on the Fluxus Technology website (<https://www.fluxus-engineering.com/>).

**ACKNOWLEDGMENTS.** We gratefully acknowledge the authors and originating and submitting laboratories of the sequences from GISAID's EpiFlu(TM) Database on which this research is based. We are grateful to Trevor Bedford (GISAID) for providing instructions and advice on the database. A table of the contributors is available in [Dataset S1, Supplementary Table 1](#). We thank Arne Röhl for assessing the network.

1. R. L. Cann, M. Stoneking, A. C. Wilson, Mitochondrial DNA and human evolution. *Nature* **325**, 31–36 (1987).
2. H. J. Bandelt, P. Forster, B. C. Sykes, M. B. Richards, Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753 (1995).
3. H. J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
4. P. Forster, C. Renfrew, Evolution. Mother tongue and Y chromosomes. *Science* **333**, 1390–1391 (2011).
5. C. Renfrew, P. Bahn, *The Cambridge World Prehistory* (Cambridge University Press, 2014).
6. P. Forster, C. Renfrew, *Phylogenetic Methods and the Prehistory of Languages* (McDonald Institute Press, 2006).
7. P. Zhou et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
8. F. Wu et al., A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
9. T. Polzin, S. V. Daneshmand, On Steiner trees and minimum-spanning trees in hypergraphs. *Oper. Res. Lett.* **31**, 12–20 (2003).