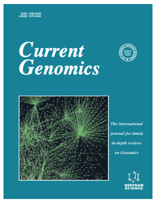
Current Genomics, 2021, 22, 541-549

RESEARCH ARTICLE



Identification of Conserved Epitopes in SARS-CoV-2 Spike and Nucleocapsid Protein

Sergio Forcelloni[1], Anna Benedetti[2,3], Maddalena Dilucca[2,*] and Andrea Giansanti[2,4]

[1]Mechanisms of Protein Biogenesis, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; [2]Sapienza University of Rome, Department of Physics, P.le A. Moro 5, 00185 Rome, Italy; [3]Sapienza University of Rome, Department AHFMO, Via A. Scarpa 14, 00161, Rome, Italy; [4]Istituto Nazionale di Fisica Nucleare, INFN, Roma1 Section 00185, Rome, Italy

Abstract: ***Background***: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel virus that first occurred in Wuhan in December 2019. The spike glycoproteins and nucleocapsid proteins are the most common targets for the development of vaccines and antiviral drugs.

Objective: We herein analyze the rate of evolution along with the sequences of spike and nucleocapsid proteins in relation to the spatial locations of their epitopes, previously suggested to contribute to the immune response caused by SARS-CoV-2 infections.

Methods: We compare homologous proteins of seven human coronaviruses: HCoV-229E, HCoV--OC43, SARS-CoV, HCoV-NL63, HCoV-HKU1, MERS-CoV, and SARS-CoV-2. We then focus on the local, structural order-disorder propensity of the protein regions where the SARS-CoV-2 epitopes are located.

Results: We show that most of nucleocapsid protein epitopes overlap the RNA-binding and dimerization domains, and some of them are characterized by a low rate of evolutions. Similarly, spike protein epitopes are preferentially located in regions that are predicted to be ordered and well-conserved, in correspondence of the heptad repeats 1 and 2. Interestingly, both the receptor-binding motif to ACE2 and the fusion peptide of spike protein are characterized by a high rate of evolution.

Conclusion: Our results provide evidence for conserved epitopes that might help develop broad--spectrum SARS-CoV-2 vaccines.

ARTICLE HISTORY

Received: August 13, 2021
Revised: November 15, 2021
Accepted: November 15, 2021

*DOI:
10.2174/1389202923666211216162605*

Keywords: SARS-CoV-2, spike glycoprotein, nucleocapsid protein, conserved epitopes, conservation score, order-disorder propensity.

1. INTRODUCTION

Coronaviruses are a large family of viruses that may cause illness in animals. In humans, seven coronaviruses are known to cause respiratory infections, ranging from the common cold to more severe diseases. The first two coronaviruses, human CoV-229E (HCoV-229E) and human CoV-OC43 (HCoV-OC43) were discovered in the 1960s, and cause relatively mild respiratory symptoms [1]. Human severe acute respiratory syndrome coronavirus (SARSr-CoV) was identified in 2003, and causes flu-like symptoms and atypical pneumonia in the worst cases [2]. The human coronavirus NL63 (HCoV-NL63), identified in 2004, and the human CoV-HKU1 (HCoV-HKU1), described in 2005 [3], generally cause upper respiratory disease in humans, which may progress in lower respiratory infections [4]. More recently, the pathogenic Middle East respiratory syndrome (MERS-CoV) coronavirus, which appeared for the first time in 2012, was identified as the sixth human coronavirus [5]. Finally, a previously unknown coronavirus probably originated in baths, SARS-CoV-2, was identified in December 2019 in Wuhan, China [6, 7]. SARS-CoV-2 caused an ongoing pandemic of severe pneumonia named coronavirus disease 19 (COVID-19), which has affected over 4 million people worldwide and caused more than 300.000 deaths as May 13, 2020 (https://ourworldindata.org/grapher/total-deaths-covid-19).

Currently, six COVID-19 vaccines have been approved by the World Health Organization: two RNA vaccines (Pfizer-BioNTech and Moderna) and four conventional attenuated/inactivated vaccines (Oxford-AstraZeneca, Johnson & Johnson, Sinovac, Sinopharm-BBIBP) [8, 9]. In this context, the viral spike (S) glycoprotein and the nucleocapsid (N) protein are two of the main targets for antibody production and the development of vaccines and antiviral drugs [9], due to their ability to trigger a dominant and long-lasting immune response.

*Address correspondence to this author at the Sapienza University of Rome, Department of Physics, P.le A. Moro 5, 00185, Rome, Italy;
Tel: +393475407737; E-mail: maddalena.dilucca@gmail.com, maddyemario@hotmail.it

1875-5488/21 $65.00+.00 © 2021 Bentham Science Publishers

The spike protein is a large type I transmembrane protein composed of approximately 1400 amino acids. S protein is an attractive target for vaccine development, as its surface expression renders it a direct target for the host immune response [10, 11]. Spike proteins assemble into trimers on the virion surface to form the distinctive crown-like structure and mediate the contact with the host cell by binding to ACE2 receptor, a process necessary for the virus entry. Spike protein contains 2 subunits: S1 N-terminal domain, responsible for ACE2 receptor binding, and S2 C-terminal domain, responsible for the fusion. The S2 subunit is the most conserved one, while the S1 subunit differs even within species of the same coronaviruses. The S1 contains two sub-domains (N-terminal and C-terminal) with receptor-binding functions. The S2 domain contains two heptad repeats composed by hydrophobic residues, responsible for the formation of an α-helical coiled-coil structure that participate in the virus-host cell membrane fusion [11].

The nucleocapsid protein regulates the viral genome transcription, replication and packaging, and it is essential for viability [12]. It contains two structural domains: the N-terminal domain, which acts as a putative RNA-binding domain, and a C-terminal domain, which acts as a dimerization domain. The N protein is of potential interest for vaccine development as it is highly immunogenic and its amino acid sequence is highly conserved [13, 14].

T cell responses against S and N proteins have been shown to be the most immunogenic and long-lasting in SARS-CoV patients. Furthermore, B-cell antibody response against S and N proteins was also reported to be effective, although short-lived compared to the T cell-response. The search of T-cell and B-cell epitopes, which can stimulate a specific immune response against S and N proteins, represents a valuable strategy to identify targets for the development of a SARS-CoV-2 vaccine [15].

In a previous study, we showed that genes encoding N and S proteins tend to evolve faster than genes encoding matrix and envelope proteins [1]. This result suggested that the higher divergence observed for these two genes could represent a significant barrier in the development of antiviral therapeutics against SARS-CoV-2. Here, we perform an accurate analysis of the position-specific rates of evolution and the order-disorder propensities of the spike glycoprotein (S) and the nucleocapsid protein (N) of SARS-CoV-2. We thus provide an *in-silico* survey of the major nucleocapsid protein and spike protein epitopes, identifying a subset of them that are well-conserved among human coronaviruses and represent reliable candidates for broad-spectrum vaccines against SARS-CoV-2.

2. RESULTS

2.1. Identification of Conserved Epitopes in Spike and Nucleocapsid Proteins

The amino acid sequences of the nucleocapsid (N) protein and spike (S) glycoprotein from the seven human coronaviruses here considered (HCoV-229E, HCoV-OC43, SARS-CoV, HCoV-NL63, HCoV-HKU1, MERS-CoV, and SARS-CoV-2.) were compared to assess the position-specific rates of evolution of these two proteins. We then investigated the relationships between the position-specific rates of evolution and the distribution of the epitopes that have previously been suggested to contribute to the immune response caused by human SARS-CoV-2 infections. For this purpose, we considered the SARS-CoV B cell and T cell linear epitopes that map identically to SARS-CoV-2 N and S proteins, as identified by Ahmed *et al.* [15]. It is worth noting that these epitopes were also considered in another study aiming to provide a molecular structural rationale of the major nucleocapsid protein epitopes for a potential role in conferring protection from SARS-CoV-2 infection [15]. Moreover, we also considered a high-quality set of previously identified linear epitopes directly extrapolated for the SARS-CoV-2, looking at the literature (see Materials and Methods). We herein focused on the SARS-CoV-2 N and S proteins because both these two proteins are the main targets of vaccines and antiviral drugs due to their dominant and long-lasting immune response previously reported against SARS-CoV [9, 15]. We aligned the homologous protein sequences in the seven human coronaviruses and used the resulting alignment to calculate a conservation profile by using the software Rate4site (see Materials and Methods) [16]. In Fig. (**1**), we report the profile obtained for the protein N, together with the functional regions/domains and the SARS-CoV-2 linear epitopes.

In this profile, values greater or less than zero reflect a faster or a slower evolution, respectively. We note that both the RNA-binding domain (region 41-186) and the dimerization domain (region 258-361) correspond to regions with values less than zero, implying higher conservation than the rest of the protein sequence. Although the general trend of the conservation profile hovers around 0, there are some regions where the conservation score is greater than one, indicating that these regions are likely to be under positive selection. Interestingly, the location of some of these regions corresponds to the presence of B cell and T cell epitopes (see Fig. **S1** for the specific location and the sequences of each epitope). We found that the vast majority of the epitopes in common between SARS-CoV and SARS-CoV-2 are positioned in high-variable regions. We suppose that the high amino acid variability of these regions might allow the virus to evade the host immune system recognition. At the same time, we suggest that the source of variability in these protein regions is likely to be the host immune response. However, we also observed the presence of some epitopes in highly conserved regions spanning residues 70-120, 150-200 and 250-315, which may potentially offer long-lasting protection against SARS-CoV-2.

Similarly, we study the conservation profile of the protein S and the distribution of the associated linear epitopes (Fig. **2**).

On the bottom of (Fig. **2**), we report the functional regions of protein S. The receptor binding domain (region 319-541) contains the receptor-binding motif to the human angiotensin-converting enzyme 2 (ACE2), an enzyme attached to the outer surface (cell membrane) of cells in lungs, arteries, heart, kidney, and intestines, and it has been identified as the functional receptor for SARS-CoV-2 [17]. As a transmembrane protein, ACE2 serves as the main entry point into cells for some coronaviruses, including HCoV-NL63, SARS-CoV, and SARS-CoV-2 [18-23]. More specifically, the binding of the spike S protein of SARS-CoV and SARS-CoV-2 to the enzymatic domain of ACE2 on the surface of cells results in endocytosis and translocation of both the virus and the enzyme into endosomes located within cells [24, 25]. Interestingly, both the receptor-binding motif to ACE2 (region 437-508) and the fusion peptide (amino acids 788-806IYKTPPIKDFGGFNFSQIL for SARS-CoV--2), the segment of the fusion protein that inserts to a target lipid bilayer and triggers virus-cell membrane fusion, are characterized by high rate of evolution. Conversely, heptad repeats 1 and 2, which are known to play a crucial role in membrane fusion and viral entry [25], show lower rates of evolution.

Moreover, we note that a large percentage of both B cell and T cell epitopes are located in the C-terminal region of spike protein in correspondence of the heptad repeats 1 and 2 in the S2 domain (see Fig. **S2** for details about the sequences of epitopes and their location along the sequence). Thus, at variance with the protein N, we note that spike protein epitopes are mainly located in the protein regions that are characterized by a lower rate of evolution. This observation suggests that the immune system has adapted to recognize slowly evolving regions of the S protein [26].

Finally, we observed that SARS-CoV derived B cell and T cell epitopes that map identically to SARS-CoV-2 proteins are more localized around functional sites than the epitopes directly extrapolated for SARS-CoV-2 proteins N and S, which are more scattered throughout the protein sequence. This observation suggests an adaptation of the immune system to recognize functional regions of proteins N and S in SARS-CoV and SARS-CoV-2 and potentially induce a long-lasting immunity against coronaviruses.

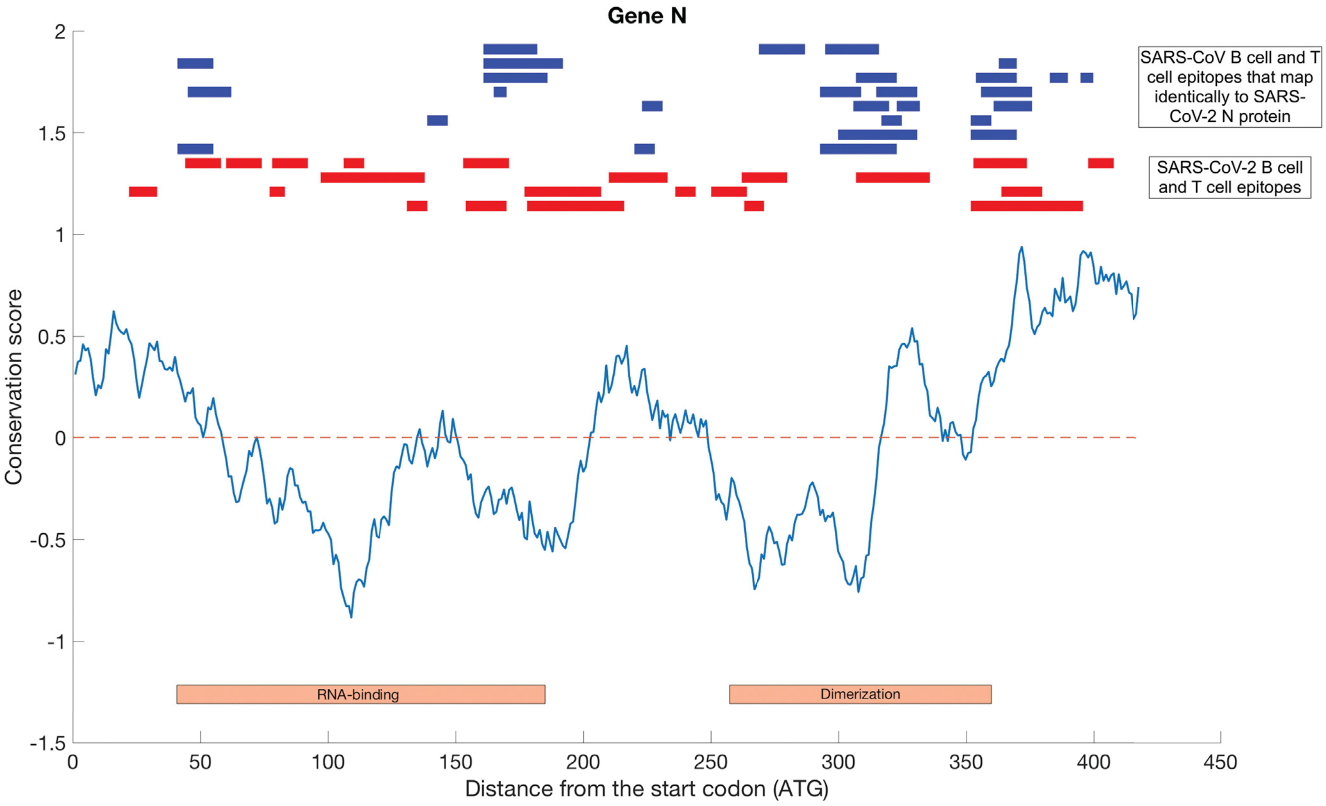


Fig. (1). Conservation profile of SARS-CoV-2 protein N. The solid blue line represents the position-specific estimations of the rate of evolution of each residue in the protein sequence as a function of the distance from the start codon. The horizontal dotted line represents the threshold value, above which the score is characteristic of highly variable regions (0 for Rate4site). On the bottom, two orange rectangles show the RNA-binding domain and the dimerization domain. On the top, the SARS-CoV derived B cell and T cell epitopes map identically to SARS-CoV-2 N protein (in blue); on the bottom, the epitopes directly extrapolate for SARS-CoV-2 (in red). Each horizontal bar represents a linear epitope, consisting of continuous residues on the protein sequence of the protein N. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

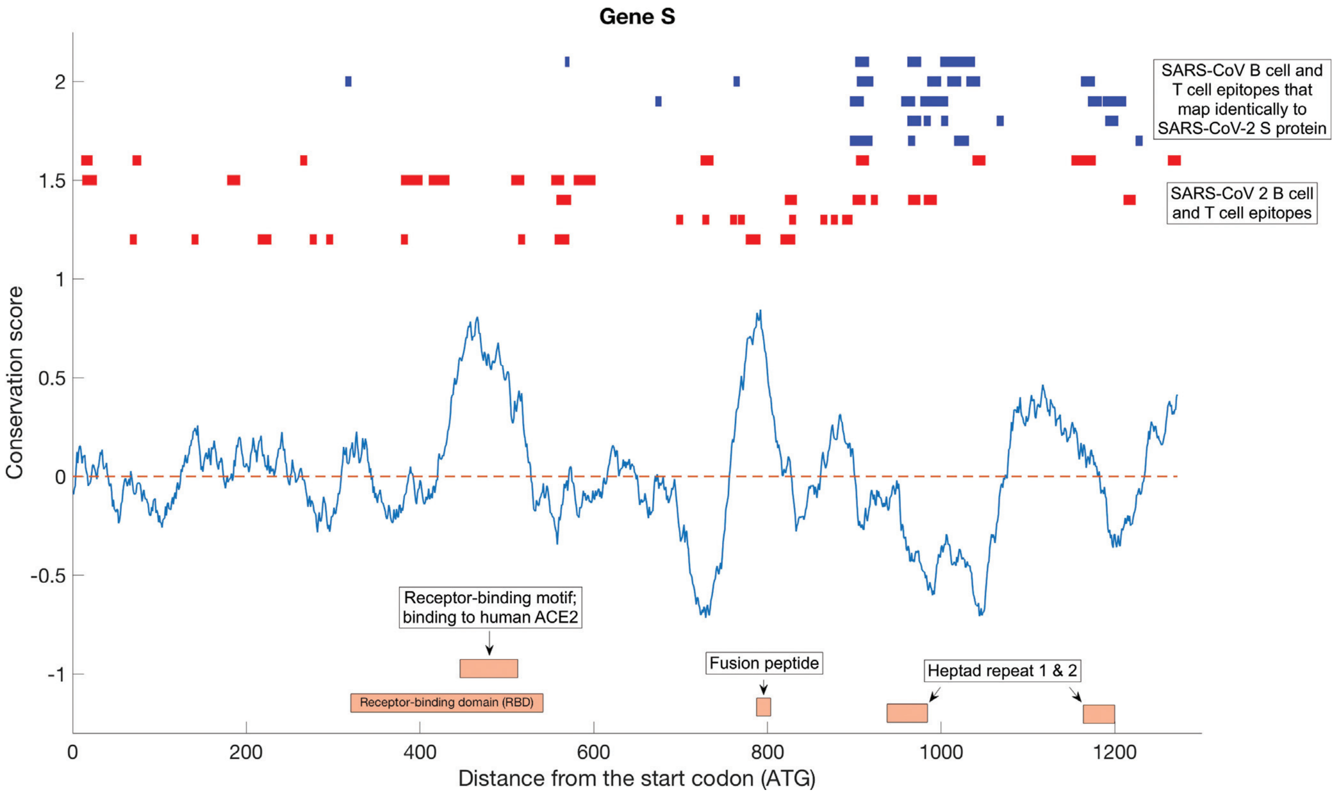


Fig. (2). Conservation profile of SARS-CoV-2 protein S. The solid blue line represents the position-specific estimations of the rate of evolution of each residue in the protein sequence, as a function of the distance from the start codon. The horizontal dotted line represents the threshold value, above which the score is characteristic of highly variable regions (0 for Rate4site). On the bottom, orange rectangles show the receptor-binding domain and its receptor binding motif to human ACE2, the fusion peptide, and the two heptad repeats. On the top, the SARS-CoV derived B cell and T cell epitopes map identically to SARS-CoV-2 N protein (in blue); on the bottom, the epitopes directly extrapolate for SARS-CoV-2 (in red). Each horizontal bar represents a linear epitope, consisting of continuous residues on the protein sequence of the protein S. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

2.2. Order-Disorder Propensities of Proteins S and N and Their Associated Epitopes

We predicted the structural order-disorder propensity proteins S and N to investigate the relationship between disordered structure and the spatial distribution of the SARS-CoV-2 derived B cell and T cell epitopes. To estimate the structural stability of a protein from its sequence without relying on the structure, we used the energy estimation approach at the core of the IUPred2A disorder prediction method (see Materials and Methods) [27]. In Fig. (**3**), we show the order-disorder propensity profile for protein N, together with the SARS-CoV-2 derived B cell and T cell epitopes.

The rationale to understand the results below is that the score of each residue in the protein sequence ranges from 0 (strong propensity for an ordered structure) to 1 (strong propensity for a disordered structure). Specifically, each residue in the sequence was classified as either ordered or disordered depending on whether the IUPred2A score is < 0.5 or > 0.5, respectively. We found a low but significant positive correlation between the order-disorder propensity profile and the conservation profile in Fig. (**1**) (Pearson correlation - coefficient = 0.3, p-value < 0.00001), implying that disordered regions of protein N tend to evolve faster than ordered ones. Both the RNA-binding domain and the dimerization domain are predicted to be ordered in a large percentage of their residues. We found that the vast majority of SARS-CoV B cell and T cell epitopes that map identically to SARS-CoV-2 N protein overlap these two functional regions that are also predicted to be conserved amino acid sites (Fig. **1** and Fig. **S3**). In contrast, some SARS-CoV--2-specific epitopes are located in predicted disordered regions outside the RNA-binding domain and the dimerization domain. In line with a previous study [28], we suggest that these disordered epitopes appear to be linear, making them ideally suited to incorporation into peptide vaccines. Nevertheless, it is worth noting that vaccines based on these epitopes may not be effective in the long term due to the high variability of corresponding protein regions (Fig. **1**).

Next, we studied the order-disorder propensity profile of S protein (Fig. **4**).

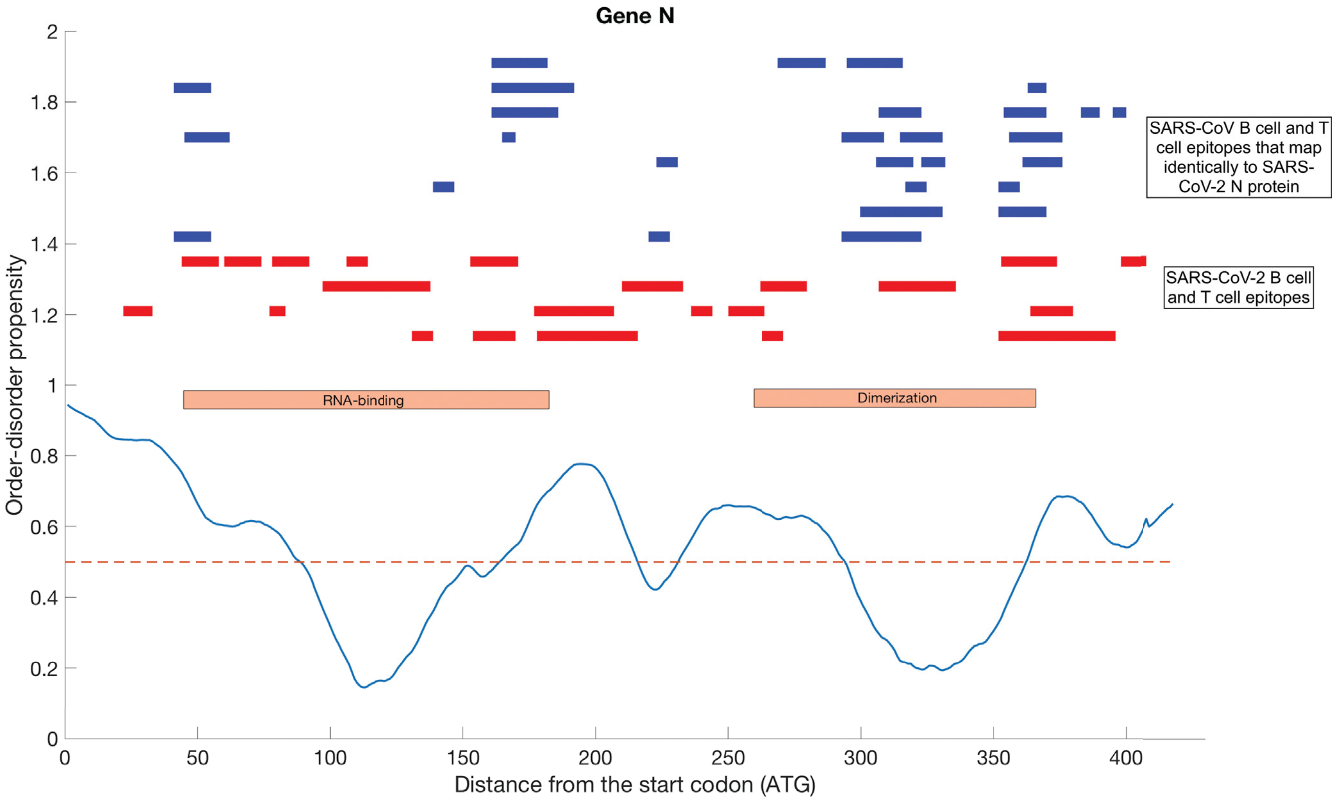


Fig. (3). Order-disorder propensity profile of SARS-CoV-2 protein N. The solid blue line represents the position-specific estimations of the structural order-disorder propensity as a function of the distance from the start codon. The horizontal dotted line represents the threshold value, above which the score is characteristic of disorder (0.5 for IUPred2A). On the top, the SARS-CoV derived B cell and T cell epitopes map identically to SARS-CoV-2 N protein (in blue); on the bottom, the epitopes directly extrapolate for SARS-CoV-2 (in red). The two orange rectangles show the RNA-binding domain and the dimerization domain. Each horizontal bar represents a linear epitope, consisting of continuous residues on the protein sequence of the protein N. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The spike protein is predicted to be ordered along the whole sequence. Both the receptor binding domain and the fusion peptide are well-structured. Also, in this case, we observe that the vast majority of the SARS-CoV-2 derived B cell and T cell epitopes are located in regions displaying reduced disorder tendency (Fig. **S4**).

3. DISCUSSION

In this study, we performed a systematic analysis of the rate of evolution and the structural order-disorder propensities of protein N and S in relation to the location of the SARS-CoV-2 epitopes derived from N and S proteins.

Identification of conserved epitopes is crucial to design broad-spectrum vaccines against the present outbreak of SARS-CoV-2 and the emergence of new SARS-CoV-2 variants that reduce the efficacy of the existing vaccines. Indeed, high-affinity neutralizing antibodies against conserved epitopes could provide immunity to SARS-CoV-2 and protection against eventual, future pandemic viruses. Conservation score measures the evolutionary conservation of an amino acid position in a set of homologous protein sequences. The rate of evolution is not constant among amino acid sites: some positions evolve slowly and are commonly referred to as “conserved”, whereas other positions evolve rapidly and are referred to as “variable”. The rate variations correspond to different levels of purifying selection acting on these sites [29]. This selection can result from geometrical constraints on protein folding and structure, constraints at amino acid sites involved in enzymatic activity, ligand binding, or protein-protein interactions.

Here, we used Rate4site to calculate the rate of evolution of each residue in the amino acid sequences of proteins N and S. We then analyzed the structural properties of the protein regions where the epitopes are located by studying their order-disorder propensity. We show the presence of both conserved epitopes and non-conserved epitopes in terms of rate of evolution (Figs. **1** and **3**). Specifically, the vast majority of the SARS-CoV-2 epitopes for the N protein are located in the RNA-binding and dimerization domains (Fig. **1**).

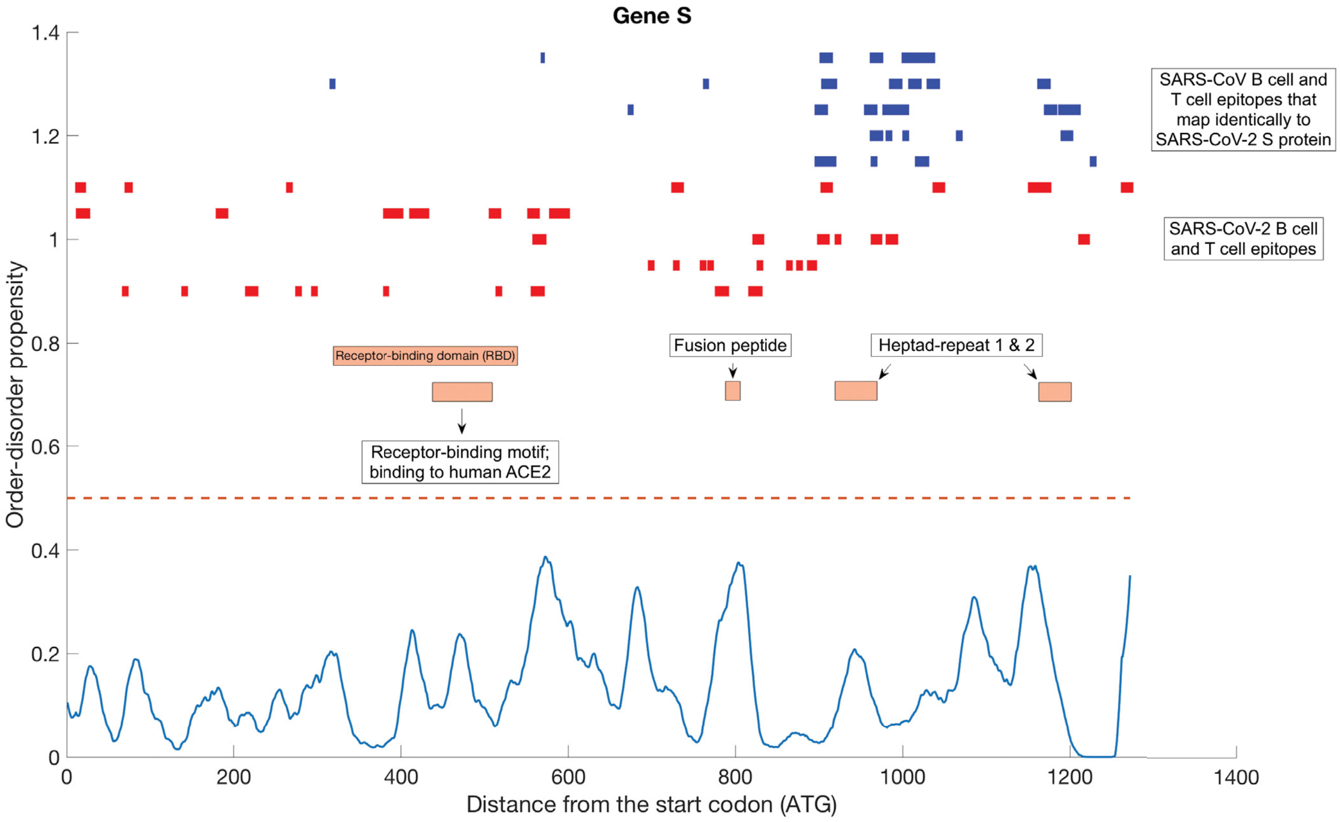


Fig. (4). Order-disorder propensity profile of SARS-CoV-2 protein S. The solid blue line represents the position-specific estimations of the structural order-disorder propensity as a function of the distance from the start codon. The horizontal dotted line represents the threshold value, above which the score is characteristic of disorder (0.5 for IUPred2A). On the top, the SARS-CoV derived B cell and T cell epitopes map identically to SARS-CoV-2 N protein (in blue); on the bottom, the epitopes directly extrapolate for SARS-CoV-2 (in red). Orange rectangles show the receptor-binding domain and its receptor binding motif to human ACE2, the fusion peptide, and the two heptad repeats 1 and 2. Each horizontal bar represents a linear epitope, consisting of continuous residues on the protein sequence of the protein S. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

In this case, we find epitopes in both ordered and disordered regions (Fig. **3**). Although we note that the vast majority of epitopes are located in regions having high rates of evolution, we also identify epitopes in conserved protein regions (Fig. **1**). Similarly, we observe the presence of SARS-CoV-2 epitopes for the S protein around the heptad repeats 1 and 2, which could be more immunogenic against SARS-CoV-2 variants because of their low rate of evolution (Fig. **2**). We thus suggest that the immune targeting of these conserved epitopes might potentially offer protection against this novel coronavirus and its variants.

Finally, it has been shown that numerous SARS S-protein-specific neutralizing antibodies recognize epitopes within the receptor-binding domain (RBD), thus blocking the RBD-ACE2 binding and preventing viral infection [30]. However, we show here that both the RBD (region 437-508) and the fusion peptide (region 788-806) are characterized by high rates of evolution, indicating a tendency for these two regions to mutate and overcome the host immunity.

4. MATERIALS AND METHODS

4.1. Data Sources

The complete coding genomic sequences of SARS-CoV-2 were obtained from NCBI viral databases, accessed as of 16th July, 2021. In this study, we considered seven human coronaviruses: human CoV-229E (HCoV-229E), human CoV-OC43 (HCoV-OC43), human Severe Acute Respiratory Syndrome Coronavirus (SARSr-CoV), human coronavirus NL63 (HCoV-NL63), human CoV-HKU1 (HCoV-HKU1), Middle East Respiratory Syndrome coronavirus (MERS-CoV), and the Severe Acute Respiratory Syndrome-related Coronavirus 2 (SARS-CoV-2). We downloaded the coding sequences of these coronaviruses from the National Center for Biotechnological Information (NCBI) (available at https://www.ncbi.nlm.nih.gov/). For each virus, we have investigated the evolutionary conservation and the structural disorder tendency of protein N (UniProt ID: P0DTC9) and S (UniProt ID: P0DTC2), because they are regarded as important targets for the development of vaccines and antiviral drugs.

4.2. Sequence Alignment

To explore the evolutionary relationship among the proteins N and S in the seven human coronaviruses here considered, the selected protein sequences were aligned by using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [31]. This tool is a multiple sequence alignment (MSA) program that uses seeded guide trees and HMM profile-profile techniques to generate alignments and phylogenetic trees of divergent sequences.

4.3. Disorder Prediction

The structural order-disorder propensity of each protein was predicted by using IUPred2A (https://iupred2a.elte.hu) [27], with the option for long disordered regions. Briefly, IUPred2A is a fast, robust, sequence-only predictor based on an energy estimation approach that allows to identify disordered protein regions. The key component of the calculations is the energy estimation matrix, a 20 by 20 matrix, whose elements characterize the general preference of each pair of amino acids to be in contact as derived from a reference set of globular proteins. This prediction associates a score to each residue in the protein sequence ranging from 0 (strong propensity for an ordered structure) to 1 (strong propensity for a disordered structure), using 0.5 as the threshold to classify residues as either ordered or disordered. In line with the original protocol [32], the position-specific estimations of the structural order-disorder propensity of each residue were then averaged over a window of 21 residues, and the average value was assigned to the central residue of the window (taking into account the limitations on both sides of the protein sequence).

4.4. Rate of Evolution for Site

We calculated the rate of evolution per-site of the SARS-CoV-2 proteins N and S relative to their orthologous proteins in other six human coronaviruses using Rate4site (https://m.tau.ac.il/~{}itaymay/cp/rate4site.html) [16]. Rate4Site calculates the evolutionary rate at each site in the MSA using a probabilistic-based evolutionary model. This allows taking into account the stochastic process underlying sequence evolution within protein families and the phylogenetic tree of the proteins in the family. The conservation score at each site in the MSA corresponds to the site's evolutionary rate. The position-specific estimations of the rate of evolution of each residue were then averaged over a window of 21 residues, and the average value was assigned to the central residue of the window (taking into account the limitations on both sides of the protein sequence). The size of the window was taken equal to that used above in section 2.3.

4.5. High Quality Set of Epitopes

In this study, we considered two groups of linear epitopes, consisting of continuous residues on the protein sequence of the proteins N and S. The first group comprises the whole set of SARS-CoV B cell and T cell epitopes that map identically to SARS-CoV-2 N and S proteins as identified by Ahmed *et al.* [15]. The second group consists of a high-quality set of previously identified epitopes directly extrapolated for SARS-CoV-2 [33-40].

CONCLUSION

In conclusion, our results suggest that targeting conserved regions of SARS-CoV-2 spike and nucleocapsid proteins with less plasticity and more structural constraint should have broader utility for antibody-based immunotherapy, neutralization, and prevention of escape variants.

AUTHORS' CONTRIBUTIONS

S.F., A.B., M.D. and A.G. conceived the study. S.F. conducted the analyses. S.F. and A.B. wrote the main manuscript. S.F., A.B., M.D. and A.G. read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article are available in the /ConservedEpitopes/ repository at: http://www.phys.uniroma1.it/doc/giansanti/BMS-CG-2021-19/ConservedEpitopes/.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Dilucca, M.; Forcelloni, S.; Georgakilas, A.G.; Giansanti, A.; Pavlopoulou, A. Codon usage and phenotypic divergences of SARS-CoV-2 genes. *Viruses,* **2020**, *12*(5), 1-21. http://dx.doi.org/10.3390/v12050498 PMID: 32366025

[2] Fouchier, R.A.; Kuiken, T.; Schutten, M.; van Amerongen, G.; van Doornum, G.J.; van den Hoogen, B.G.; Peiris, M.; Lim, W.; Stöhr, K.; Osterhaus, A.D. Aetiology: Koch's postulates fulfilled for SARS virus. *Nature,* **2003**, *423*(6937), 240. http://dx.doi.org/10.1038/423240a PMID: 12748632

[3] Woo, P.C.; Lau, S.K.; Chu, C.M.; Chan, K.H.; Tsoi, H.W.; Huang, Y.; Wong, B.H.; Poon, R.W.; Cai, J.J.; Luk, W.K.; Poon, L.L.; Wong, S.S.; Guan, Y.; Peiris, J.S.; Yuen, K.Y. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.,* **2005**, *79*(2), 884-895. http://dx.doi.org/10.1128/JVI.79.2.884-895.2005 PMID: 15613317

[4] van der Hoek, L.; Pyrc, K.; Jebbink, M.F.; Vermeulen-Oost, W.; Berkhout, R.J.; Wolthers, K.C.; Wertheim-van Dillen, P.M.; Kaandorp, J.; Spaargaren, J.; Berkhout, B. Identification of a new human coronavirus. *Nat. Med.,* **2004**, *10*(4), 368-373. http://dx.doi.org/10.1038/nm1024 PMID: 15034574

[5] Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.; Fouchier, R.A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.,* **2012**, *367*(19), 1814-1820. http://dx.doi.org/10.1056/NEJMoa1211721 PMID: 23075143

[6] Andersen, K.G.; Rambaut, A.; Lipkin, W.I.; Holmes, E.C.; Garry, R.F. The proximal origin of SARS-CoV-2. *Nat. Med.,* **2020**, *26*(4), 450-452. http://dx.doi.org/10.1038/s41591-020-0820-9 PMID: 32284615

[7] Gorbalenya, A.E.; Baker, S.C.; Baric, R.S.; de Groot, R.J.; Drosten, C.; Gulyaev, A.A. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.,* **2020**, *5*(4), 536-544. http://dx.doi.org/10.1038/s41564-020-0695-z PMID: 32123347

[8] Shrotri, M.; Swinnen, T.; Kampmann, B.; Parker, E.P.K. An interactive website tracking COVID-19 vaccine development. *Lancet Glob. Health,* **2021**, *9*(5), e590-e592. http://dx.doi.org/10.1016/S2214-109X(21)00043-7 PMID: 33667404

[9] Zhang, J.; Zeng, H.; Gu, J.; Li, H.; Zheng, L.; Zou, Q. Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines (Basel),* **2020**, *8*(2), 153. http://dx.doi.org/10.3390/vaccines8020153 PMID: 32235387

[10] Du, L.; He, Y.; Zhou, Y.; Liu, S.; Zheng, B.J.; Jiang, S. The spike protein of SARS-CoV-a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.,* **2009**, *7*(3), 226-236. http://dx.doi.org/10.1038/nrmicro2090 PMID: 19198616

[11] Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell,* **2020**, *181*(2), 281-292.e6. http://dx.doi.org/10.1016/j.cell.2020.02.058 PMID: 32155444

[12] Surjit, M.; Lal, S.K. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. *Infect. Genet. Evol.,* **2008**, *8*(4), 397-405. http://dx.doi.org/10.1016/j.meegid.2007.07.004 PMID: 17881296

[13] Sheikh, A.; Al-Taher, A.; Al-Nazawi, M.; Al-Mubarak, A.I.; Kandeel, M. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. *J. Virol. Methods,* **2020**, *277*, 113806. http://dx.doi.org/10.1016/j.jviromet.2019.113806 PMID: 31911390

[14] Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Musella, V.; Britti, D.; Bonizzi, L.; Urbani, A.; Roncada, P. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect.,* **2020**, *22*(4-5), 188-194. http://dx.doi.org/10.1016/j.micinf.2020.04.002 PMID: 32302675

[15] Ahmed, S.F.; Quadeer, A.A.; McKay, M.R. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses,* **2020**, *12*(3), 254. http://dx.doi.org/10.3390/v12030254 PMID: 32106567

[16] Pupko, T.; Bell, R.E.; Mayrose, I.; Glaser, F.; Ben-Tal, N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics,* **2002**, *18*(Suppl. 1), S71-S77. http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.S71 PMID: 12169533

[17] Hamming, I.; Timens, W.; Bulthuis, M.L.; Lely, A.T.; Navis, G.; van Goor, H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.,* **2004**, *203*(2), 631-637. http://dx.doi.org/10.1002/path.1570 PMID: 15141377

[18] Fehr, A.R.; Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.,* **2015**, *1282*, 1-23. http://dx.doi.org/10.1007/978-1-4939-2438-7_1 PMID: 25720466

[19] Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res.,* **2013**, *100*(1), 246-254. http://dx.doi.org/10.1016/j.antiviral.2013.08.014 PMID: 23994189

[20] Li, M.Y.; Li, L.; Zhang, Y.; Wang, X.S. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect. Dis. Poverty,* **2020**, *9*(1), 45. http://dx.doi.org/10.1186/s40249-020-00662-x PMID: 32345362

[21] Kuba, K.; Imai, Y.; Rao, S.; Gao, H.; Guo, F.; Guan, B.; Huan, Y.; Yang, P.; Zhang, Y.; Deng, W.; Bao, L.; Zhang, B.; Liu, G.; Wang, Z.; Chappell, M.; Liu, Y.; Zheng, D.; Leibbrandt, A.; Wada, T.; Slutsky, A.S.; Liu, D.; Qin, C.; Jiang, C.; Penninger, J.M. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat. Med.,* **2005**, *11*(8), 875-879. http://dx.doi.org/10.1038/nm1267 PMID: 16007097

[22] Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; Chen, H.D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.D.; Liu, M.Q.; Chen, Y.; Shen, X.R.; Wang, X.; Zheng, X.S.; Zhao, K.; Chen, Q.J.; Deng, F.; Liu, L.L.; Yan, B.; Zhan, F.X.; Wang, Y.Y.; Xiao, G.F.; Shi, Z.L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature,* **2020**, *579*(7798), 270-273. http://dx.doi.org/10.1038/s41586-020-2012-7 PMID: 32015507

[23] Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci. China Life Sci.,* **2020**, *63*(3), 457-460. http://dx.doi.org/10.1007/s11427-020-1637-5 PMID: 32009228

[24] Wang, H.; Yang, P.; Liu, K.; Guo, F.; Zhang, Y.; Zhang, G.; Jiang, C. SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. *Cell Res.,* **2008**, *18*(2), 290-301. http://dx.doi.org/10.1038/cr.2008.15 PMID: 18227861

[25] Millet, J.K.; Whittaker, G.R. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology,* **2018**, *517*, 3-8. http://dx.doi.org/10.1016/j.virol.2017.12.015 PMID: 29275820

[26] Liu, S.; Xiao, G.; Chen, Y.; He, Y.; Niu, J.; Escalante, C.R.; Xiong, H.; Farmar, J.; Debnath, A.K.; Tien, P.; Jiang, S. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet,* **2004**, *363*(9413), 938-947. http://dx.doi.org/10.1016/S0140-6736(04)15788-7 PMID: 15043961

[27] Mészáros, B.; Erdős, G.; Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.,* **2018**, *46*(W1), W329-W337. http://dx.doi.org/10.1093/nar/gky384 PMID: 29860432

[28] MacRaild, C.A.; Seow, J.; Das, S.C.; Norton, R.S. Disordered epitopes as peptide vaccines. *Pept. Sci. (Hoboken),* **2018**, *110*(3), e24067. http://dx.doi.org/10.1002/pep2.24067 PMID: 32328540

[29] Forcelloni, S.; Giansanti, A. Evolutionary forces and codon bias in different flavors of intrinsic disorder in the human proteome. *J. Mol. Evol.,* **2020**, *88*(2), 164-178. http://dx.doi.org/10.1007/s00239-019-09921-4 PMID: 31820049

[30] Sui, J.; Deming, M.; Rockx, B.; Liddington, R.C.; Zhu, Q.K.; Baric, R.S.; Marasco, W.A. Effects of human anti-spike protein receptor binding domain antibodies on severe acute respiratory syndrome coronavirus neutralization escape and fitness. *J. Virol.,*

2014, *88*(23), 13769-13780.
http://dx.doi.org/10.1128/JVI.02232-14 PMID: 25231316

[31] Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.,* **2019**, *47*(W1), W636-W641.
http://dx.doi.org/10.1093/nar/gkz268 PMID: 30976793

[32] Dosztányi, Z.; Csizmók, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.,* **2005**, *347*(4), 827-839.
http://dx.doi.org/10.1016/j.jmb.2005.01.071 PMID: 15769473

[33] Lee, E.; Sandgren, K.; Duette, G.; Stylianou, V.V.; Khanna, R.; Eden, J.S.; Blyth, E.; Gottlieb, D.; Cunningham, A.L.; Palmer, S. Identification of SARS-CoV-2 nucleocapsid and spike T-cell epitopes for assessing T-cell immunity. *J. Virol.,* **2021**, *95*(6), e02002-e02020.
http://dx.doi.org/10.1128/JVI.02002-20 PMID: 33443088

[34] Poh, C.M.; Carissimo, G.; Wang, B.; Amrun, S.N.; Lee, C.Y.; Chee, R.S.; Fong, S.W.; Yeo, N.K.; Lee, W.H.; Torres-Ruesta, A.; Leo, Y.S.; Chen, M.I.; Tan, S.Y.; Chai, L.Y.A.; Kalimuddin, S.; Kheng, S.S.G.; Thien, S.Y.; Young, B.E.; Lye, D.C.; Hanson, B.J.; Wang, C.I.; Renia, L.; Ng, L.F.P. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.,* **2020**, *11*(1), 2806.
http://dx.doi.org/10.1038/s41467-020-16638-2 PMID: 32483236

[35] Singh, P.; Tripathi, M.K.; Shrivastava, R. *In silico* identication of linear B-cell epitope in Coronavirus 2019 (SARS-CoV-2) surface glycoprotein: a prospective towards peptide vaccine. *Minerva Biotechnol. Biomol. Res.,* **2021**, *33*, 29-35.
http://dx.doi.org/10.23736/S2724-542X.20.02659-2

[36] Hisham, Y.; Ashhab, Y.; Hwang, S.H.; Kim, D.E. Identification of highly conserved SARS-CoV-2 antigenic epitopes with wide coverage using reverse vaccinology approach. *Viruses,* **2021**, *13*(5), 787.
http://dx.doi.org/10.3390/v13050787 PMID: 33925069

[37] Oliveira, S.C.; de Magalhães, M.T.Q.; Homan, E.J. Immunoinformatic analysis of SARS-CoV-2 nucleocapsid protein and identification of COVID-19 vaccine targets. *Front. Immunol.,* **2020**, *11*, 587615.
http://dx.doi.org/10.3389/fimmu.2020.587615 PMID: 33193414

[38] Lu, S.; Xie, X.X.; Zhao, L.; Wang, B.; Zhu, J.; Yang, T.R.; Yang, G.W.; Ji, M.; Lv, C.P.; Xue, J.; Dai, E.H.; Fu, X.M.; Liu, D.Q.; Zhang, L.; Hou, S.J.; Yu, X.L.; Wang, Y.L.; Gao, H.X.; Shi, X.H.; Ke, C.W.; Ke, B.X.; Jiang, C.G.; Liu, R.T. The immunodominant and neutralization linear epitopes for SARS-CoV-2. *Cell Rep.,* **2021**, *34*(4), 108666.
http://dx.doi.org/10.1016/j.celrep.2020.108666 PMID: 33503420

[39] Saini, S.K.; Hersby, D.S.; Tamhane, T.; Povlsen, H.R.; Amaya Hernandez, S.P.; Nielsen, M.; Gang, A.O.; Hadrup, S.R. SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8$^+$ T cell activation in COVID-19 patients. *Sci. Immunol.,* **2021**, *6*(58), eabf7550.
http://dx.doi.org/10.1126/sciimmunol.abf7550 PMID: 33853928

[40] Amrun, S.N.; Lee, C.Y.; Lee, B.; Fong, S.W.; Young, B.E.; Chee, R.S.; Yeo, N.K.; Torres-Ruesta, A.; Carissimo, G.; Poh, C.M.; Chang, Z.W.; Tay, M.Z.; Chan, Y.H.; Chen, M.I.; Low, J.G.; Tambyah, P.A.; Kalimuddin, S.; Pada, S.; Tan, S.Y.; Sun, L.J.; Leo, Y.S.; Lye, D.C.; Renia, L.; Ng, L.F.P. Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine,* **2020**, *58*, 102911.
http://dx.doi.org/10.1016/j.ebiom.2020.102911 PMID: 32711254