Potential immune epitope map for structural proteins of SARS-CoV-2

Yengkhom Damayanti Devi
Department of Molecular Biology and Biotechnology, Tezpur University
Himanshu Ballav Goswami
Department of Molecular Biology and Biotechnology, Tezpur University
Sushmita Konwar
Department of Molecular Biology and Biotechnology, Tezpur University
Chandrima Doley
Department of Molecular Biology and Biotechnology, Tezpur University
Anutee Dolley
Department of Molecular Biology and Biotechnology, Tezpur University
Arpita Devi
Department of Molecular Biology and Biotechnology, Tezpur University
Chen Chongtham
National Institute of Immunology, Aruna Asaf Ali Marg, Jawaharlal Nehru University
Vashkar Biswa
Department of Biotechnology, Bodoland University
Latonglila Jamir
Department of Environmental Science, Nagaland University (Central)
Aditya Kumar
Department of Molecular Biology and Biotechnology, Tezpur University
Siddhartha Shankar Satapathy
Department of Computer Science and Engineering, Tezpur University
Suvendra Kumar Ray
Department of Molecular Biology and Biotechnology, Tezpur University
Ramesh Chandra Deka
Department of Chemical Sciences, Tezpur University
Robin Doley
Department of Molecular Biology and Biotechnology, Tezpur University
Manabendra Mandal
Department of Molecular Biology and Biotechnology, Tezpur University
Sandeep Das
Department of Biotechnology, Bodoland University
Ch. Shyamsunder Singh

Department of Paediatrics, Regional Institute of Medical Sciences

Partha Pratim Borah

Department of Paediatrics and Neonatology, Pratiksha Hospital

Pabitra Nath

Department of Physics, Tezpur University

Nima D. Namsa (namsa@tezu.ernet.in)

Tezpur University, Assam, India

Research Article

Keywords: SARS-CoV-2, immune epitope mapping, in silico predictions, structural proteins, MHC class I and class II binding epitopes

Posted Date: August 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-47663/v1

Version of Record: A version of this preprint was published at PLOS ONE on August 18th, 2020. See the published version at https://doi.org/10.1371/journal.pone.0258645.

Abstract

Researchers around the world are developing more than 145 vaccines (DNA/mRNA/whole-virus/viral-vector/protein-based/repurposed vaccine) against the SARS-CoV-2 and 21 vaccines are in human trials. However, a limited information is available about which SARS-CoV-2 proteins are recognized by human B- and T-cell immune responses. Using a comprehensive computational prediction algorithm and stringent selection criteria, we have predicted and identified potent B- and T-cell epitopes in the structural proteins of SARS-CoV and SARS-CoV-2. The amino acid residues spanning the predicted linear B-cell epitope in the RBD of S protein (370-NSASFSTFKCYGVSPTKLNDLCFTNV-395) have recently been identified for interaction with the CR3022, a previously described neutralizing antibody known to neutralize SARS-CoV-2 through binding to the RBD of the S protein. Intriguingly, most of the amino acid residues spanning the predicted B-cell epitope (aa 331-NITNLCPFGEVFNATRFASVYAWNRK-356, 403-RGDEVRQIAPGQTGKIADYNYKLPD-427 and aa 437- NSNNLDSKVGGNYNYLYRLFRKSNL-461) of the S protein have been experimentally verified to interact with the cross-neutralizing mAbs (S309 and CB6) in an ACE2 receptor-S protein interaction independent-manner. In addition, we found that computationally predicted epitope of S protein (370-395) is likely to function as both linear B-cell and MHC class II epitope. Similarly, 403-27 and 437-461 peptides of S protein were predicted as linear B cell and MHC class I epitope while, 177-196 and 1253-1273 peptides of S protein were predicted as linear and conformational B cell epitope. We found MHC class I epitope 316-GMSRIGMEV-324 predicted as high affinity epitope (HLA-A*02:03, HLA-A*02:01, HLA-A*02:06) common to N protein of both SARS-CoV-2 and SARS-CoV (N317-325) was previously shown to induce interferon-gamma (IFN-γ) in PBMCs of SARS-recovered patients. Interestingly, two MHC class I epitopes, 1041-GVVFLHVTY-1049 (HLA-A*11:01, HLA-A*68:01, HLA-A*03:01) and 1202-FIAGLIAIV-1210 (HLA-A*02:06, HLA-A*68:02) derived from SARS-CoV S protein with epitope conservancy between 85 to 100% with S protein of SARS-CoV-2 was experimentally verified using PBMCs derived from SARS-CoV patients. We observed that HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*11:01, HLA-A*30:01, HLA-A*68:01, HLA-A*68:02, HLA-B*15:01 and HLA-B*35:01 have been predicted to bind to the maximum number of MHC class I epitope (based on the criterion of allele predicted to bind more than 30 epitopes) of S protein of SARS-CoV-2. Similarly, we observed that HLA-A*02:06, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*68:01, HLA-A*68:02, HLA-B*15:01 and HLA-B*35:01 are predicted to bind to the maximum number of MHC class I epitope of N protein of SARS-CoV-2. We found that HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*13:02, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DQA1*04:01, DQB1*04:02, HLA-DPA1*02:01, DPB1*01:01, HLA-DPA1*01:03, DPB1*02:01, HLA-DPA1*01:03, DPB1*04:01, HLA-DPA1*03:01, DPB1*04:02, HLA-DPA1*02:01, DPB1*05:01, HLA-DPA1*02:01, and DPB1*14:01 are predicted to bind to the maximum number of MHC class II epitope of S protein of SARS-CoV-2. Alleles such as HLA-DRB1*04:01, HLA-DRB1*07:01, HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*13:02, HLA-DRB3*02:02, HLA-DRB5*01:01, HLA-DQA1*01:02, DQB1*06:02, DPB1*05:01 and HLA-DPA1*02:01 are found to interact with the maximum number of MHC class II epitope of N protein of SARS-CoV-2. Using the IEDB tool we found the occurrence of HLA alleles with population coverage of around 99% throughout the world. The findings of computational predictions of

mega-pool of B- and T-cell epitopes identified in the four main structural proteins of SARS-CoV-2 provides a platform for future experimental validations and the results of present works support the use of RBD or the full-length S and N proteins in an effort towards designing of recombinant protein-based vaccine and a serological diagnostic assay for SARS-CoV-2.

Introduction

COVID-19, a disease caused by a newly identified strain of coronavirus called the 'severe acute respiratory syndrome coronavirus' 2 (SARS-CoV-2) was responsible for pneumonia-like symptoms in Wuhan, Hubei province, China in the early December of 2019. At a time while we communicate the findings of computational work, the ongoing pandemic that first started in December 2019 in Wuhan, China and spread worldwide across 216 countries and territories as on May 23, 2020 as per the World Health Organization has now confirmed cases of 50,61,476 and confirmed deaths of 3,31,475. Phylogeny of whole genome sequences (~29,000 to 30,000 nucleotides) isolated from infected patients revealed that SARS-CoV-2 was highly similar to the genome of RaTG13 with an overall identity of nearly 96.2%, but distant from SARS-CoV BJ01 (about 79.6% sequence identity, GenBank accession number AY278488.2) and MERS-CoV (about 50% sequence identity, GenBank accession number NC_019843.3)[1,2].

The genome of SARS-CoV-2 encodes for structural and nonstructural proteins (NSPs). The first ORFs (ORF1a/b) encode for 16 NSPs (NSP1-16), except gamma Coronavirus that lacks NSP1. There is a -1 frameshift between ORF1a and ORF1b leading to the production of two polypeptides (pp1a and pp1ab). These polyproteins are post-translationally processed by viral encoded chymotrypsin-like protease (3CLpro) or main protease (Mpro) and one or two papain-like proteases into 16 NSPs. The structural proteins include spike glycoprotein (S), envelope protein (E), membrane protein (M) and nucleocapsid protein (N). SARS-CoV (2003) and Middle East respiratory syndrome coronavirus (MERS-CoV) (2012) have caused world-wide outbreaks in the past two decades. The surface spike glycoprotein (S) has two functional subunits that mediate cell attachment (S1 subunit, existing of four core domains S1A through S1D) and virus-host membrane fusion (S2 subunit). The S proteins of SARS-CoV-2 and SARS-CoV are phylogenetically closely related having an amino-acid sequence identity of around 77%[1] and uses the same cellular receptor called the angiotensin-converting enzyme 2 (ACE2) for entry into cells[3].

The serum antibody response of patients with SARS have found that the antibody response to SARS-CoV appears to be dominated by antibodies to S and N proteins with fewer antibodies against E and M proteins[4,5,6]. Currently, the main diagnostic test for SARS-CoV-2 relies on the detection of the SARS-CoV-2 RNA genome using the real-time reverse-transcriptase polymerase chain reaction (RT-PCR). *In silico* identification of B- and T-cell epitopes can predict potential candidate targets for immune responses and would assist in the selection of specific target antigens for the development of a vaccine or diagnostic test for SARS-CoV-2. Here, we have used validated IEDB and online bioinformatic resources to predict B- and T-cell epitopes of four structural proteins (S, E, M and N) of SARS-CoV and SARS-CoV-2.

Bioinformatic prediction has led to the identification of potent common and viral-species specific B- and T-cell epitope that are likely to be recognized in humans and also determined the conservancy of predicted epitopes across different species of coronaviruses (CoVs).

Results

Prediction of continuous B-cell epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV

A recently published paper revealed that S proteins of SARS-CoV-2 and SARS-CoV are closely related with an amino-acid sequence identity of around 77%[1] and SARS-CoV-2 uses ACE2 as the cellular receptor for entry into cells[3]. In this work, we have identified and predicted potential B- and T-cell epitopes in the structural proteins (S, E, M and N) of SARS-CoV and SARS-CoV-2 using the validated prediction tools (Fig.1). BepiPred[8] has identified and predicted around 22 B-cell epitope in SARS-CoV-2 S protein, 14 of these epitopes are likely to be antigenic as predicted by vaxigen (Table 1a). Similarly, Bcepred[9] has predicted around 27 B cell epitope in S protein of which 21 of them are likely antigenic epitopes. BepiPred have predicted around 17 B-cell epitope in SARS-CoV-2 N protein which are likely to be antigenic as predicted whereas Bcepred have predicted around 14 B-cell epitope in N protein of which 10 of them are likely to be antigenic epitopes (Table 1a). Both Bepipred and Bcepred have predicted a similar number of epitopes for S and N protein of SARS-CoV (Table 1a). As expected, a smaller number of B-cell epitopes were predicted for E and M proteins of SARS-CoV-2 and SARS-CoV. The prediction results are suggestive of a good degree of correlation between the size of proteins and total number of predicted epitopes. A heat map was generated using R programming for the predicted B- and T-cell epitope. Both Bepipred and Bcepred predicted linear B cell epitope showed the distribution of antigenic epitope in the S1 domain of S protein of SARS-CoV-2 (Fig.2a). As there are no known bioinformatic prediction server that could predict the common epitopes, we have manually analyzed and sorted out B-and T-cell epitopes that are common to SARS-CoV-2 and SARS-CoV. We have found a conserved epitope of S (407-VRQIAPGQTG-416, 421-YNYKLPD-427,1028-KMSECVLGQSKRVDFCGKGYHL-1049, 1254-CKFDEDDSEPVLKGVKLHYT-1273), N (173-AEGSRGGSQASSRSSSR-189 & 235-SGKGQQQQGQTVT-247), M (163-DLPKEITVATSRTLSYYKLG-182) and E (58-VYSRVKNLNSS-68) proteins that are common to both SARS-CoV and SARS-CoV-2 (Table 1b).

The accessibility of linear B-cell epitopes of S, E, M and N proteins predicted as antigenic, non-allergenic and nearly conserved properties were visualized using BIOVIA Discovery Studio 2017 R2 and the results revealed that epitopes are likely to be localized on the surface of 3D-structure of proteins (Table 1c and Fig. 3a-e). The predicted linear B-cell epitopes were enriched in the S and N proteins, whereas less enriched in the M and E proteins of SARS-CoV and SARS-CoV-2 (Table 1a & 1c). Most of the predicted epitope peptides showed good helical content as predicted based on Agadir score[14,15]. In addition to Bepipred and Bcepred, we have predicted the presence of continuous (linear) B-cell epitope of S, E, M and N proteins of SARS-CoV-2 using the ABCpred server. ABCpred is used for prediction of B cell epitope in an antigen sequence with 65.93% accuracy using recurrent neural networks[27]. In this study, we applied a

window length of 18 amino acid (aa) with a threshold setting of 0.7, the predicted epitopes were further verified by VexiJen V2.0 to find out antigenic B-cell epitopes. A total of 52 antigenic linear B cell epitopes were predicted for S, 18 for N, 7 for M and 3 for E proteins of SARS-CoV-2 (Table 1d), many of these epitopes were also found to be predicted by Bepipred and Bcepred (Table 1c).

We have carried out sequence alignment of RBDs of S protein of SARS-CoV and SARS-CoV-2 using clustal omega (Fig.4). Inset table of Fig.4. shows the linear and conformation B cell epitope predicted and identified in the RBD of the S protein of SARS-CoV-2 and notably, we have found a linear B-cell epitope of S protein (370-NSASFSTFKCYGVSPTKLNDLCFTNV-395, inset table of Fig. 4) that was recently found to interact with CR3022. CR3022 is a previously described neutralizing antibody isolated from a convalescent SARS patient have now found to neutralize SARS-CoV-2[28,29]. Co-crystallization of CR3022 and RBD of the S protein of SARS-CoV-2 have identified a conserved epitope residues enabling cross-reactive neutralization between SARS-CoV-2 and SARS-CoV without hampering ACE2 binding to RBD of S protein[28]. Many of the predicted linear and conformational B cell epitopes (Inset table of Fig.4) predicted in this work are found to be located in the RBD of S protein of SARS-CoV-2 ($S1_B$ residues 338-506) and a receptor-binding subdomain (residues 438-498). 47D11 is the first report of a human monoclonal antibody that neutralizes SARS-CoV-2 most likely by binding to the conserved core structure of the $S1_B$ (residues 338-506) using a mechanism independent of receptor-binding inhibition[30]. Similarly, several virus-specific memory B cells recognizing the RBD of the S protein of SARS-CoV-2 have been identified and only two (of the total 206 SARS-CoV-2 RBD-specific monoclonal antibodies) antibodies were found to block the viral entry which correlated with high competing capacity against ACE2 receptor[31]. The findings are suggestive of virus species-specific antibody response to RBDs of SARS and cross-recognition target regions outside the RBDs. Amanat et al (2020) have developed a serological assay using the S protein expressed in insect (iSpike/iRBD) and mammalian (mSpike/mRBD) expression system. They observed that all COVID-19 plasma/serum samples reacted strongly to both RBD and full-length spike protein. Reactivity of COVID-19 sera was, in general, stronger against the full-length S protein than against the RBD. They have tested an additional 12 serum samples from patients with acute COVID19 disease, as well as convalescent participants, for reactivity to mRBD and mSpike. All 12 samples reacted with both RBD and spike protein[32].

Prediction of conformational B-cell epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV

The Cryo-EM structure of S protein of SARS-CoV-2 (PDB:6VSB[3]) is available from residues 1 to 1208 that had missing 65 residues of C-terminal region of the protein. Similarly, the Cryo-EM solved structure (residues 1 to 1195) of S protein of SARS-CoV had missing residues of 60 amino acids towards the C-terminal end (PDB:5WRG[33]) due to limited electron density. We have overcome this problem using I-TASSER modelled structures of S protein of SARS-CoV and SARS-CoV-2[34]. We compared modelled structures with cryo-EM solved structure[3,34] and observed high similarity with root mean square deviation (RMSD) of around 1.3 Å (SARS-CoV) and 2.0 Å (SARS-CoV-2) (data not shown). Similarly, the structures of E, M and N proteins were modelled and used as input structures to identify potent conformational B-

cell epitopes using CBTOPE, DiscoTope, ElliPro and EPSVR. We identified a cluster of conformational B-cell epitopes predominately enriched in S and N proteins of SARS-CoV and SARS-CoV-2 (Table 2a & 2c). We have generated a heat map showing the distribution of antigenic epitopes in the S1 domain of S protein of SARS-CoV-2 (Fig.2b). Notably, we found conformational B-cell epitope of S, E, M and N proteins common to both SARS-CoV-2 and SARS-CoV (Table 2b). The findings of this work revealed that predicted conformational epitopes are likely to be localized on the accessible region of 3D-structure of proteins as visualized by BIOVIA Discovery Studio 2017 R2 (Fig. 5a-e).

Prediction of MHC I binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV

SARS-CoV is closely related human beta-CoV to SARS-CoV-2 and a previous data from SARS-CoV patients in 2003-2004 has identified $CD8^+$ T and $CD4^+$ T cell responses using the whole proteome[35], a likely possibility has been that substantial CD4+ T cell, CD8+ T cell, and neutralizing antibody responses develop to SARS-CoV-2 and all contribute to clearance of the acute infection and protective immunity against SARS-CoV-2 infection[36,37]. The availability of information of SARS-CoV-2 proteins and epitopes recognized by human T-cell responses can greatly assist researchers in selecting potential epitopes or target proteins for the design of the candidate vaccine. All the four structural proteins (S, E, M and N) of SARS-CoV-2 and SARS-CoV were screened and epitopes were predicted that are likely to be presented on 27 HLA class I molecules (HLA-A and HLA-B) molecules using three different servers, proteasomal cleavage/TAP transport/MHC class I combined predictor[17], MHC-NP[18] and TepiTool[16].

We selected the high affinity-ranked peptides and found nearly 683 MHC I epitope for S protein of SARS-CoV-2, as predicted using proteasomal cleavage/TAP transport/MHC class I combined predictor, 683 by MHC-NP and 341 by Tepitool (Table 3a). Similarly, we observed likely number of MHC I epitopes of S protein of SARS-CoV predicted using the proteasomal cleavage/TAP transport/MHC class I combined predictor (674), MHC-NP (673) and TepiTool (331). This observations give us some confidence in the ability of three independent servers with a varying operating algorithm to predict a consistent number of unique MHC I epitopes across all four proteins (S, E, M & N) and corroborate with the overall high sequence similarity of SARS-CoV and SARS-CoV-2 (Fig. 6a). The predicted epitopes were further screened for their ability to elicit an IFN-γ response using the IFNepitope tool[22] and the epitopes that are predicted as likely conserved across the species of CoVs, antigenic, non-allergenic have been shortlisted. It is also worth mentioning that the same epitope is being predicted by more than two servers employed in this study. Notably, we observed and identified 6 CD8+ T cell epitopes in S, 6 epitopes in N, 5 epitopes each in M and E proteins that are found common to both SARS-CoV-2 and SARS-CoV (Table 3b). Table 3c shows the list of top MHC I binding epitopes that are antigenic, non-allergenic, a positive score for IFN-γ and the majority of the listed epitopes are predicted by more than two servers used in this study. A total of 41 MHC I epitope was found for S protein, 19 each for N & M and 10 for E protein of SARS-CoV-2 (Table 3c).

We found that one epitope of N protein (316-GMSRIGMEV-324) common to both SARS-CoV-2 and SARS-CoV (N317-325, Table 3b) predicted as high affinity epitope (HLA-A*02:03, HLA-A*02:01, HLA-A*02:06) was previously known to induce interferon-gamma (IFN-γ) in PBMCs of SARS-recovered patients[38].

Interestingly, two epitopes, 1041-GVVFLHVTY-1049 (HLA-A*11:01, HLA-A*68:01, HLA-A*03:01) and 1202-FIAGLIAIV-1210 (HLA-A*02:06, HLA-A*68:02, Table 3c) derived from SARS-CoV S protein having epitope conservancy between 85 to 100% with S protein of SARS-CoV-2 was previously reported as immunogenic in PBMCs derived from SARS-CoV patients[38]. CD4+ and CD8+T cells can recognize antigen only when it is presented by a self-MHC molecule and MHC molecules are extremely polymorphic. Selecting multiple peptides with different HLA binding specificities will afford increased coverage of the patient population targeted by peptide-based vaccines and diagnostic methods. Using the IEDB tool we found the occurrence of HLA alleles with population coverage of around 99% throughout the world except for central America where the population coverage of HLA allele was less than 10% (Fig.7a). We found that HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*11:01, HLA-A*30:01, HLA-A*68:01, HLA-A*68:02, HLA-B*15:01 and HLA-B*35:01 have been predicted to bind to the maximum number of epitope (based on the criterion of allele predicted to bind more than 30 epitopes) of S protein of SARS-CoV-2 (Table S1a). Similarly, we observed that HLA-A*02:06, HLA-A*30:01, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*68:01, HLA-A*68:02, HLA-B*15:01 and HLA-B*35:01 are predicted to bind to the maximum number of epitope (based on the criterion of allele predicted to bind more than 10 epitopes) of N protein of SARS-CoV-2 (Table S1a).

Prediction of MHC II binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV

Helper T cells are required for adaptive immune responses and they help activate B cells to secrete antibodies and help activate cytotoxic T cells to kill infected target cells. NetMHCIIpan 4.0[19] and TepiTool[16] have been used to predict and identify high-affinity MHC II binding epitopes based on 23 HLA class II molecules (HLA-DP, HLA-DQ and HLA-DR) molecules. We found nearly 884 and 519 MHC II epitope for S protein of SARS-CoV-2 predicted using NetMHCIIpan 4.0 and TepiTool (Table 4a). A similar number of epitopes were predicted for S protein of SARS-CoV, the findings of computational predictions are suggestive of reliability and good correlation. Interestingly, we found a total of eight (8) CD4+ T cell epitope for S, six (6) for N, two (2) for M and four (4) for E proteins that are common to both SARS-CoV-2 and SARS-CoV with epitope conservancy of approximately between 80% to 100% (Table 4b). We have selected the top list of MHC II epitopes based on antigenicity, non-allergenicity and conservancy and all these epitopes were predicted to induce IFN-gamma as shown in Table 4c. The allele-wise distribution of predicted epitopes for S, E, M and N proteins of SARS-CoV-2 have been presented in a heat map (Fig.6b). Using the IEDB tool we found the occurrence of HLA alleles with population coverage of around 99% throughout the world except for South Africa where the population coverage of HLA class II allele was around 30% (Fig.7b).

We found that HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*13:02, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01, HLA-DQA1*04:01, DQB1*04:02, HLA-DPA1*02:01, DPB1*01:01, HLA-DPA1*01:03, DPB1*02:01, HLA-DPA1*01:03, DPB1*04:01, HLA-DPA1*03:01, DPB1*04:02, HLA-DPA1*02:01, DPB1*05:01, HLA-DPA1*02:01, and DPB1*14:01 are predicted to bind to the maximum number of epitope (based on the criterion of allele predicted to bind more than 30 epitopes) of S protein of SARS-CoV-2 (Table S1b). Similarly, we observed that HLA-DRB1*04:01, HLA-DRB1*07:01,

HLA-DRB1*08:02, HLA-DRB1*09:01, HLA-DRB1*11:01, HLA-DRB1*13:02, HLA-DRB3*02:02, HLA-DRB5*01:01, HLA-DQA1*01:02, DQB1*06:02, DPB1*05:01 and HLA-DPA1*02:01 are found to interact with the maximum number of epitope (based on the criterion of allele predicted to bind more than 10 epitopes) of N protein of SARS-CoV-2 (Table S1b). The findings of computational predictions are suggestive that T-cell based immunity might be generated largely against S and N protein of SARS-CoV-2 and could potentially be selected as target candidate for recombinant protein-based vaccine. On careful observations of computationally predicted epitopes (Table 5), we found epitope in $S1_B$ of spike protein (370-NSASFSTFKCYGVSPTKLNDLCFTNV-395) likely to function as both linear B-cell and MHC class II epitope. 403-RGDEVRQIAPGQTGKIADYNYKLPD-427 and 437-NSNNLDSKVGGNYNYLYRLFRKSNL-461 peptides of S protein were predicted as linear B cell and MHC class I epitope. 177-MDLEGKQGNFKNLREFVFKN-196 and 1253-CCKFDEDDSEPVLKGVKLHYT-1273 peptides of S protein were predicted as linear and conformational B cell epitope. Similar overlapping B- and T-cell epitopes were predicted for E, M and N proteins of SARS-CoV-2 (Table 5).

Utilizing bioinformatic approaches a few researchers have identified specific peptides in SARS-CoV-2 with increased probability of being T-cell targets[37,39] and have identified circulating SARS-CoV-2-specific CD8+ and CD4+ T-cells in approximately 70% and 100% of COVID-19 convalescent patients, respectively[40]. CD4+ T cell responses were observed predominately in the S protein and the robustness of T cell responses was correlated with the magnitude of IgG and IgA titers of SARS-CoV-2. Among the structural proteins, the S and M proteins were mainly recognized as possible targets for CD8+ T cells of SARS-CoV-2[40].

Identification of potential B and T cell epitope of structural proteins for serological assay and multi-epitope-based vaccine

The spike protein of SARS-CoV-2 and SARS-CoV share an amino-acid sequence identity of around 77% and are phylogenetically closely related[1]. Previously characterized immune response of patients with SARS have found maximum neutralizing antibodies against the S and N proteins[4,5,6] and the recent findings showed S protein as an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients[32,41,42]. Therefore, S and N proteins are the major targets for development of vaccine and serological assay because SARS-CoV-2 neutralization assays are time-consuming and require BSL-3 containment facilities. In this study, we have designed a multi-epitope chimeric constructs by utilizing the computationally predicted putative and potent B- and T-cell epitopes of S, E, M and N proteins of SARS-CoV-2 (Table 6). In the multi-epitope constructs we have included MHC class I (CTL) and class II binding epitopes (HTL) having the positive score for IFN-γ induction. CTL and HTL epitopes were linked together by AAY and KK cleavable linkers whereas the B- cell epitopes (linear and conformational) were linked together with GGGGS flexible linker (Fig. 8a-f). Poly-Gly-rich linkers can be considered as independent units and do not affect the function of the individual proteins. A total of six (6) multi-epitope chimeric constructs namely, (i) RBD of S protein (B- and T-cell epitopes), (ii) Full-length S protein (B- and T-cell epitopes), (iii) Structural protein construct (B- and T-cell epitopes of S, E, M and N proteins), (iv) Chimeric

construct of S and N proteins (B- and T-cell epitopes of S and N), (v) S protein (B-cell epitopes of S protein) and (vi) Nucleocapsid protein (B-cell epitopes of N protein) were designed containing N-CTL, HTL and B-cell epitopes (Fig.8a-f). The structure of chimeric multi-epitope constructs has been modelled using I-TASSER (Fig.9a-f) and validated by RAMPAGE server to generate a Ramachandran plot[43]. Most of the amino acid residues of epitope constructs were found in the favourable region (Fig.S1, Inset table) except two constructs (RBD of S and nucleocapsid protein). Various physicochemical parameters that included number of residues, theoretical pI, molecular weight, aliphatic index and grand average of hydrophobicity (GRAVY) were analysed by ProtParam[44]. Based on the aliphatic index scores, the multi-epitope constructs could be considered moderately thermostable (Table S2). Gravy scores obtained were of a negative value for all the constructs indicating the likelihood of the chimeric multi-epitope antigen being globular and hydrophilic in nature. The secondary structure of multi-epitope constructs was predicted by online server PSIPRED[45] and helical and strand content of the constructs are provided in Table S2 and Fig.S2. Cathepsin and carboxypeptidase involved in proteolytic processing of endocytosed proteins in the MHC class II pathway display preferential cleavage of dibasic (RR, KK, KR or RK) sites[46]. Proteases, which provide peptide ligands for the MHC class II antigenic presentation pathway display preferential cleavage of hydrophobic motifs (AAY). The cleavable linkers are required to be accessible for the proteases associated with MHC I and II antigen processing pathway. It is observed that the cleavable linker residues (AAY and KK) used in multi-epitope subunit vaccines were surface accessible based on computational prediction algorithms indicating high probability of T-cell epitopes presentation by MHC molecules as visualized by discovery studio (Fig.10a-f). The results of C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.php) prediction revealed the ability of multi-epitope constructs to simulate IFN-γ production following an immunization with the peptide (Fig. 11a-f). *In-silico* immune simulation of the multi-epitope constructs {(A) RBD of spike protein, (B) Spike protein, (C) Structural protein construct and (D) Chimeric construct of S and N proteins)} showed consistent results with the actual immune responses as observed by the primary response of high levels of IgM. This is followed by marked increase in B-cell populations and levels of IgG_1 + IgG_2, IgM, and IgG + IgM antibodies as a part of the secondary and tertiary responses (Fig. 12, A-D (i-iv). A similarly high response was observed in the T_H (helper) and T_C (cytotoxic) cell populations with corresponding memory development especially for constructs made of structural proteins (S, E, M and N) and chimeric construct of S and N proteins (Fig.12C & D).

Discussion

In this study, using validated bioinformatics resources, we comprehensively screened and identified potential B- and T-cell epitope predominately clustered in S and N proteins of SARS-CoV and SARS-CoV-2. This is significant for the fact that knowledge of the innate and adaptive immunity to SARS-CoV-2 can provide crucial information on the vaccine and serological assays for coronavirus disease 2019 (COVID-19) and its pathogenesis. Limited information is available about which SARS-CoV-2 proteins are recognized by human T cell immune responses. Both B- and T-cells are the weapons of adaptive immune response and computationally predicted epitopes are likely to provide information about the possible targets of the immune response to SARS-CoV-2.

Our bioinformatics analysis of structural proteins has revealed that most of the predicted B cell epitopes were found in S and N proteins of SARS-CoV and SARS-CoV-2 (Table 1a & Table 2a). Importantly, five highly conserved amino acid (aa) regions were identified, of which three are located in the RBD of S protein {aa 370-395 (linear B cell epitope and MHC II), aa 403-427 and aa 437-461 (linear B cell epitope and MHC I)}, one in the N (aa 177-196) and C-terminal region of S protein (aa 1253-1273) were predicted and likely to function as both linear and conformational B-cell epitopes (Table 5). Similarly, five conserved amino acid regions were identified for N protein, of which aa 32-46 and aa 366-394 were predicted as linear and conformational B-cell epitopes and MHC I. The peptide aa 274-283 was predicted as linear and conformational B-cell epitopes and the peptide aa 173-190 and aa 227-266 were predicted as linear B-cell epitopes and MHC I. Three conserved amino acid regions (aa 35-48, aa 101-119 and aa 160-182) were identified for M protein are likely to function as both linear and conformational B-cell epitope and MHC I. One conserved amino acid region (aa 57-75) was identified for E protein which is likely to function as both linear and conformational B-cell epitope. One epitope (aa1253-1273) for S, two (aa 173-190, aa 274-283) for N, two for M (aa 101-119, aa 160-182) proteins with high conservancy (~80% to 100%) and found common to both SARS-CoV and SARS-CoV-2 (Table 5). These findings are in agreement with that of Ju et al. (2020) who have isolated and characterized around 206 monoclonal antibodies specific to RBD of S protein derived from single B cells of eight SARS-CoV-2 infected individuals. Crystal structure analysis of the RBD-bound antibody revealed that the neutralization activity correlates with their competitive capacity with ACE2 for RBD binding. The findings are suggestive that anti-RBD antibodies are viral species-specific inhibitors without cross-reactivity to RBDs of SARS-CoV and MERS-CoV[31]. In line with our bioinformatic observations, Wang et al. (2020) have identified a human monoclonal antibody (47D11) that neutralizes SARS-CoV-2 and SARS-CoV through binding to RBDs without hampering ACE2 receptor interaction in cell culture model system[30]. Similarly, Amanat et al. (2020) have developed a serological assay using S protein expressed in insect and mammalian expression system and findings indicated that the plasma/serum samples derived from COVID-19 patients reacted strongly to both RBD and full-length S protein[32]. We found a linear B-cell epitope in the RBD of S protein (370-NSASFSTFKCYGVSPTKLNDLCFTNV-395, inset table of Fig. 4) which was recently shown to interact with the CR3022, a previously described antibody from the SARS patient and known to be neutralized by binding to RBD of S protein of SARS-CoV and SARS-CoV-2 without hindering S protein interaction with the cellular ACE2 receptor[28,29]. In a recent preprint study, Nanda et al. (2020) have developed a modified vaccinia Ankara (MVA) based-vaccines expressing either the full-length or secreted S1 domain of S glycoprotein of SARS-CoV-2. The findings revealed that, although both immunogens induce strong IgG antibodies to purified S protein, only the full-length S protein induced a strong neutralizing antibody response against the SARS-CoV-2[47]. In a similar MVA-based vaccine, the full-length S protein was found to induce a broader spectrum of neutralizing antibodies in comparison with fragmented S protein of SARS-CoV[48].

Of the epitopes identified in the RBD of S protein (Fig.4 inset table and Table 1c), the B cell peptide spanning the region (aa 403-RGDEVRQIAPGQTGKIADYNYKLPD-427, aa 437-NSNNLDSKVGGNYNYLYRLFRKSNL-461 and aa 483-VEGFNCYFPLQ-493) have recently been identified as

residues responsible for interaction between CB6 mAb (a human monoclonal neutralizing antibody isolated from a convalescent COVID-19 patient) and the RBD of S protein of SARS-CoV-2. A co-crystallized structure has revealed that CB6 recognizes an epitope that overlaps with ACE2-binding sites in the RBD of SARS-CoV-2 leading to disruption of the S-ACE2 receptor complex formation[49]. Intriguingly, most of the amino acid residues spanning the B-cell epitope (aa 331-NITNLCPFGEVFNATRFASVYAWNRK-356 and aa 437- NSNNLDSKVGGNYNYLYRLFRKSNL-461, Fig.4, inset table) identified in the S protein have recently been shown to interact with the cross-neutralizing mAb, S309 (previously isolated from SARS-CoV patient) in a ACE2 receptor-independent manner[50]. In a study, using the recombinant SARS-CoV-2 RBD antigen, a strong correlation between levels of RBD binding antibodies and SARS-CoV-2 neutralizing antibodies have been observed in patients on day 9 after the onset of symptoms[51]. Similar investigations carried out using samples collected from SARS-CoV-2 patients of USA, Europe and Hong Kong have reported that the full-length S protein and the RBD performed well for specific and sensitive antibody detection[32,42,41]. The findings of these works support the use of RBD or the full-length S protein in serological diagnostic assays.

In a preprint by Farrera-Soler et al. (2020), the three linear epitopes (655-672, 787-822 and 1147-1158) of S protein were most abundantly detected in all the SARS-CoV-2 positive samples tested, but each epitope is detected in >40% of the positive patients. The 655-672 epitope was the most detected in the positive samples and corresponds to a peptide that is not part of secondary structures. In our bioinformatics prediction we have found the same linear B-cell epitope being predicted by both the server used- BepiPred and Bcepred and a poor Agadir score (0.03) collaborating a lack of helical content as reported previously[52]. Similarly, these linear B-cell epitopes of S protein (aa 652- GAEHVNNSYECDIPIGAG-669, aa 660- YECDIPIGAGICASYQTQ-677, aa 780-EVFAQVKQIYKTPPIKDF-797, aa 798- GGFNFSQILPDPSKPSKR-815, aa 819-EDLLFNKVTLADAGFIKQ-836, aa 1149- KEELDKYFKNHTSPDVDL-1166, Table 1d) was found predicted as potent B-cell epitopes using ABCpred server thereby collaborating with the recent findings of detection of the above epitope in the sera of SARS-CoV-2 positive patients[52]. Poh and colleagues (2020) have identified and characterized neutralizing antibodies spanning the two immunodominant regions on S protein (aa 553-TESNKKFLPFQQFGRDIA-570, aa 809-PSKPSKRSFIEDLLFNKV-826) of SARS-CoV-2 that are recognised by convalescent sera of COVID-19 patients[53]. Using the ABCpred server, we have identified these B cell epitopes (aa 543-FNGLTGTGVLTESNKKFL-560, aa 559-FLPFQQFGRDIADTTDAV-576, aa 798-GGFNFSQILPDPSKPSKR-815, aa 819-EDLLFNKVTLADAGFIKQ-836, Table 1d) in the S protein of SARS-CoV-2. We further showed that these linear B-cell epitopes (inset table of Fig.13) are localized on the accessible region using modelled and cryo-EM solved (PDB: 6VSB) structure of the S protein of SARS-CoV-2 (Fig.13A and B).

Among those B cell peptides identified in S protein of SARS-CoV-2 (aa 590-CSFGGVSVITPGTNTSNQ-607, aa 607-QVAVLYQDVNCTEVPVAI-624, Table 1d) and SARS-CoV (aa 602-NCTDVSTAIHADQLTPAWRIYSTGNNVFQT-631, aa 1160- NIQKEIDRLNEVAKNLNESL-1179, aa 1179-LIDLQELGKYEQYIKWP-1195, data not show), of particular interest is the region of two tandem epitopes (S597-603 and S604-625) that showed a high level of serological reactivity and found to have the

capacity to induce a long-term B-cell memory response with antisera from convalescent SARS patients[54,55,56]. The linear B cell peptide corresponding to the region (aa 1236-CKFDEDDSEPVLKGVKLHYT-1255) of S protein is found conserved (~84%, Table 1b and 1c) and common to both SARS-CoV and SARS-CoV-2. This peptide was earlier shown as one of the immunodominant antigenic peptides in the S protein of SARS-CoV, supporting further exploration of this highly conserved C-terminal region among the S protein of SARS-CoV-2 will be important[55].

The nucleocapsid (N) protein is the most abundant and highly conserved structural protein across CoVs and the N protein of SARS-CoV-2 shares ~90% amino acid identity with N protein of SARS-CoV suggesting that previously described antibodies against the N protein of SARS-CoV would likely recognize and bind the N protein of SARS-CoV-2 as well. The earlier work by Leung et al. (2004) showed that neutralizing antibodies were detected early and with high specificity and predominantly against N protein in patients with SARS-CoV[5]. The B-cell epitope predicted and identified (aa 359-DAYKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDD-403, Table 1c) in the N protein of SARS-CoV was previously shown to have the highest affinity to form peptide-antibody complexes with SARS serum[57,58]. Notably, the same epitope was found localized on the surface of N protein of SARS-CoV-2 (366-TEPKKDKKKKADETQALPQRQKKQQTVTL-394, Table 1c and Fig. 3c), thus providing scope for future investigations and might contributes to the understanding of the immunogenicity and persistence of SARS-CoV-2 coronavirus. Therefore, N371 (371-KDKKKKTDEAQPLPQRQKKQ-390) and N385 (N385-QRQKKQPTVTLLPAADMDDFSRQ-407) located at the C-terminal region of the N protein of SARS-CoV appear to be highly immunogenic and may be useful for serologic assays for SARS-CoV-2 given the conservancy of the previously characterized epitope of N protein[57,59,58]. Similarly, synthetic peptide M132-161 derived from the C-terminal epitope of M protein (aa 131-PLMESELVIGAVIIRGHLRM-150, Table 1c) was found highly reactive using the convalescent-sera of SARS patients, suggesting the potential application for serologic diagnosis of SARS[60].

In this work we have identified and listed the top CD8+ and CD4+ T cell epitopes based on antigenicity, conservancy, non-allergenicity and positive score for IFN-gamma, of which 41 MHC class I epitopes were predicted for the S protein, 19 each for the N and M proteins, and 10 for the E protein of SARS-CoV-2 (Table 3c), while we found 26 MHC class II epitopes predicted for the S protein, 8 epitopes for the N protein, 6 and 2 epitopes for the M and E proteins, respectively (Table 4c). A recent finding showed that the S, M and N proteins of SARS-CoV-2 each accounted for 11-27% of the total CD4+ response, while for CD8+ T cells, S and M were recognized including some non-structural proteins[40]. In a recent preprint study, Takuya et al. (2020) have shown that SARS-CoV-2 elicits robust T-cell mediated immune responses in both symptomatic and asymptomatic individuals, even in the absence of a positive antibody test[61]. It is generally expected that B- and T-cell epitopes can be predicted from both structural and non-structural proteins and in this study we have considered the four structural proteins (S, E, M and N) of SARS-CoV-2 based on the existing knowledge of innate and adaptive immune responses to SARS-CoV. Analysis of T-cell epitopes has revealed that peptides corresponding to residues 1-30, 86-100, 306-320, and 351-365 (Table S3 and S4) of N protein of SARS-CoV (isolate BJ01) have been reported earlier as

immunodominant T-cell epitopes. It was previously observed that peptides corresponding to residues aa 336-350 were capable of stimulating IFN-γ production in T-cell cultures derived from peripheral blood mononuclear cells and the peptide spanning the region (aa 81-PDDQIGYYRRATRRV-95) have confirmed that YYRRATRRV is a very potent functional CD8+ T-cell epitope of N protein[61]. Intriguingly, this functional CD8+ T-cell epitope (aa 86-GYYRRATRR-94) is found to bind with two alleles (HLA-A*31:01, HLA-A*33:0) is conserved to both N protein of SARS-CoV-2 and SARS-CoV (~85%, Table S3). Additionally, the epitope corresponding to aa 350-NVILLNKHIDAYKTF-364, aa 305- IAQFAPSASAFFGMS-319, aa 352-ILLNKHIDA-360 were found common to N protein of both SARS-CoV-2 and SARS-CoV (~80% conservancy, Table S3 and Table S4). Some of the present findings of B- and T-cell epitope predicted using computational algorithms corroborate with the experimentally verified immunogenic epitopes of SARS-CoV and SARS-CoV-2. The potential for cross-protection also exists as the selected proteins and predicted epitopes used in generating the chimeric multi-epitopes exhibited considerable conservation across structural proteins of SARS-CoV and SARS-CoV-2.

Designing synthetic peptides for use as vaccines to induce both humoral and cell-mediated immunity requires an understanding of the nature of T- and B-cell epitopes. Ideally, vaccines for inducing humoral immunity should include peptides that form immunodominant B-cell epitopes. A successful vaccine must also generate a population of memory T cells; therefore the peptide should include immunodominant T-cell epitopes. Bioinformatic methods used for prediction of B-cell and MHC binding peptides are of very high quality and predict epitopes with high accuracy. However, it remains to be investigated whether a mega-pool of B- and T-cell epitopes identified in this work can stimulate T-cells by measuring the amount of released cytokines both *in vitro* and *in vivo* experimental system, thus providing a platform for future investigations.

Materials And Methods

Structural protein sequences of SARS-CoV-2 and SARS-CoV

We have used NCBI GenBank SARS-CoV-2 (isolate WIV02, accession number MN996527.1[1]) and SARS-CoV (isolate BJ01, accession number AY278488.2[7]) to download coding sequences for structural proteins such as spike (S), nucleocapsid (N), membrane (M) and envelope (E) for prediction of B- and T-cell epitopes using bioinformatics methods. Sequence alignment of the RBDs of the S protein of SARS-CoV-2 (QHR63250.2) and SARS-CoV (QHR63300.2) was performed using Clustal omega.

B-cell epitope prediction for SARS-CoV-2 and SARS-CoV

BepiPred-2.0[8] and Bcepred[9] servers were employed to predict linear B cell epitopes. BepiPred-2.0 is a web-based server that utilizes a random forest algorithm trained on epitopes annotated from antibody-antigen protein structures. Bcepred predicts B-cell epitopes based on the physico-chemical properties such as hydrophilicity, flexibility, accessibility, polarity, exposed surface, turns and antigenic propensity. EPSVR, Discotope, CBTOPE and Ellipro servers were used to predict conformational B-cell epitopes. EPSVR uses a

Support Vector Regression (SVR) method to predict antigenic B cell epitopes[10]. DiscoTope server predicts discontinuous B cell epitopes from protein-3D structures. The method utilizes calculation of surface accessibility and a novel epitope propensity amino acid score[11]. CBTOPE server predicts conformational B-cell epitope with an accuracy of more than 85% using antigen primary sequence in the absence of any homology with the known structures[12]. ElliPro[13] predicts linear and discontinuous antibody epitopes based on the protein structure and homology-based model of the amino acid sequence. The helical behaviour of predicted monomeric peptides was computed using Agadir server[14,15].

T-cell epitope prediction for SARS-CoV-2 and SARS-CoV

TepiTool is an interactive and easy to use tool to predict potential peptides binding to MHC class I and class II molecules. In Tepitool prediction we have used a panel of 27 most frequent alleles that are distributed among populations worldwide[16]. Proteasomal cleavage/TAP transport/MHC class I combined predictor[17] was used to identify MHC class I epitope based on proteasomal processing and TAP transport in the cell whereas MHC-NP predicts T cell peptides naturally processed by the MHC[18]. We selected the top 2 % ranked epitopes as strong binders with 9-mer peptide length. NetMHCIIpan 4.0 server predicts peptide binding to any MHC II molecule of the known sequence using artificial neural networks[19]. We applied the threshold of top 2% ranked epitopes based on the prediction score as strong binders to HLA class II molecules with 15-mer peptide length. All 9- and 10-mer peptides were predicted for their binding affinity to 27 MHC class I and 23 MHC class II molecules which account for 97% of HLA A and B allelic variants in most ethnicities[20]. We used the Immune Epitope Database and Analysis Resource (IEDB) analysis resource tool to analyse population coverage of predicted MHC class I and class II epitopes using allele frequency database of 115 countries and 21 different ethnicities grouped into 16 different geographical areas[21]. The ability of the predicted T-cell epitopes to induce interferon-gamma (IFN-γ) response was assessed using the IFNepitope server[22]. IFNepitope uses the algorithm based on three models such as motif-based, SVM based, and hybrid approach for prediction of IFN-γ inducing epitopes.

Antigenicity, allergenicity, and conservancy of predicted B- and T-cell epitope

The antigenicity and allergenicity of epitopes were predicted using VaxiJen v2.0 and AllerTop online servers. VaxiJen predict protective antigens and subunit vaccines based on an alignment-independent method[23]. AllerTop is a server for *in silico* prediction of allergens in a given protein or antigens[24]. We used the IEDB analysis resource epitope conservancy analysis tool to computes the degree of conservancy of an epitope within a given protein sequence[25].

Immune simulation

To determine the immunogenicity and immune response profile of the multi-epitope protein constructs containing B- and T-cell epitopes, *in-silico* immune simulation was carried out using the C-ImmSim server. C-ImmSim server uses a position-specific scoring matrix and machine learning techniques for prediction of epitope and immune interactions. The server simulates three components of immune system found in

mammals[26]: (i) the bone marrow, where hematopoietic stem cells are simulated and produce new lymphoid and myeloid cells; (ii) the thymus, where naive T-cells are selected to avoid auto immunity; and (iii) a tertiary lymphatic organ, such as a lymph node. All simulation parameters were set at default with time steps set at 1, 84, and 168 (each time step is 8 hours and time step 1 is injection at time = 0). Therefore, three injections were given at four weeks apart without lipopolysaccharide (LPS). The Simpson index, D (a measure of diversity) was interpreted from the plot.

Conclusion

One of the limitations of the present work is the lack of information on B- and T-cell epitopes of non-structural proteins of SARS-CoV-2. In summary, the bioinformatics analysis of the structural proteins has led to prediction and identification of a pool of B- and T-cell epitopes, which are likely to facilitate researchers to select appropriate epitope or region of proteins (especially the S and N proteins) in an effort to develop a novel drugs, vaccines, and serological assays for the detection, treatment, and management of SARS-CoV-2.

Declarations

Acknowledgments

I as corresponding author had full access to all the data in this study and take responsibility for the integrity of the data and the accuracy of the data analysis and had final responsibility for the decision to submit for publication. I have not been paid to write this article by any pharmaceutical company or other agency.

Author contributions

YDD, HBG, SK, CD and AD contributed equally as first authors. YDD and HBG prepared figures and tables. NDN conceived and designed the study and performed literature search. NDN supervised data collection and analysis. YDD, HBG, SK, CD and AD collected data. YDD, HBG and NDN analysed and interpreted the data. HBG have helped in reference formatting. AD and CC have helped to generate heat and surface localization of predicted epitopes on modelled structures. VB and SD have helped to prepare figures and table and manuscript editing. LJ has edited the manuscript. AK, SSS, SKR, RCD, RD and MM have helped in analysis of data. Ch. SS and PPB are a part of clinical collaborators. YDD and NDN wrote the manuscript. All authors contributed to reviewing and editing the manuscript.

Declaration of competing interests

We declare no competing interests. This part of work at present has no funding support. However, NDN has submitted project on ‘development of multi-plex RT-PCR and serological assay for SARS-CoV-2’ to DST, SERB, Govt. of India for financial support.

References

1. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
2. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
3. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.).* **367**, 1260–1263 (2020).
4. Sui, J. *et al.* Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2536–2541 (2004).
5. Leung, D. T. M. *et al.* Antibody Response of Patients with Severe Acute Respiratory Syndrome (SARS) Targets the Viral Nucleocapsid. *J. Infect. Dis.* **190**, 379–386 (2004).
6. Tripp, R. A. *et al.* Monoclonal antibodies to SARS-associated coronavirus (SARS-CoV): Identification of neutralizing and antibodies reactive to S, N, M and e viral proteins. *J. Virol. Methods* **128**, 21–28 (2005).
7. Wu, Q. *et al.* The E protein is a multifunctional membrane protein of SARS-CoV. *Genomics, proteomics Bioinforma. / Beijing Genomics Inst.* **1**, 131–144 (2003).
8. Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **45**, W24–W29 (2017).
9. Saha, S. & Raghava, G. P. S. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **3239**, 197–204 (2004).
10. Liang, S. *et al.* EPSVR and EPMeta: Prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* **11**, (2010).
11. Kringelum, J. V., Lundegaard, C., Lund, O. & Nielsen, M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. *PLoS Comput. Biol.* **8**, (2012).
12. Ansari, H. R. & Raghava, G. P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res.* **6**, 6 (2010).
13. Ponomarenko, J. *et al.* ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* **9**, 1–8 (2008).
14. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. II†. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**, 275–296 (1995).
15. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters: III. Temperature and pH dependence. *J. Mol. Biol.* **245**, 297–308 (1995).
16. Paul, S., Sidney, J., Sette, A. & Peters, B. TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Curr. Protoc. Immunol.* **2016**, 18.19.1-18.19.24 (2016).

17. Tenzer, S. *et al.* Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell. Mol. Life Sci.* **62**, 1025–1037 (2005).
18. Giguère, S. *et al.* MHC-NP: Predicting peptides naturally processed by the MHC. *J. Immunol. Methods* **400**–**401**, 30–36 (2013).
19. Reynisson, B. *et al.* Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J. Proteome Res.* **19**, 2304–2315 (2020).
20. Sette, A. & Sidney, J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* **50**, 201–212 (1999).
21. Bui, H. H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* **7**, 1–5 (2006).
22. Dhanda, S. K., Vir, P. & Raghava, G. P. S. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct* **8**, 1–15 (2013).
23. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 1–7 (2007).
24. Dimitrov, I., Flower, D. R. & Doytchinova, I. AllerTOP - a server for in silico prediction of allergens. *BMC Bioinformatics* **14**, S4 (2013).
25. Bui, H. H., Sidney, J., Li, W., Fusseder, N. & Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* **8**, 1–6 (2007).
26. Rapin, N., Lund, O., Bernaschi, M. & Castiglione, F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* **5**, (2010).
27. Saha S, R. G. Prediction of Continuous B-Cell Epitopes in an Antigen Using Recurrent Neural Network. *Proteins Struct. Funct. Bioinforma.* **65**, 40–48 (2006).
28. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science (80-.).* **368**, 630–633 (2020).
29. Tian, X. *et al.* Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect.* **9**, 382–385 (2020).
30. Wang, C. *et al.* A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **11**, 1–6 (2020).
31. Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature.* https://doi.org/10.1038/s41586-020-2380-z (2020).
32. Amanat, F. *et al.* A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* https://doi.org/10.1038/s41591-020-0913-5 (2020).
33. Gui, M. *et al.* Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017).

34. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
35. Li, C. K. *et al.* T Cell Responses to Whole SARS Coronavirus in Humans. *J. Immunol.* **181**, 5490–5500 (2008).
36. Guo, X. *et al.* Long-Term Persistence of IgG Antibodies in SARS-CoV Infected Healthcare Workers. *medRxiv* 2020.02.12.20021386; https://doi.org/10.1101/2020.02.12.20021386 (2020).
37. Grifoni, A. *et al.* A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671-680.e2 (2020).
38. Tsao, Y. P. *et al.* HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. *Biochem. Biophys. Res. Commun.* **344**, 63–71 (2006).
39. Kiyotani, K., Toyoshima, Y., Nemoto, K. & Nakamura, Y. Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. *J. Hum. Genet.* **65**, 569–575 (2020).
40. Grifoni, A. *et al.* Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**, 1489-1501.e15 (2020).
41. Perera, R. A. P. M. *et al.* Serological assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), March 2020. *Eurosurveillance* **25**, (2020).
42. Okba, N. M. A. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2-Specific Antibody Responses in Coronavirus Disease 2019 Patients. *Emerg. Infect. Dis.* **26**, (2020).
43. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
44. Gasteiger, E. *et al.* The Proteomics Protocols Handbook. *Proteomics Protoc. Handb.* 571–608 doi:10.1385/1592598900 (2005).
45. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
46. Moudgil, K. D. *et al.* Antigen processing and T cell repertoires as crucial aleatory features in induction of autoimmunity. *J. Autoimmun.* **9**, 227–234 (1996).
47. Routhu NK, Gangadhara S, Cheedarla N, Shiferaw A, Rahman SA, Sahoo A, Shi PY, Menachery VD, Floyd K, Fischinger S, Atyeo C, Alter G, Suthar MS, A. R. Modified Vaccinia Ankara Based SARS-CoV-2 Vaccine Expressing Full-Length Spike Induces Strong Neutralizing Antibody Response. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.27.175166 (2020).
48. Chen, Z. *et al.* Recombinant Modified Vaccinia Virus Ankara Expressing the Spike Glycoprotein of Severe Acute Respiratory Syndrome Coronavirus Induces Protective Neutralizing Antibodies Primarily Targeting the Receptor Binding Region. *J. Virol.* **79**, 2678–2688 (2005).
49. Shi, R. *et al.* A human neutralizing antibody targets the receptor binding site of SARS-CoV-2. *Nature.* https://doi.org/10.1038/s41586-020-2381-y (2020).
50. Pinto, D. *et al.* Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature.* https://doi.org/10.1038/s41586-020-2349-y (2020).

51. Premkumar, L. *et al.* The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci. Immunol.* **5**, 1–10 (2020).
52. Farrera-soler, L. *et al.* Identification of immunodominant linear epitopes from SARS- CoV-2 patient plasma. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.15.20131391 (2020).
53. Poh, C. M. *et al.* Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.* **11**, (2020).
54. Hu, H. *et al.* Screening and identification of linear B-cell epitopes and entry-blocking peptide of severe acute respiratory syndrome (SARS)-associated coronavirus using synthetic overlapping peptide library. *J. Comb. Chem.* **7**, 648–656 (2005).
55. Chow, S. C. S. *et al.* Specific epitopes of the structural and hypothetical proteins elicit variable humoral responses in SARS patients. *J. Clin. Pathol.* **59**, 468–476 (2006).
56. Wang, Q. *et al.* Immunodominant SARS coronavirus epitopes in humans elicited both enhancing and neutralizing effects on infection in non-human primates. *ACS Infect. Dis.* **2**, 361–376 (2016).
57. Wang, J. *et al.* Assessment of Immunoreactive Synthetic Peptides from the Structural Proteins of Severe Acute Respiratory Syndrome Coronavirus. *Clin. Chem.* **49**, 1989–1996 (2003).
58. Liu, S. J. *et al.* Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates. *Vaccine* **24**, 3100–3108 (2006).
59. He, Y. *et al.* Mapping of antigenic sites on the nucleocapsid protein of the severe acute respiratory syndrome coronavirus. *J. Clin. Microbiol.* **42**, 5309–5314 (2004).
60. He, Y., Zhou, Y., Siddiqui, P., Niu, J. & Jiang, S. Identification of immunodominant epitopes on the membrane protein of the severe acute respiratory syndrome-associated coronavirus. *J. Clin. Microbiol.* **43**, 3718–3726 (2005).
61. Sekine, T., Perez-potti, A., Rivera-ballesteros, O. & Strålin, K. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.29.174888 (2020).

Tables

Due to technical limitations, full-text HTML conversion of the tables could not be completed. However, the tables can be downloaded and accessed as a PDF in the supplementary files section.

Figures

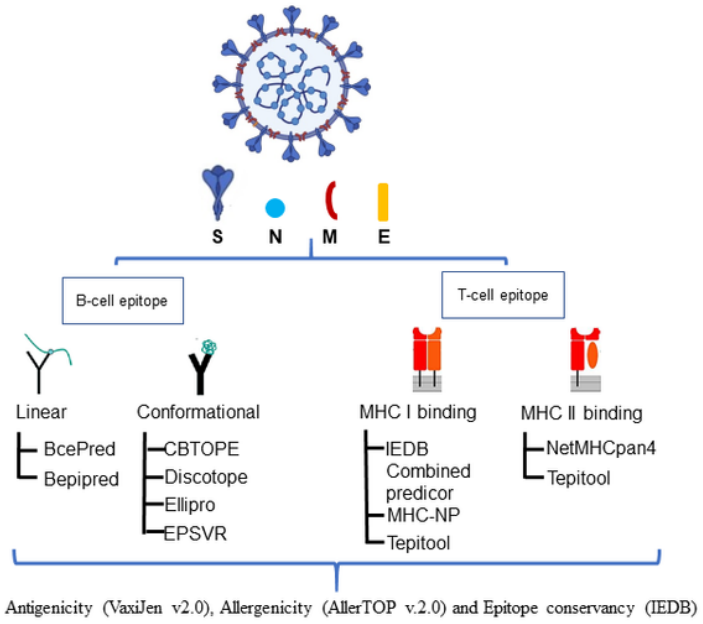


Figure 1

Schematic representation of an in-silico workflow utilized for prediction of potential B- and T-cell epitopes of structural proteins {spike (S), envelope (E), membrane (M) and nucleocapsid (N)}of SARS-CoV-2 and SARS-CoV.

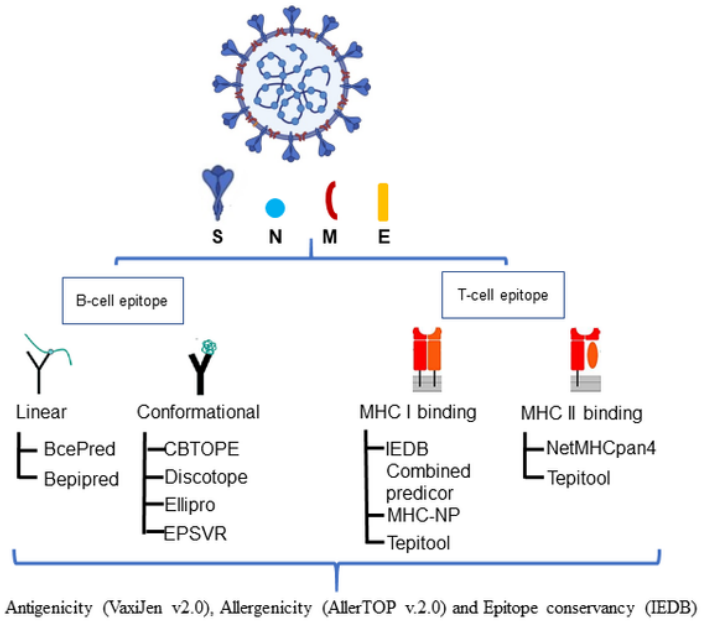


Figure 1

Schematic representation of an in-silico workflow utilized for prediction of potential B- and T-cell epitopes of structural proteins {spike (S), envelope (E), membrane (M) and nucleocapsid (N)}of SARS-CoV-2 and SARS-CoV.

Fig.2

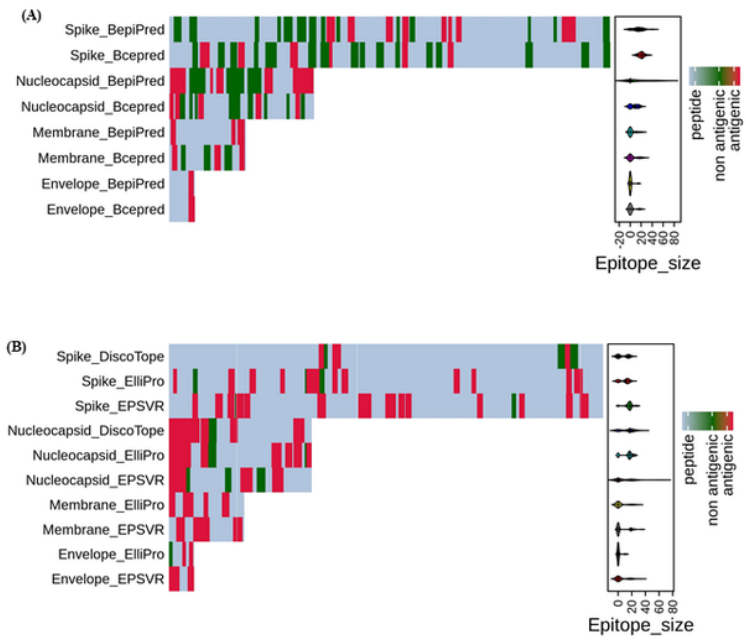


Figure 2

Summary of SARS-CoV-2-derived B-cell epitopes. Heat map showing the distribution of (A) linear (continuous) and (B) conformational (discontinuous) B-cell epitopes across the protein sequences of spike (1273 aa), nucleocapsid (419 aa), membrane (222 aa) and envelope (75 aa) proteins of SARS-CoV-2. Red color represent likely antigenic epitopes that were predicted using the methods described in figure 1.

Fig.2

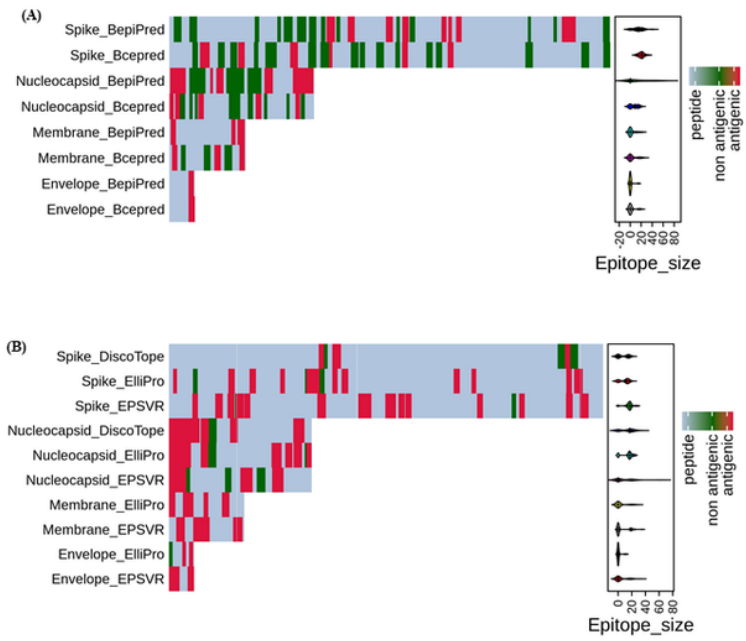
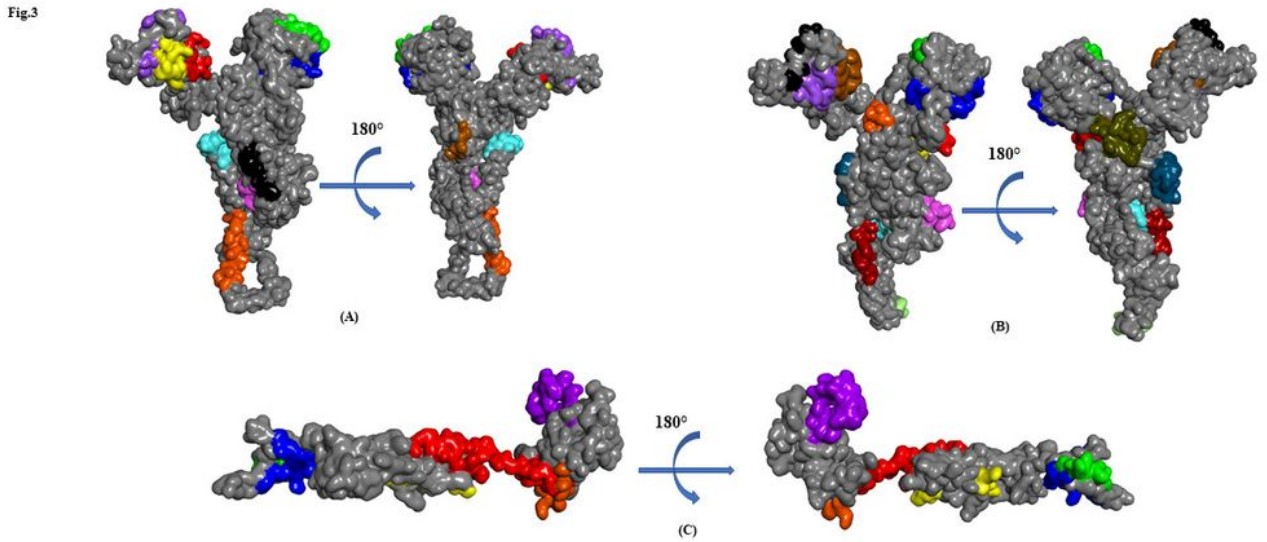


Figure 2

Summary of SARS-CoV-2-derived B-cell epitopes. Heat map showing the distribution of (A) linear (continuous) and (B) conformational (discontinuous) B-cell epitopes across the protein sequences of spike (1273 aa), nucleocapsid (419 aa), membrane (222 aa) and envelope (75 aa) proteins of SARS-CoV-2. Red color represent likely antigenic epitopes that were predicted using the methods described in figure 1.



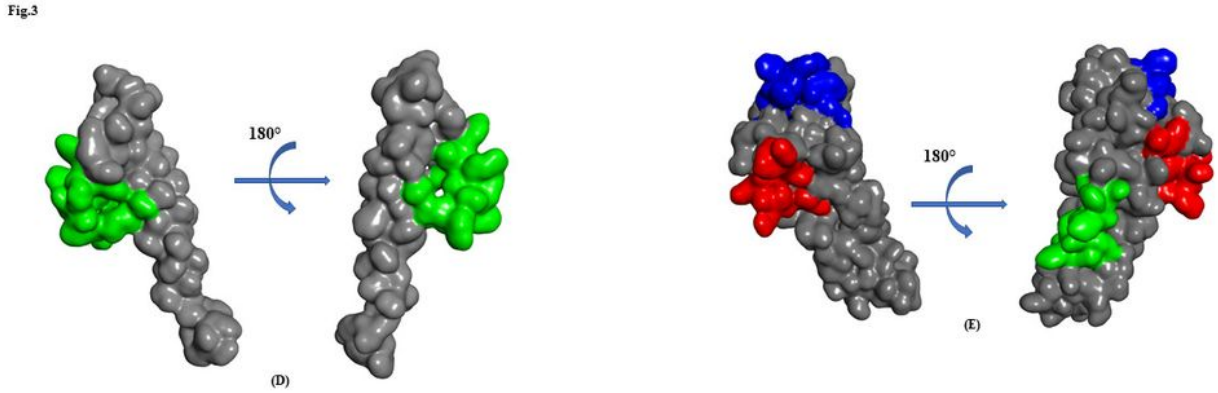
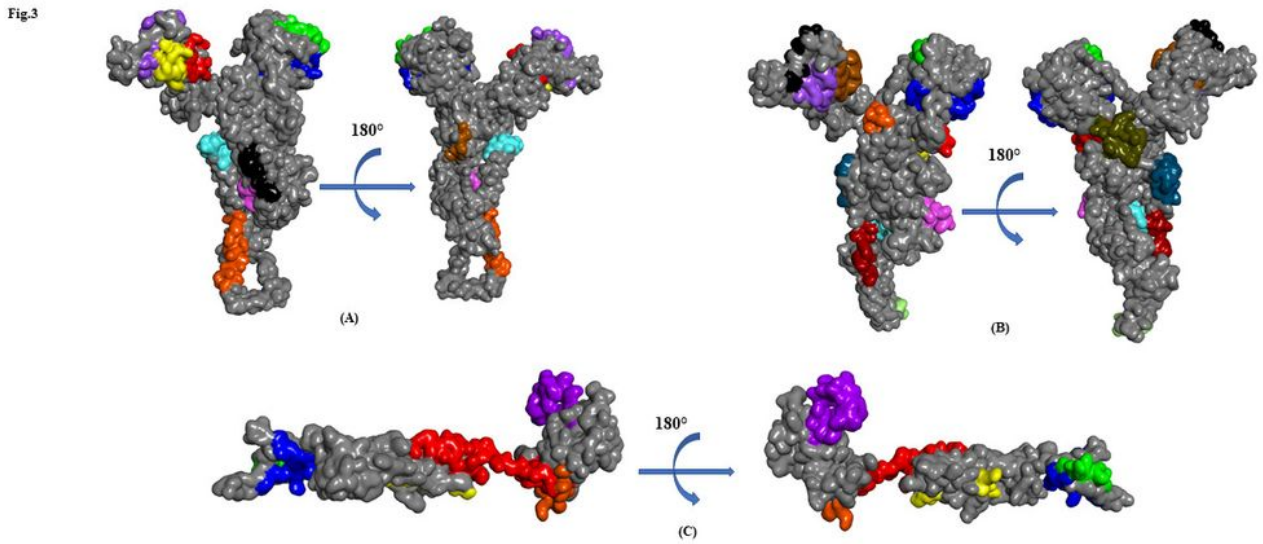


Figure 3

Localization of the linear B-cell epitopes on the modelled structure of SARS-CoV-2 proteins. (A) Spike {Green: GVYYHKNNKSWMESEFRVY; Blue: MDLEGKQGNFKNLREFVFKN; Red:NSASFSTFKCYGVSPTKLNDLCFTNV; Yellow: RGDEVRQIAPGQTGKIADYNYKLPD; Purple: NSNNLDSKVGGNYNYLYRLFRKSNL; Brown: AEHVNNSYECDIPI; Black: IAVEQDKNTQEVFAQVKQ; Pink: KMSECVLGQSKRVDFCGKGYHL; orange: KNLNESLIDLQELGKYEQYIK; Cyan:

CCKFDEDDSEPVLKGVKLHYT} and (B) Spike protein of SARS-CoV (Green: NVVRGWVFGSTMNNKSQSV; Blue: FSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPID and DVSEKSGNFKHLREFVFKNKDGFLY; Red: AELKCSVKSFEIDKGIYQTSNFR; Yellow: KGDDVRQIAPGQTGVIADYNYKLPD; Purple: DATSTGNYNYKYRYLRHGKLRPFERDI; Black: QILPDPLKPTKRSFIED; Pink: KMSECVLGQSKRVDFCGKGYHL; Orange: LIDLQELGKYEQYIKWP; Cyan: CKFDEDDSEPVLKGVKLHYT) using modelled structure. (C) Nucleocapsid. (Green: SDNGPQNQRNAP; Blue: RSGARSKQRRPQGLP; Yellow: AEGSRGGSQASSRSSSRS; Red: LNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATK; Orange: FGRRGPEQTQ; Purple: TEPKKDKKKKADETQALPQRQKKQQTVTL. (D) Envelope (Green: YVYSRVKNLNSSRVPDLLV) and (E) Membrane (Green: LQFAYANRNRFLYI; Blue: RLFARTRSMWSFNPETNIL; Red: DIKDLPKEITVATSRTLSYYKLG).



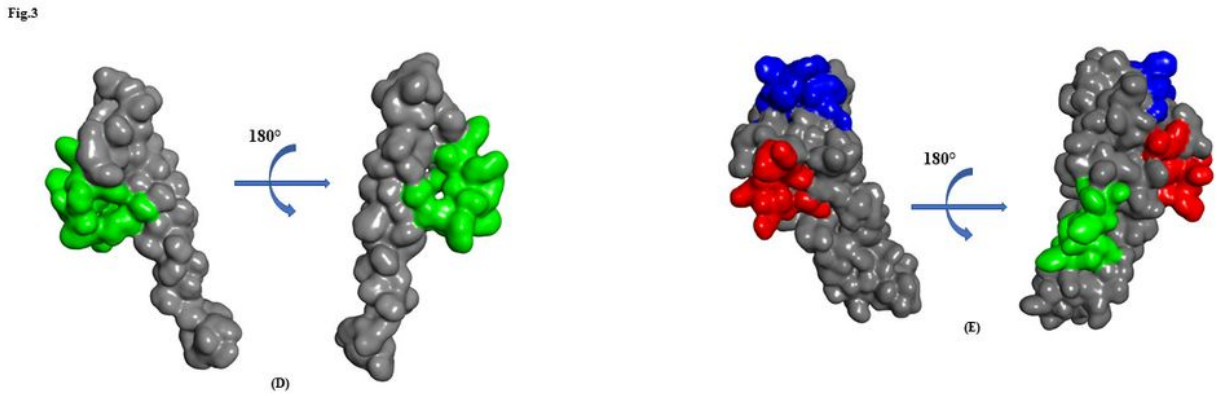


Figure 3

Localization of the linear B-cell epitopes on the modelled structure of SARS-CoV-2 proteins. (A) Spike {Green: GVYYHKNNKSWMESEFRVY; Blue: MDLEGKQGNFKNLREFVFKN; Red:NSASFSTFKCYGVSPTKLNDLCFTNV; Yellow: RGDEVRQIAPGQTGKIADYNYKLPD; Purple: NSNNLDSKVGGNYNYLYRLFRKSNL; Brown: AEHVNNSYECDIPI; Black: IAVEQDKNTQEVFAQVKQ; Pink: KMSECVLGQSKRVDFCGKGYHL; orange: KNLNESLIDLQELGKYEQYIK; Cyan:

CCKFDEDDSEPVLKGVKLHYT} and (B) Spike protein of SARS-CoV (Green: NVVRGWVFGSTMNNKSQSV; Blue: FSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPID and DVSEKSGNFKHLREFVFKNKDGFLY; Red: AELKCSVKSFEIDKGIYQTSNFR; Yellow: KGDDVRQIAPGQTGVIADYNYKLPD; Purple: DATSTGNYNYKYRYLRHGKLRPFERDI; Black: QILPDPLKPTKRSFIED; Pink: KMSECVLGQSKRVDFCGKGYHL; Orange: LIDLQELGKYEQYIKWP; Cyan: CKFDEDDSEPVLKGVKLHYT) using modelled structure. (C) Nucleocapsid. (Green: SDNGPQNQRNAP; Blue: RSGARSKQRRPQGLP; Yellow: AEGSRGGSQASSRSSSRS; Red: LNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATK; Orange: FGRRGPEQTQ; Purple: TEPKKDKKKKADETQALPQRQKKQQTVTL. (D) Envelope (Green: YVYSRVKNLNSSRVPDLLV) and (E) Membrane (Green: LQFAYANRNRFLYI; Blue: RLFARTRSMWSFNPETNIL; Red: DIKDLPKEITVATSRTLSYYKLG).

Fig.4

```
SARS-CoV      RBD   306 RVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFK 365
SARS-CoV-2    RBD   319 RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFK 378

SARS-CoV      RBD   366 CYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT 425
SARS-CoV-2    RBD   379 CYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNS 438

SARS-CoV      RBD   426 RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTP-PALNCYWPLNDYGFY 484
SARS-CoV-2    RBD   439 NNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQ 498

SARS-CoV      RBD   485 TTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNF 529
SARS-CoV-2    RBD   499 PTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFN- 542
```

Name	Predicted epitope of SARS-CoV-2	Length	Server	Antigenicity by VexiJen (T=0.4)	Allergeicity by AllerTope v. 2.0	Agadir Score	Conservancy analysis by IEDB tool	
							SARS-CoV-2	SARS-CoV
Linear B-cell epitopes	313-YQTSNFRVQP-322	10	1	Antigen (1.1866)	Non-allergen	0.04	100.00% (175/175)	0.31% (1/326)
	331-NITNLCPFGEVFNATRFASVYAWNRK-356	26	1 & 2	Antigen (0.5525)	Allergen	0.55	100.00% (175/175)	0.00% (0/326)
	370-NSASFSTFKCYGVSPTKLNDLCFTNV-395	26	1	Antigen (1.3609)	Non-allergen	0.39	100.00% (175/175)	0.00% (0/326)
	403-RGDEVRQIAPGQTGKIADYNYKLPD-427	25	1 & 2	Antigen (1.0356)	Non-allergen	0.45	100.00% (175/175)	0.00% (0/326)
	437-NSNNLDSKVGGNYNYLYRLFRKSNL-461	25	1 & 2	Antigen (0.4015)	Non-allergen	4.42	100.00% (175/175)	0.00% (0/326)
	483-VEGFNCYFPLQ-493	11	1	Antigen (0.5612)	Allergen	0.16	100.00% (175/175)	0.00% (0/326)
	525-CGPKKSTNLVKNKCVNFNFNGL-546	22	1 & 2	Antigen (0.7688)	Allergen	0.28	100.00% (175/175)	0.31% (1/326)
Conformational B-cell epitopes (6VSB)	K304,S305,F306,T307,V308,E309,K310,G311,I312,Y313,Q314,T315,S316,N317,F318,R319,V320,Q321,P322,T323,E324,S325	22	1	Antigen (0.5575)	Non-allergen	0.14	100.00% (175/175)	0.00% (0/326)
	L390,C391,F392,T393,N394,V395,Y396,A397,D398,S399,F400,V401,I402,R403	14	3	Antigen (0.5206)	Non-allergen	0.07	100.00% (175/175)	0.00% (0/326)
	G496,F497,Q498,P499,T500	5	2	Antigen (0.4499)	Non-allergen	0.00	100.00% (175/175)	0.00% (0/326)
	P507,Y508,R509,V510,V511,V512,L513,S514,F515	9	1 & 3	Antigen (1.0281)	Non-allergen	0.06	100.00% (175/175)	85.89% (280/326)
	S530,T531,N532,L533,V534,K535,N536,K537,N544,T553,E554,S555,N556,K557,F559,L560,P561,F562,Q563,Q564,A575,V576,R577,D578,P579,Q580,T581,L582,E583,I584,L585	31	3	Antigen (0.4327)	Non-allergen	0.13	100.00% (175/175)	0.31% (1/326)

Figure 4

Sequence alignment of RBDs of SARS-CoV-2 and SARS-CoV spike (S) proteins. GenBank accession numbers are QHR63250.2 (SARS-CoV-2) and AAP30030.1 (SARS-CoV BJ01). ACE2-binding residues are colored magenta. CR3022 epitope residues are colored bright green. The alignment was performed using Clustal Omega. Inset table shows the likely antigenic linear and conformational B-cell epitopes identified and predicted in the RBD of spike protein of SARS-CoV-2.

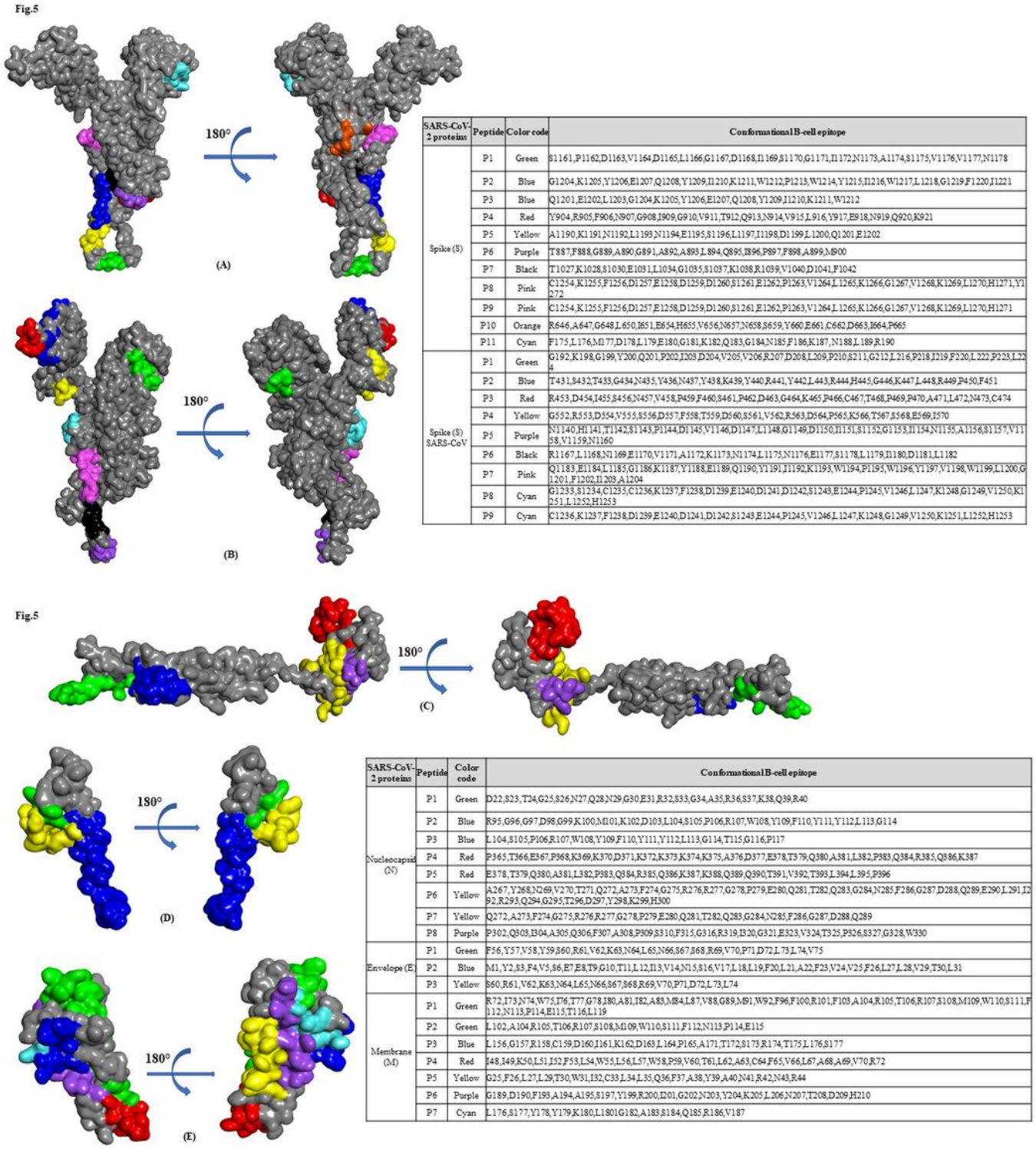
Fig.4

```
SARS-CoV     RBD  306 RVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFK 365
SARS-CoV-2   RBD  319 RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFK 378

SARS-CoV     RBD  366 CYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT 425
SARS-CoV-2   RBD  379 CYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNS 438

SARS-CoV     RBD  426 RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTP-PALNCYWPLNDYGFY 484
SARS-CoV-2   RBD  439 NNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQ 498

SARS-CoV     RBD  485 TTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNF 529
SARS-CoV-2   RBD  499 PTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFN- 542
```

Name	Predicted epitope of SARS-CoV-2	Length	Server	Antigenicity by VexiJen (T=0.4)	Allergeicity by AllerTopev. 2.0	Agadir Score	Conservancy analysis by IEDB tool	
							SARS-CoV-2	SARS-CoV
Linear B-cell epitopes	313-YQTSNFRVQP-322	10	1	Antigen (1.1866)	Non-allergen	0.04	100.00% (175/175)	0.31% (1/326)
	331-NITNLCPFGEVFNATRFASVYAWNRK-356	26	1 & 2	Antigen (0.5525)	Allergen	0.55	100.00% (175/175)	0.00% (0/326)
	370-NSASFSTFKCYGVSPTKLNDLCFTNV-395	26	1	Antigen (1.3609)	Non-allergen	0.39	100.00% (175/175)	0.00% (0/326)
	403-RGDEVRQIAPGQTGKIADYNYKLPD-427	25	1 & 2	Antigen (1.0356)	Non-allergen	0.45	100.00% (175/175)	0.00% (0/326)
	437-NSNNLDSKVGGNYNYLYRLFRKSNL-461	25	1 & 2	Antigen (0.4015)	Non-allergen	4.42	100.00% (175/175)	0.00% (0/326)
	483-VEGFNCYFPLQ-493	11	1	Antigen (0.5612)	Allergen	0.16	100.00% (175/175)	0.00% (0/326)
	525-CGPKKSTNLVKNKCVNFNFNGL-546	22	1 & 2	Antigen (0.7688)	Allergen	0.28	100.00% (175/175)	0.31% (1/326)
Conformational B-cell epitopes (6VSB)	K304,S305,F306,T307,V308,E309,K310,G311,I312,Y313,Q314,T315,S316,N317,F318,R319,V320,Q321,P322,T323,E324,S325	22	1	Antigen (0.5575)	Non-allergen	0.14	100.00% (175/175)	0.00% (0/326)
	L390,C391,F392,T393,N394,V395,Y396,A397,D398,S399,F400,V401,I402,R403	14	3	Antigen (0.5206)	Non-allergen	0.07	100.00% (175/175)	0.00% (0/326)
	G496,F497,Q498,P499,T500	5	2	Antigen (0.4499)	Non-allergen	0.00	100.00% (175/175)	0.00% (0/326)
	P507,Y508,R509,V510,V511,V512,L513,S514,F515	9	1 & 3	Antigen (1.0281)	Non-allergen	0.06	100.00% (175/175)	85.89% (280/326)
	S530,T531,N532,L533,V534,K535,N536,K537,N544,T553,E554,S555,N556,K557,F559,L560,P561,F562,Q563,Q564,A575,V576,R577,D578,P579,Q580,T581,L582,E583,I584,L585	31	3	Antigen (0.4327)	Non-allergen	0.13	100.00% (175/175)	0.31% (1/326)

Figure 4

Sequence alignment of RBDs of SARS-CoV-2 and SARS-CoV spike (S) proteins. GenBank accession numbers are QHR63250.2 (SARS-CoV-2) and AAP30030.1 (SARS-CoV BJ01). ACE2-binding residues are colored magenta. CR3022 epitope residues are colored bright green. The alignment was performed using Clustal Omega. Inset table shows the likely antigenic linear and conformational B-cell epitopes identified and predicted in the RBD of spike protein of SARS-CoV-2.

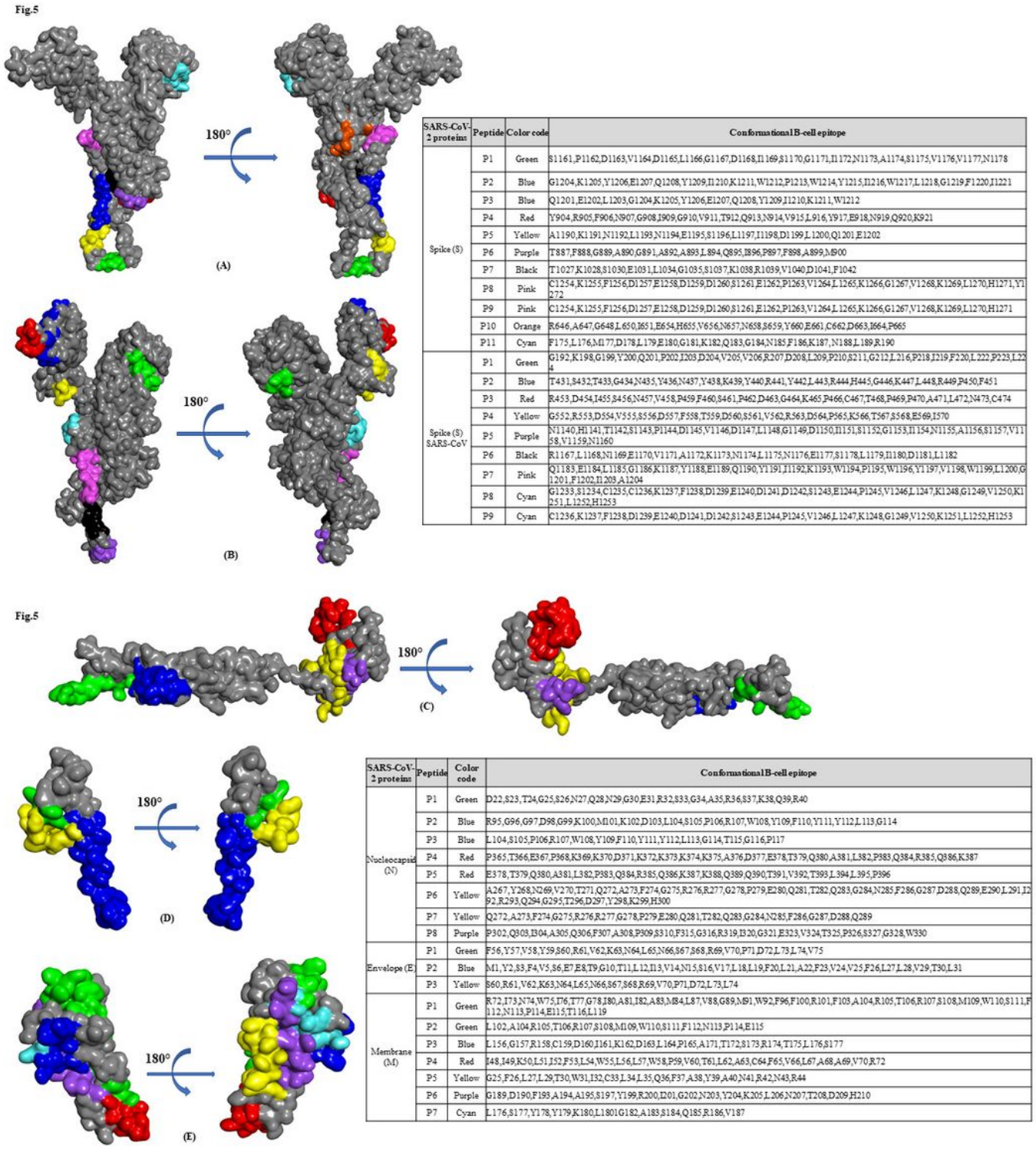


SARS-CoV-2 proteins	Peptide	Color code	Conformational B-cell epitope
Spike (S)	P1	Green	S1161,P1162,D1163,V1164,D1165,L1166,G1167,D1168,I1169,S1170,G1171,I1172,N1173,A1174,S1175,V1176,V1177,N1178
	P2	Blue	G1204,K1205,Y1206,E1207,Q1208,Y1209,I1210,K1211,W1212,P1213,W1214,Y1215,I1216,W1217,L1218,G1219,F1220,I1221
	P3	Blue	Q1201,E1202,L1203,G1204,K1205,Y1206,E1207,Q1208,Y1209,I1210,K1211,W1212
	P4	Red	Y904,R905,F906,N907,G908,I909,G910,V911,T912,Q913,N914,V915,L916,Y917,E918,N919,Q920,K921
	P5	Yellow	A1190,K1191,N1192,L1193,N1194,E1195,S1196,L1197,I1198,D1199,L1200,Q1201,E1202
	P6	Purple	T887,F888,G889,A890,G891,A892,A893,L894,Q895,I896,P897,F898,A899,M900
	P7	Black	T1027,K1028,S1030,E1031,L1034,G1035,S1037,K1038,R1039,V1040,D1041,F1042
	P8	Pink	C1254,K1255,F1256,D1257,E1258,D1259,D1260,S1261,E1262,P1263,V1264,L1265,K1266,G1267,V1268,K1269,L1270,H1271,Y1272
	P9	Pink	C1254,K1255,F1256,D1257,E1258,D1259,D1260,S1261,E1262,P1263,V1264,L1265,K1266,G1267,V1268,K1269,L1270,H1271
	P10	Orange	R646,A647,G648,L650,I651,E654,H655,V656,N657,N658,S659,Y660,E661,C662,D663,I664,P665
	P11	Cyan	F175,L176,M177,D178,L179,E180,G181,K182,Q183,G184,N185,F186,K187,N188,L189,R190
Spike (S) SARS-CoV	P1	Green	G192,K198,G199,Y200,Q201,P202,I203,D204,V205,V206,R207,D208,L209,P210,S211,G212,L216,P218,I219,F220,L222,P223,L224
	P2	Blue	T431,S432,T433,G434,N435,Y436,N437,Y438,K439,Y440,R441,Y442,L443,R444,H445,G446,K447,L448,R449,P450,F451
	P3	Red	R453,D454,I455,S456,N457,V458,P459,F460,S461,P462,D463,G464,K465,P466,C467,T468,P469,P470,A471,L472,N473,C474
	P4	Yellow	G552,R553,D554,V555,S556,D557,F558,T559,D560,S561,V562,R563,D564,P565,K566,T567,S568,E569,I570
	P5	Purple	N1140,H1141,T1142,S1143,P1144,D1145,V1146,D1147,L1148,G1149,D1150,I1151,S1152,G1153,I1154,N1155,A1156,S1157,V1158,V1159,N1160
	P6	Black	R1167,L1168,N1169,E1170,V1171,A1172,K1173,N1174,L1175,N1176,E1177,S1178,L1179,I1180,D1181,L1182
	P7	Pink	Q1183,E1184,L1185,G1186,K1187,Y1188,E1189,Q1190,Y1191,I1192,K1193,W1194,P1195,W1196,Y1197,V1198,W1199,L1200,G1201,F1202,I1203,A1204
	P8	Cyan	G1233,S1234,C1235,C1236,K1237,F1238,D1239,E1240,D1241,D1242,S1243,E1244,P1245,V1246,L1247,K1248,G1249,V1250,K1251,L1252,H1253
	P9	Cyan	C1236,K1237,F1238,D1239,E1240,D1241,D1242,S1243,E1244,P1245,V1246,L1247,K1248,G1249,V1250,K1251,L1252,H1253

SARS-CoV-2 proteins	Peptide	Color code	Conformational B-cell epitope
Nucleocapsid (N)	P1	Green	D22,S23,T24,G25,S26,N27,Q28,N29,G30,E31,R32,S33,G34,A35,R36,S37,K38,Q39,R40
	P2	Blue	R95,G96,G97,D98,G99,K100,M101,K102,D103,L104,S105,P106,R107,W108,Y109,F110,Y111,Y112,L113,G114
	P3	Blue	L104,S105,P106,R107,W108,Y109,F110,Y111,Y112,L113,G114,T115,G116,P117
	P4	Red	P365,T366,E367,P368,K369,K370,D371,K372,K373,K374,K375,A376,D377,E378,T379,Q380,A381,L382,P383,Q384,R385,Q386,K387
	P5	Red	E378,T379,Q380,A381,L382,P383,Q384,R385,Q386,K387,K388,Q389,Q390,T391,V392,T393,L394,L395,P396
	P6	Yellow	A267,Y268,N269,V270,T271,Q272,A273,F274,G275,R276,R277,G278,P279,E280,Q281,T282,Q283,G284,N285,F286,G287,D288,Q289,E290,L291,I292,R293,Q294,G295,T296,D297,Y298,K299,H300
	P7	Yellow	Q272,A273,F274,G275,R276,R277,G278,P279,E280,Q281,T282,Q283,G284,N285,F286,G287,D288,Q289
	P8	Purple	P302,Q303,I304,A305,Q306,F307,A308,P309,S310,F315,G316,R319,I320,G321,E323,V324,T325,P326,S327,G328,W330
Envelope (E)	P1	Green	F56,Y57,V58,Y59,S60,R61,V62,K63,N64,L65,N66,S67,S68,R69,V70,P71,D72,L73,L74,V75
	P2	Blue	M1,Y2,S3,F4,V5,S6,E7,E8,T9,G10,T11,L12,I13,V14,N15,S16,V17,L18,L19,F20,L21,A22,F23,V24,V25,F26,L27,L28,V29,T30,L31
	P3	Yellow	S60,R61,V62,K63,N64,L65,N66,S67,S68,R69,V70,P71,D72,L73,L74
Membrane (M)	P1	Green	R72,I73,N74,W75,I76,T77,G78,I80,A81,I82,A83,M84,L87,V88,G89,M91,W92,F96,F100,R101,F103,A104,R105,T106,R107,S108,M109,W110,S111,F112,N113,P114,E115,T116,L119
	P2	Green	L102,A104,R105,T106,R107,S108,M109,W110,S111,F112,N113,P114,E115
	P3	Blue	L156,G157,R158,C159,D160,I161,K162,D163,L164,P165,A171,T172,S173,R174,T175,L176,S177
	P4	Red	I48,I49,K50,L51,I52,F53,L54,W55,L56,L57,W58,P59,V60,T61,L62,A63,C64,F65,V66,L67,A68,A69,V70,R72
	P5	Yellow	G25,F26,L27,L29,T30,W31,I32,C33,L34,L35,Q36,F37,A38,Y39,A40,N41,R42,N43,R44
	P6	Purple	G189,D190,F193,A194,A195,S197,Y199,R200,I201,G202,N203,Y204,K205,L206,N207,T208,D209,H210
	P7	Cyan	L176,S177,Y178,Y179,K180,L1801G182,A183,S184,Q185,R186,V187

Figure 5

Localization of conformational B-cell epitopes on the modelled structure of SARS-CoV-2 proteins. S proteins of SARS-CoV-2 (A), SARS-CoV (B), Nucleocapsid (C), Envelope (D) and Membrane (E) proteins of SARS-CoV-2. The selected B-cell epitopes are provided in the inset table and the corresponding color shows the localization on the modelled structures of S, E, M and N proteins of SARS-CoV-2.



SARS-CoV-2 proteins	Peptide	Color code	Conformational B-cell epitope
Spike (S)	P1	Green	S1161,P1162,D1163,V1164,D1165,L1166,G1167,D1168,I1169,S1170,G1171,I1172,N1173,A1174,S1175,V1176,V1177,N1178
	P2	Blue	G1204,K1205,Y1206,E1207,Q1208,Y1209,I1210,K1211,W1212,P1213,W1214,Y1215,I1216,W1217,L1218,G1219,F1220,I1221
	P3	Blue	Q1201,E1202,L1203,G1204,K1205,Y1206,E1207,Q1208,Y1209,I1210,K1211,W1212
	P4	Red	Y904,R905,F906,N907,G908,I909,G910,V911,T912,Q913,N914,V915,L916,Y917,E918,N919,Q920,K921
	P5	Yellow	A1190,K1191,N1192,L1193,N1194,E1195,S1196,L1197,I1198,D1199,L1200,Q1201,E1202
	P6	Purple	T887,F888,G889,A890,G891,A892,A893,L894,Q895,I896,P897,F898,A899,M900
	P7	Black	T1027,K1028,S1030,E1031,L1034,G1035,S1037,K1038,R1039,V1040,D1041,F1042
	P8	Pink	C1254,K1255,F1256,D1257,E1258,D1259,D1260,S1261,E1262,P1263,V1264,L1265,K1266,G1267,V1268,K1269,L1270,H1271,Y1272
	P9	Pink	C1254,K1255,F1256,D1257,E1258,D1259,D1260,S1261,E1262,P1263,V1264,L1265,K1266,G1267,V1268,K1269,L1270,H1271
	P10	Orange	R646,A647,G648,L650,I651,E654,H655,V656,N657,N658,S659,Y660,E661,C662,D663,I664,P665
	P11	Cyan	F175,L176,M177,D178,L179,E180,G181,K182,Q183,G184,N185,F186,K187,N188,L189,R190
Spike (S) SARS-CoV	P1	Green	G192,K198,G199,Y200,Q201,P202,I203,D204,V205,V206,R207,D208,L209,P210,S211,G212,L216,P218,I219,F220,L222,P223,L224
	P2	Blue	T431,S432,T433,G434,N435,Y436,N437,Y438,K439,Y440,R441,Y442,L443,R444,H445,G446,K447,L448,R449,P450,F451
	P3	Red	R453,D454,I455,S456,N457,V458,P459,F460,S461,P462,D463,G464,K465,P466,C467,T468,P469,P470,A471,L472,N473,C474
	P4	Yellow	G552,R553,D554,V555,S556,D557,F558,T559,D560,S561,V562,R563,D564,P565,K566,T567,S568,E569,I570
	P5	Purple	N1140,H1141,T1142,S1143,P1144,D1145,V1146,D1147,L1148,G1149,D1150,I1151,S1152,G1153,I1154,N1155,A1156,S1157,V1158,V1159,N1160
	P6	Black	R1167,L1168,N1169,E1170,V1171,A1172,K1173,N1174,L1175,N1176,E1177,S1178,L1179,I1180,D1181,L1182
	P7	Pink	Q1183,E1184,L1185,G1186,K1187,Y1188,E1189,Q1190,Y1191,I1192,K1193,W1194,P1195,W1196,Y1197,V1198,W1199,L1200,G1201,F1202,I1203,A1204
	P8	Cyan	G1233,S1234,C1235,C1236,K1237,F1238,D1239,E1240,D1241,D1242,S1243,E1244,P1245,V1246,L1247,K1248,G1249,V1250,K1251,L1252,H1253
	P9	Cyan	C1236,K1237,F1238,D1239,E1240,D1241,D1242,S1243,E1244,P1245,V1246,L1247,K1248,G1249,V1250,K1251,L1252,H1253

SARS-CoV-2 proteins	Peptide	Color code	Conformational B-cell epitope
Nucleocapsid (N)	P1	Green	D22,S23,T24,G25,S26,N27,Q28,N29,G30,E31,R32,S33,G34,A35,R36,S37,K38,Q39,R40
	P2	Blue	R95,G96,G97,D98,G99,K100,M101,K102,D103,L104,S105,P106,R107,W108,Y109,F110,Y111,Y112,L113,G114
	P3	Blue	L104,S105,P106,R107,W108,Y109,F110,Y111,Y112,L113,G114,T115,G116,P117
	P4	Red	P365,T366,E367,P368,K369,K370,D371,K372,K373,K374,K375,A376,D377,E378,T379,Q380,A381,L382,P383,Q384,R385,Q386,K387
	P5	Red	E378,T379,Q380,A381,L382,P383,Q384,R385,Q386,K387,K388,Q389,Q390,T391,V392,T393,L394,L395,P396
	P6	Yellow	A267,Y268,N269,V270,T271,Q272,A273,F274,G275,R276,R277,G278,P279,E280,Q281,T282,Q283,G284,N285,F286,G287,D288,Q289,E290,L291,L292,R293,Q294,G295,T296,D297,Y298,K299,H300
	P7	Yellow	Q272,A273,F274,G275,R276,R277,G278,P279,E280,Q281,T282,Q283,G284,N285,F286,G287,D288,Q289
	P8	Purple	P302,Q303,I304,A305,Q306,F307,A308,P309,S310,F315,G316,R319,I320,G321,E323,V324,T325,P326,S327,G328,W330
Envelope (E)	P1	Green	F56,Y57,V58,Y59,S60,R61,V62,K63,N64,L65,N66,S67,S68,R69,V70,P71,D72,L73,L74,V75
	P2	Blue	M1,Y2,S3,F4,V5,S6,E7,E8,T9,G10,T11,L12,I13,V14,N15,S16,V17,L18,L19,F20,L21,A22,F23,V24,V25,F26,L27,L28,V29,T30,L31
	P3	Yellow	S60,R61,V62,K63,N64,L65,N66,S67,S68,R69,V70,P71,D72,L73,L74
Membrane (M)	P1	Green	R72,I73,N74,W75,I76,T77,G78,I80,A81,I82,A83,M84,L87,V88,G89,M91,W92,F96,F100,R101,F103,A104,R105,T106,R107,S108,M109,W110,S111,F112,N113,P114,E115,T116,L119
	P2	Green	L102,A104,R105,T106,R107,S108,M109,W110,S111,F112,N113,P114,E115
	P3	Blue	L156,G157,R158,C159,D160,I161,K162,D163,L164,P165,A171,T172,S173,R174,T175,L176,S177
	P4	Red	I48,I49,K50,L51,I52,F53,L54,W55,L56,L57,W58,P59,V60,T61,L62,A63,C64,F65,V66,L67,A68,A69,V70,R72
	P5	Yellow	G25,F26,L27,L29,T30,W31,I32,C33,L34,L35,Q36,F37,A38,Y39,A40,N41,R42,N43,R44
	P6	Purple	G189,D190,F193,A194,A195,S197,Y199,R200,I201,G202,N203,Y204,K205,L206,N207,T208,D209,H210
	P7	Cyan	L176,S177,Y178,Y179,K180,L1801G182,A183,S184,Q185,R186,V187

Figure 5

Localization of conformational B-cell epitopes on the modelled structure of SARS-CoV-2 proteins. S proteins of SARS-CoV-2 (A), SARS-CoV (B), Nucleocapsid (C), Envelope (D) and Membrane (E) proteins of SARS-CoV-2. The selected B-cell epitopes are provided in the inset table and the corresponding color shows the localization on the modelled structures of S, E, M and N proteins of SARS-CoV-2.

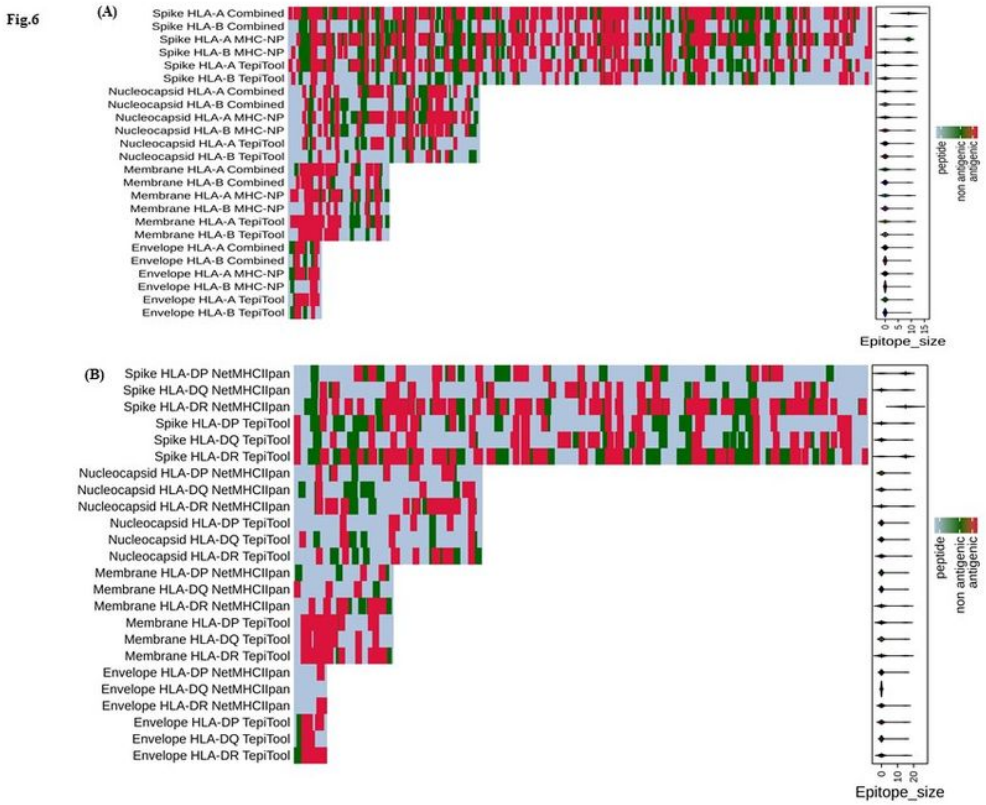


Figure 6

Summary of SARS-CoV-2-derived T-cell epitopes. Heat map showing the distribution of (A) HLA-class I and (B) II epitopes across the protein sequences of spike (1273 aa), nucleocapsid (419 aa), membrane (222 aa) and envelope (75 aa) proteins of SARS-CoV-2. Red color represents likely antigenic epitopes that were predicted using the methods described in figure 1. Strong binding affinity epitopes with <0.5% rank and 2% rank, to HLA class I and class II, respectively, for each HLA molecule are represented here.

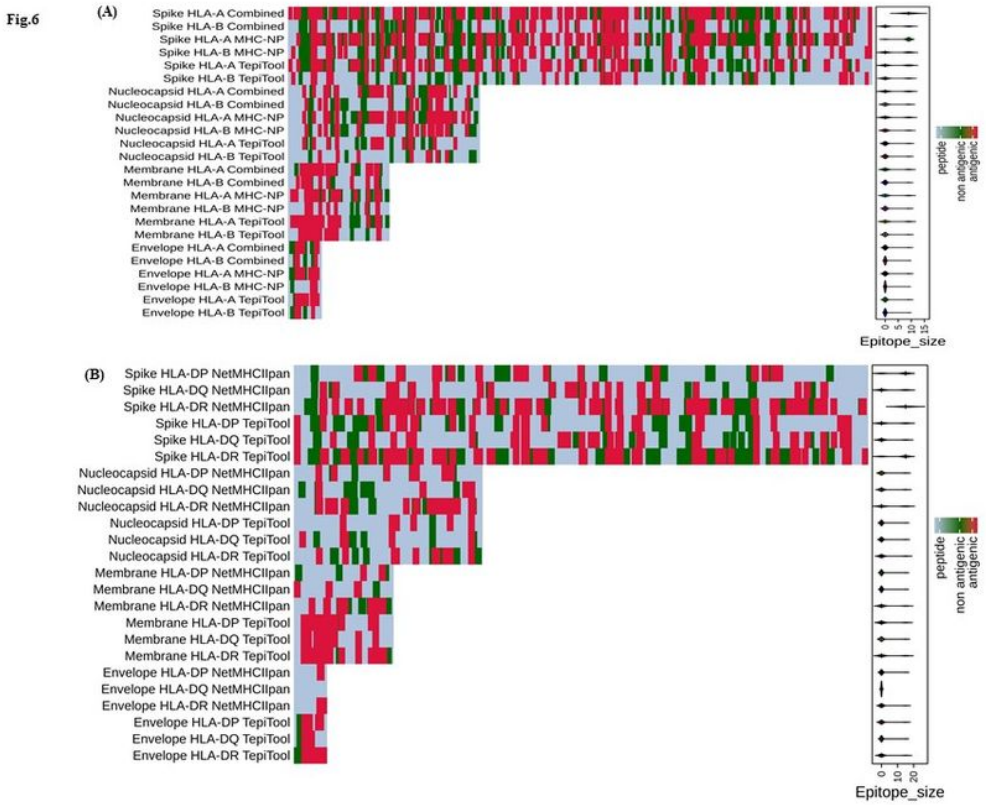


Figure 6

Summary of SARS-CoV-2-derived T-cell epitopes. Heat map showing the distribution of (A) HLA-class I and (B) II epitopes across the protein sequences of spike (1273 aa), nucleocapsid (419 aa), membrane (222 aa) and envelope (75 aa) proteins of SARS-CoV-2. Red color represents likely antigenic epitopes that were predicted using the methods described in figure 1. Strong binding affinity epitopes with <0.5% rank and 2% rank, to HLA class I and class II, respectively, for each HLA molecule are represented here.

Fig.7A

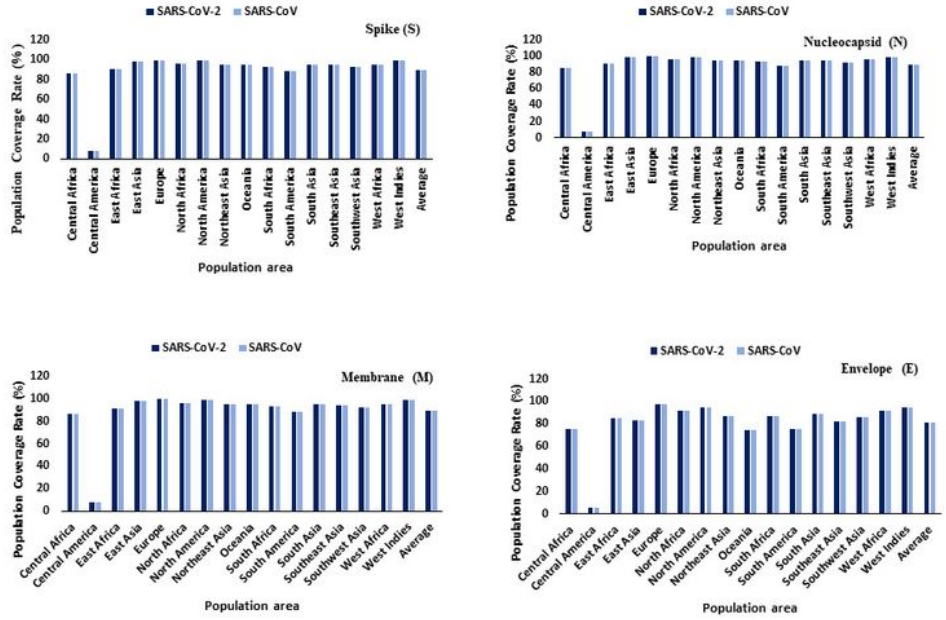


Fig.7B

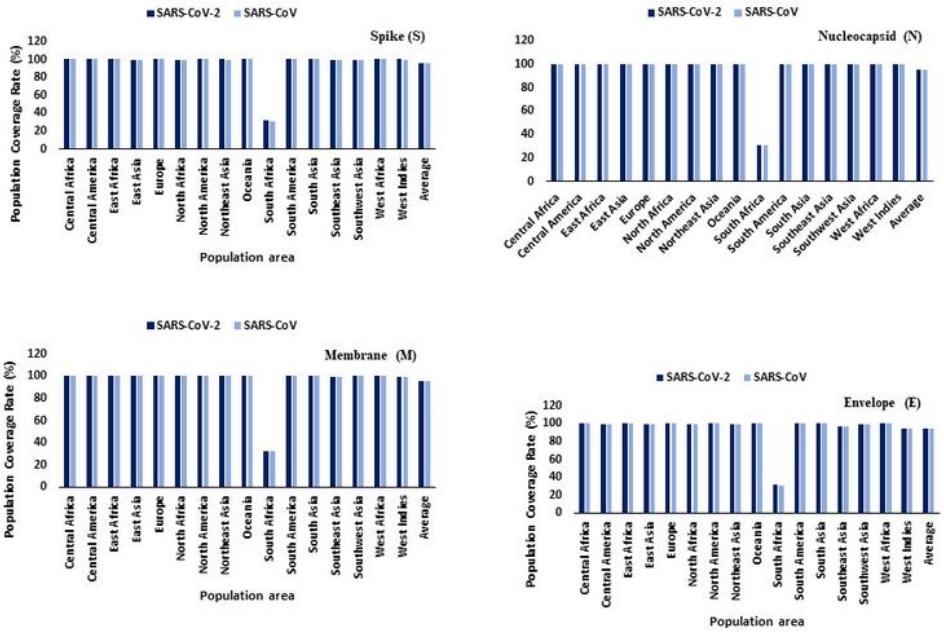


Figure 7

(A) Predicted population coverage rate (%) of predicted MHC I binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV. All the epitopes provided in table 3b and 3c have been analysed for population coverage analysis using IEDB resource tool. (B) Predicted population coverage rate (%) of predicted MHC II binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV. All the

epitopes provided in table 4b and 4c have been analysed for population coverage analysis using IEDB resource tool.

Fig.7A

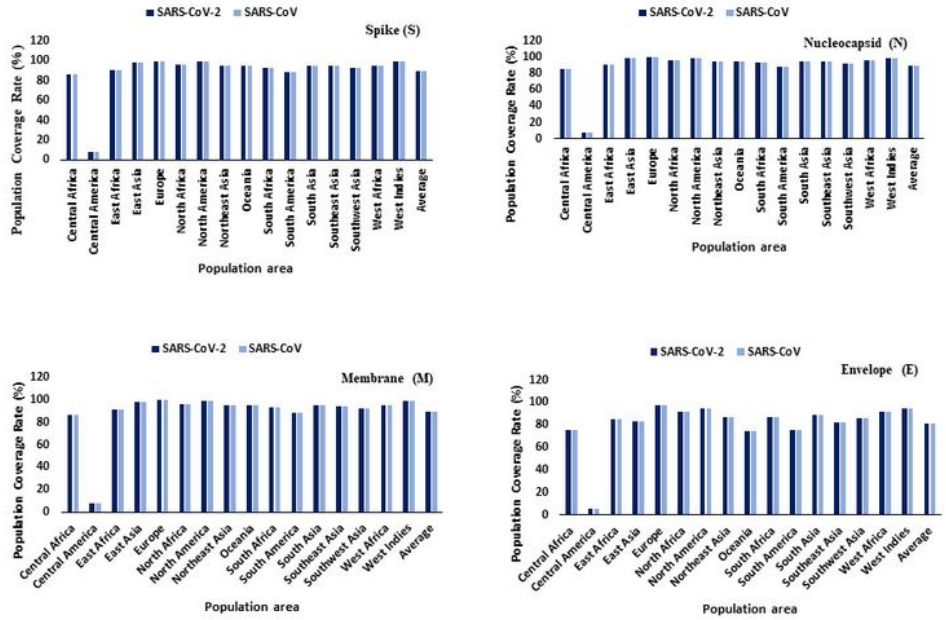


Fig.7B

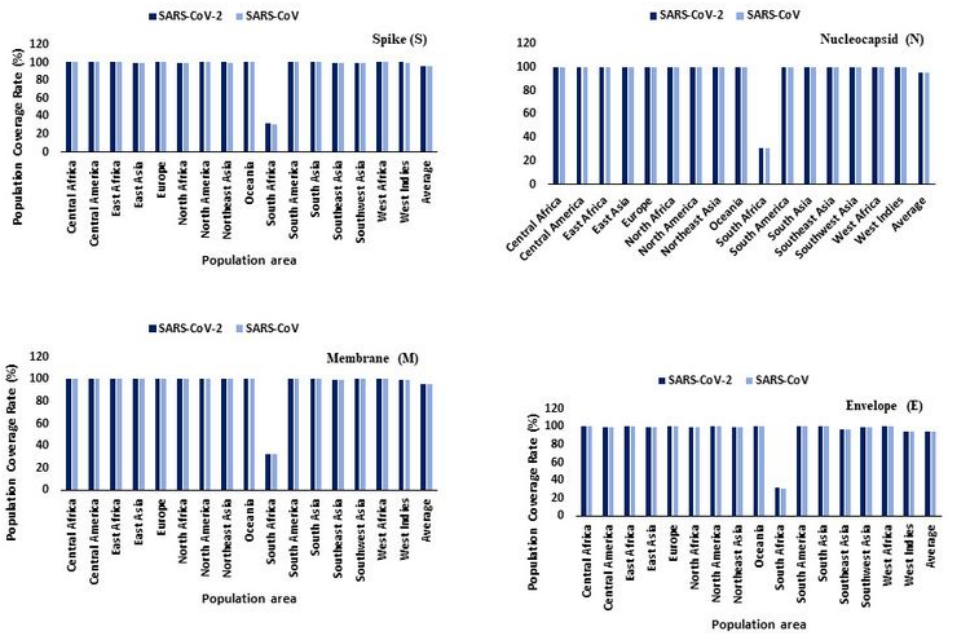


Figure 7

(A) Predicted population coverage rate (%) of predicted MHC I binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV. All the epitopes provided in table 3b and 3c have been analysed for population coverage analysis using IEDB resource tool. (B) Predicted population coverage rate (%) of

predicted MHC II binding epitopes of S, E, M and N proteins of SARS-CoV-2 and SARS-CoV. All the epitopes provided in table 4b and 4c have been analysed for population coverage analysis using IEDB resource tool.

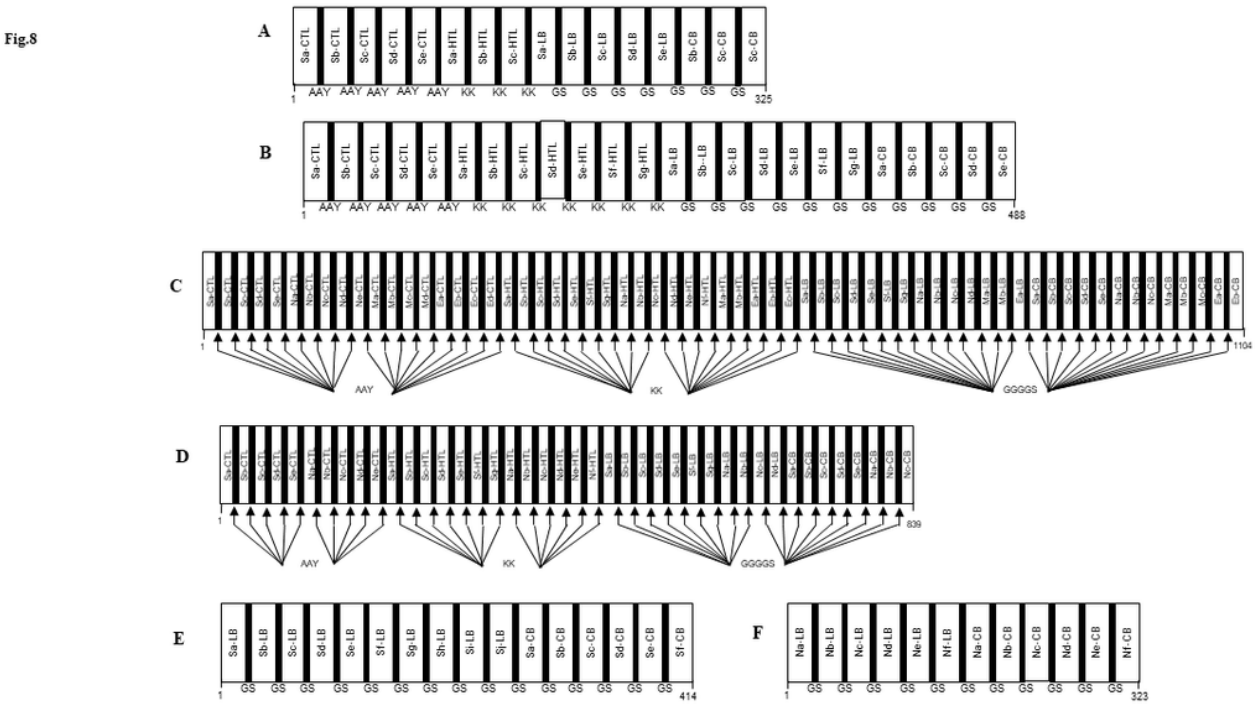


Figure 8

Schematic diagram of multi-epitope protein constructs of SARS-CoV-2. The multi-epitope constructs encompassing predicted B- and T-cell epitopes. CTL epitopes has been joined by AAY linker, whereas, KK and GGGGS linkers were used to join the HTL and linear and conformational B-cell epitopes, respectively. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

Fig.8

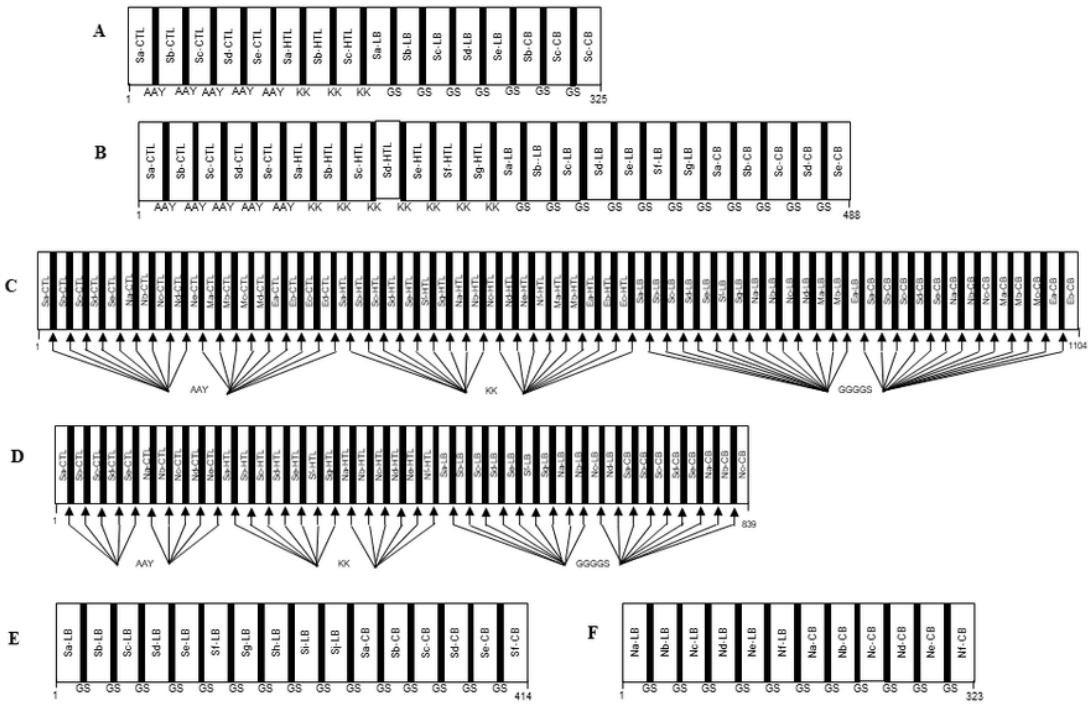


Figure 8

Schematic diagram of multi-epitope protein constructs of SARS-CoV-2. The multi-epitope constructs encompassing predicted B- and T-cell epitopes. CTL epitopes has been joined by AAY linker, whereas, KK and GGGGS linkers were used to join the HTL and linear and conformational B-cell epitopes, respectively. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

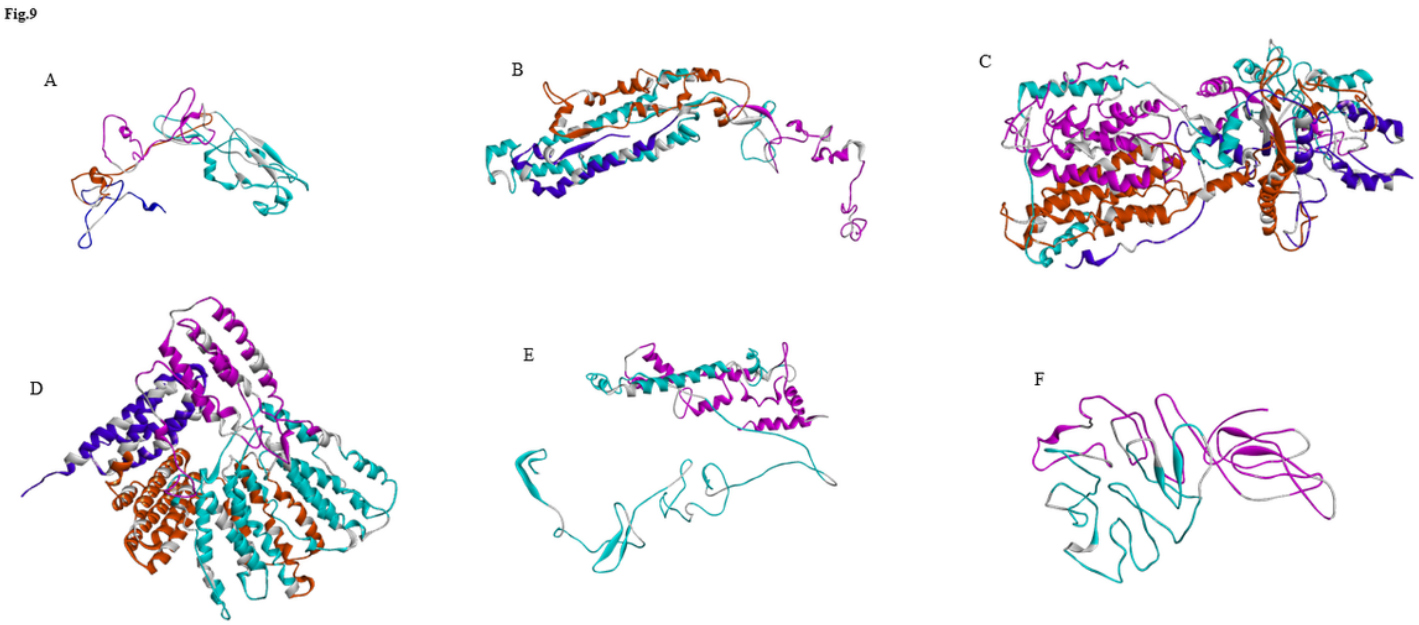


Figure 9

Schematic representation of modelled structure of multi-epitope protein constructs of SARS-CoV-2. Blue color represents CTL epitopes, orange represents HTL epitope, cyan represents linear B-cell epitopes and conformational B-cell epitope is highlighted with magenta. A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

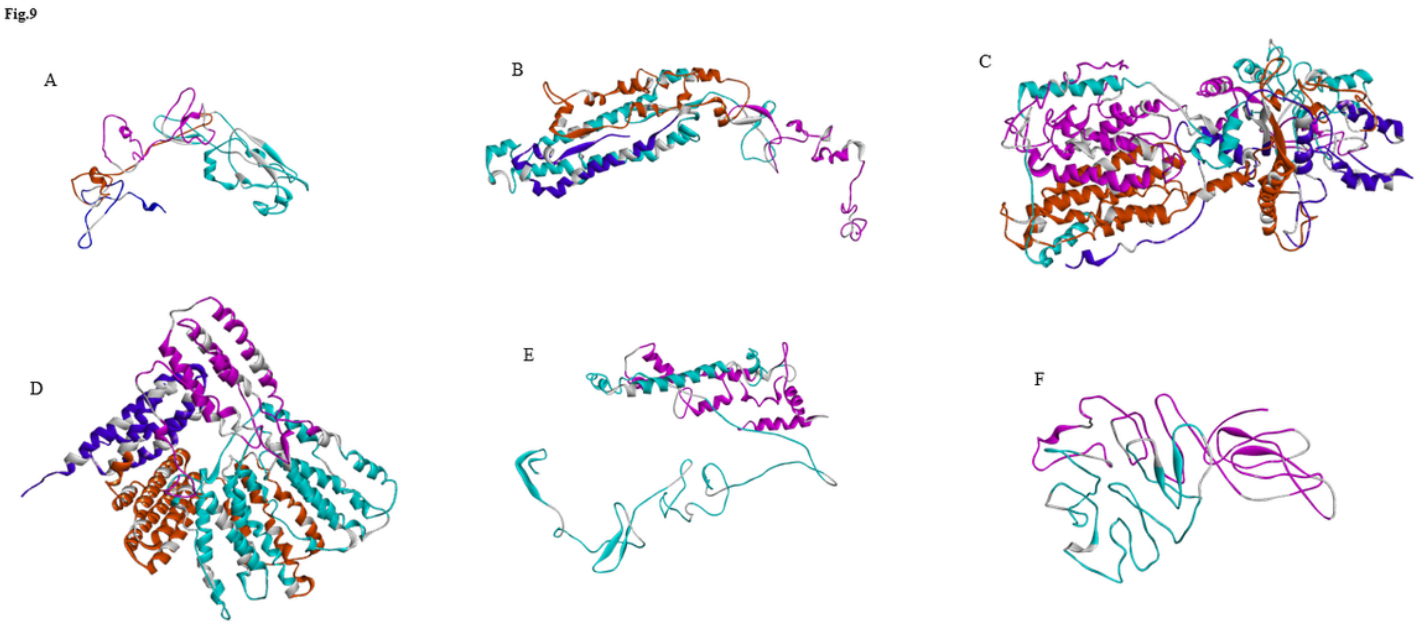


Figure 9

Schematic representation of modelled structure of multi-epitope protein constructs of SARS-CoV-2. Blue color represents CTL epitopes, orange represents HTL epitope, cyan represents linear B-cell epitopes and conformational B-cell epitope is highlighted with magenta. A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

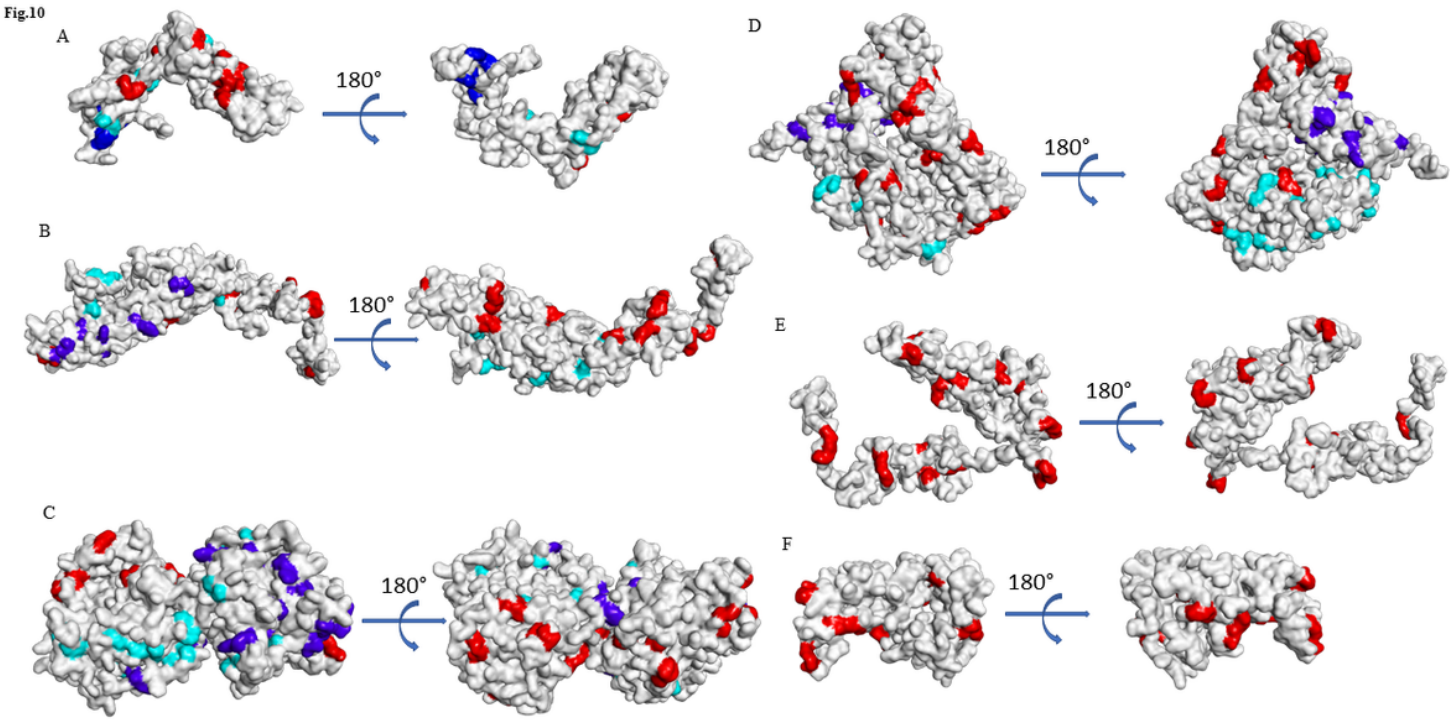


Figure 10

Accessibility of linkers of multi-epitope protein constructs of SARS-CoV-2. Blue color represents AAY linker, cyan represents KK linker and GGGGS is represented by red color. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

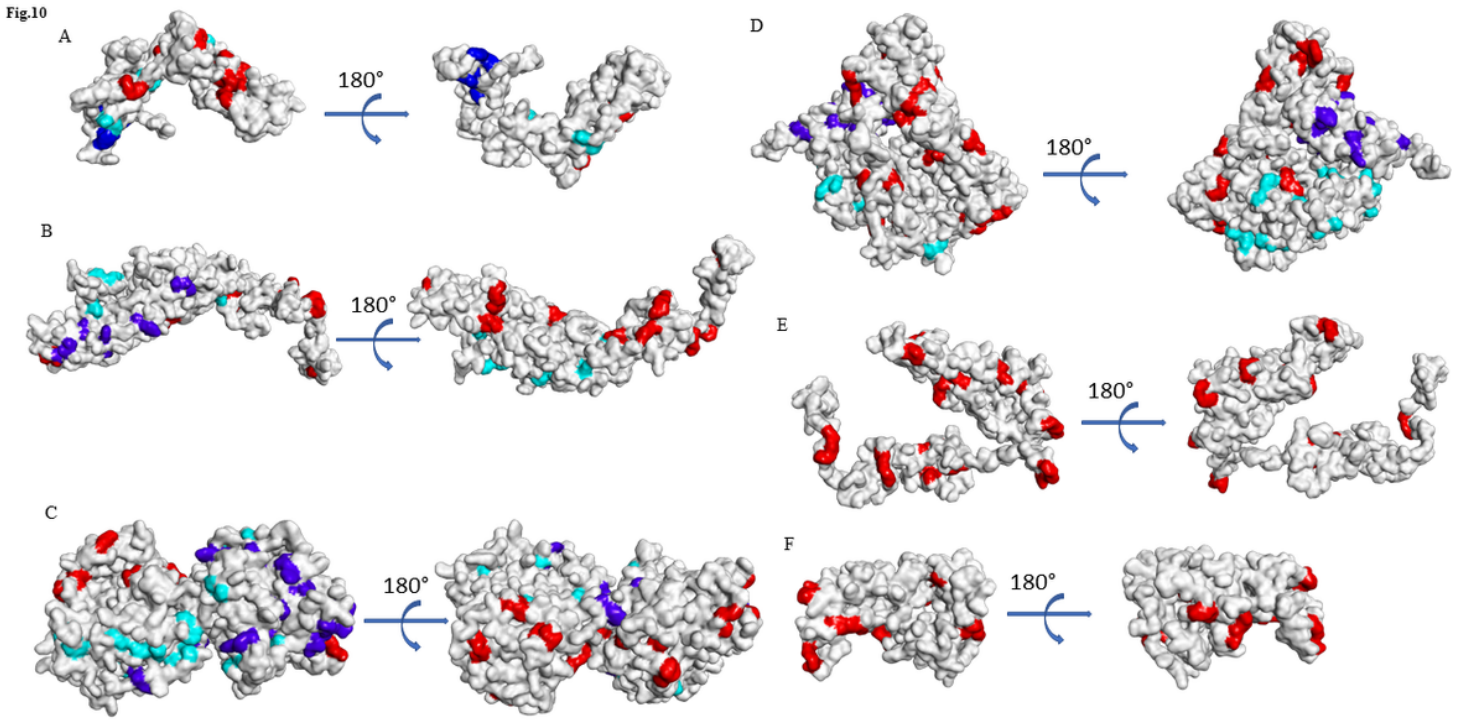


Figure 10

Accessibility of linkers of multi-epitope protein constructs of SARS-CoV-2. Blue color represents AAY linker, cyan represents KK linker and GGGGS is represented by red color. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

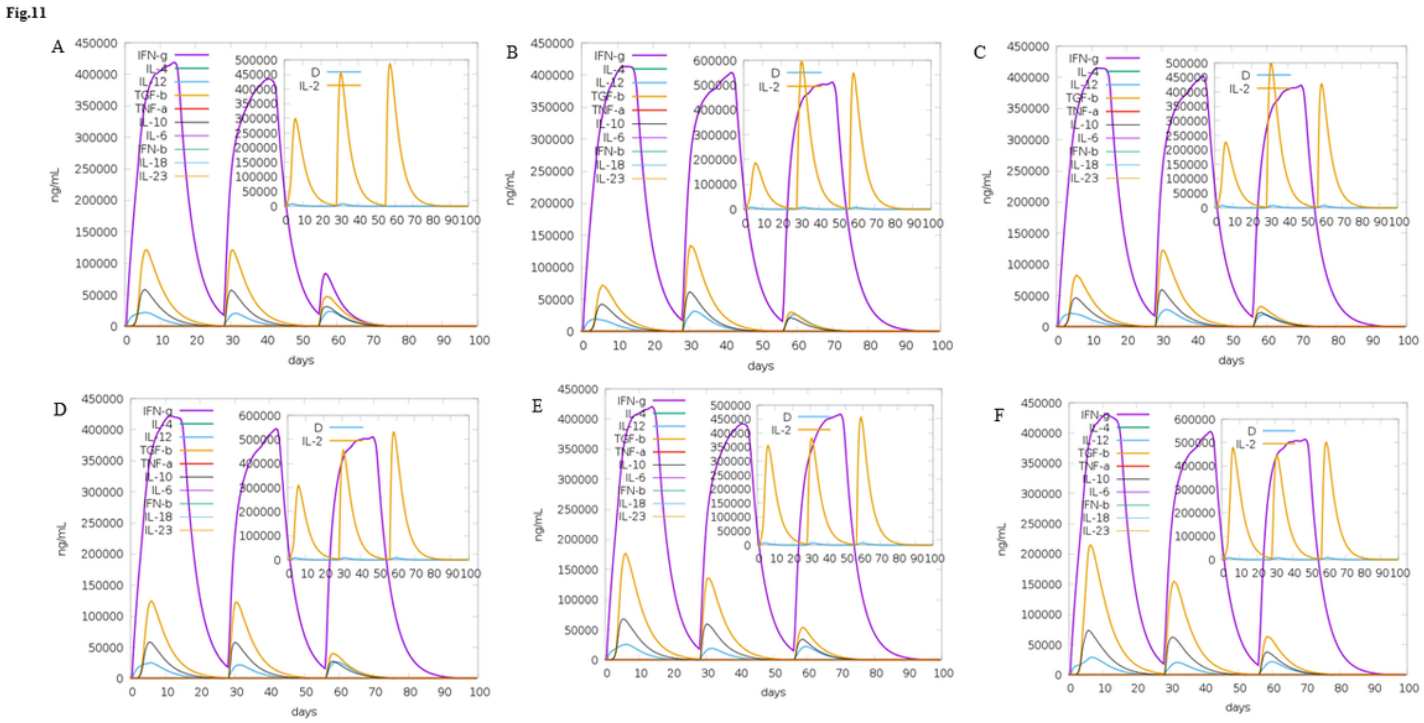


Figure 11

In-silico immune simulation of multi-epitope protein constructs of SARS-CoV-2. C-ImmSim simulation of the cytokine levels induced by three injections given at 4 weeks apart. The main plot shows cytokine levels after the injections. The inset plot shows IL-2 level with the Simpson index, D indicated by the dotted line. D is a measure of diversity. Increase in D over time indicates emergence of different epitope-specific dominant clones of T-cells. The smaller the D value, the lower the diversity. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

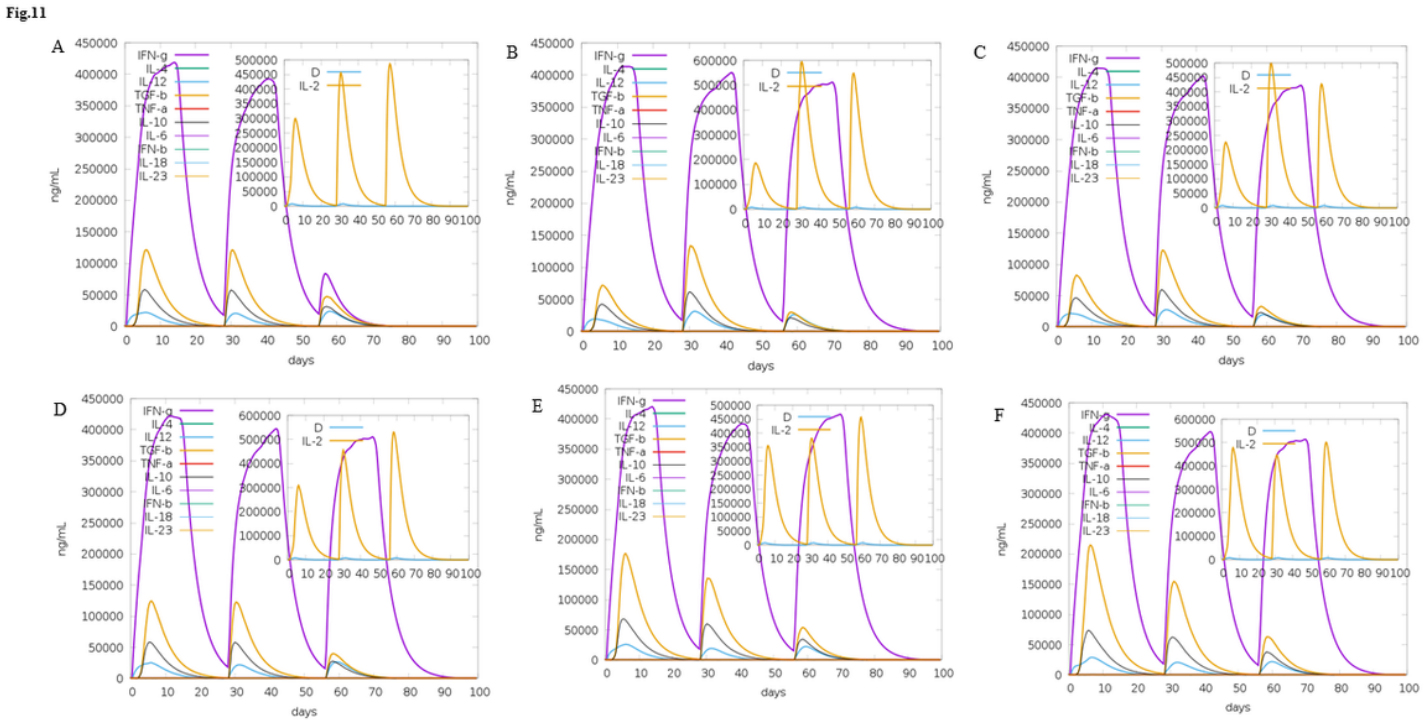


Figure 11

In-silico immune simulation of multi-epitope protein constructs of SARS-CoV-2. C-ImmSim simulation of the cytokine levels induced by three injections given at 4 weeks apart. The main plot shows cytokine levels after the injections. The inset plot shows IL-2 level with the Simpson index, D indicated by the dotted line. D is a measure of diversity. Increase in D over time indicates emergence of different epitope-specific dominant clones of T-cells. The smaller the D value, the lower the diversity. (A) RBD of spike protein (B- and T-cell epitopes), (B) Spike protein (B- and T-cell epitopes), (C) Structural protein construct (B- and T-cell epitopes of S, N, M and E protein), (D) Chimeric construct of S and N proteins (B- and T-cell epitopes of S and N), (E) spike protein (B-cell epitopes of S protein) and (F) Nucleocapsid protein (B-cell epitopes of N protein).

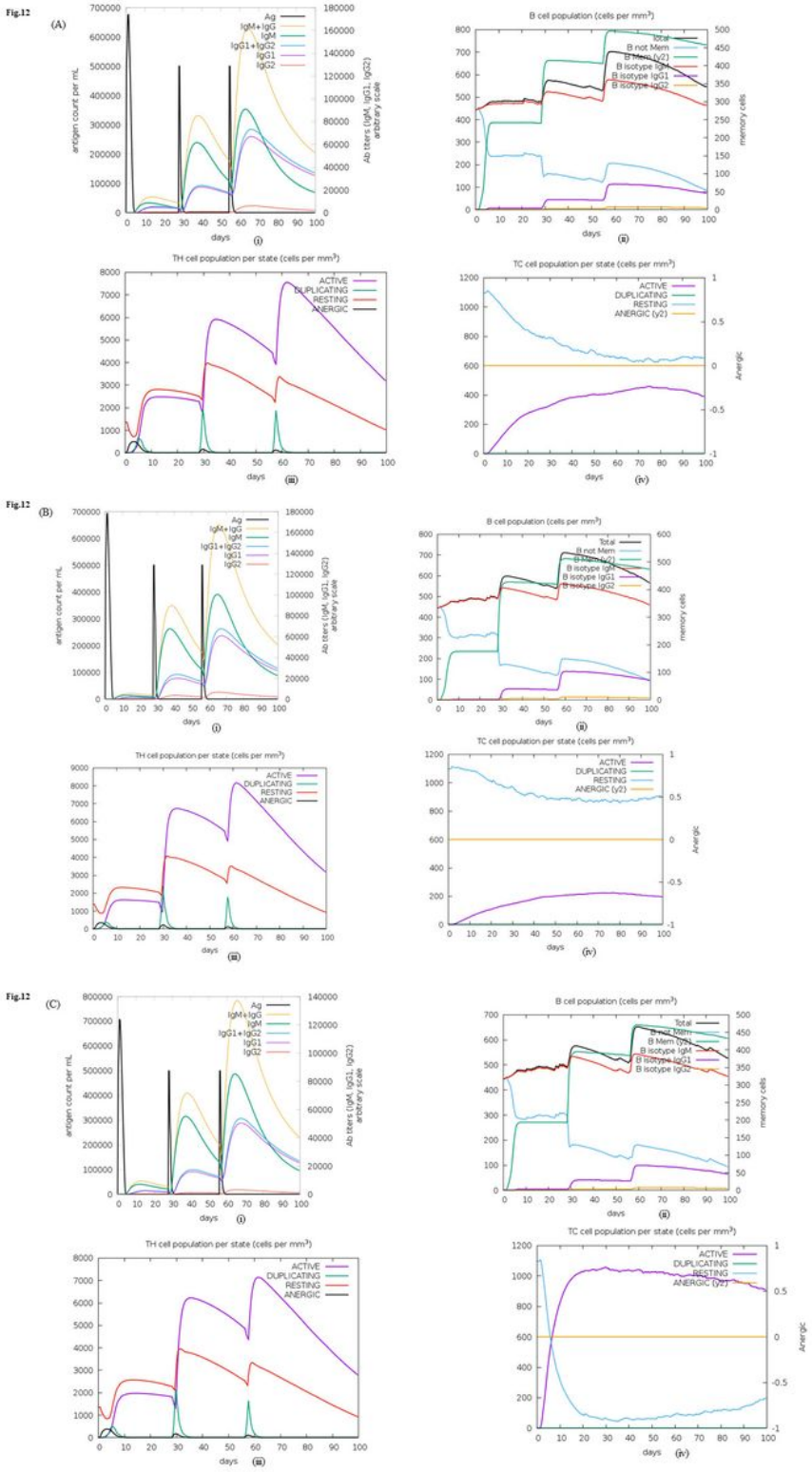


Figure 12

C-ImmSim presentation of an in-silico immune simulation with the multi-epitope protein constructs of SARS-CoV-2. (A) RBD of spike protein, (B) Spike protein, (C) Structural protein construct and (D) Chimeric construct of S and N proteins. (i) nImmunoglobulin production in response to antigen injections (black vertical lines); specific subclasses are indicated as coloured peaks. (ii) The evolution of B-cell populations after the three injections. (iii) The evolution of T-helper, and (iv) T-cytotoxic cell populations per state after

the injections with constructs. The resting state represents cells not presented with the antigen while the anergic state represents tolerance of the T-cells to the antigen due to repeated exposures.

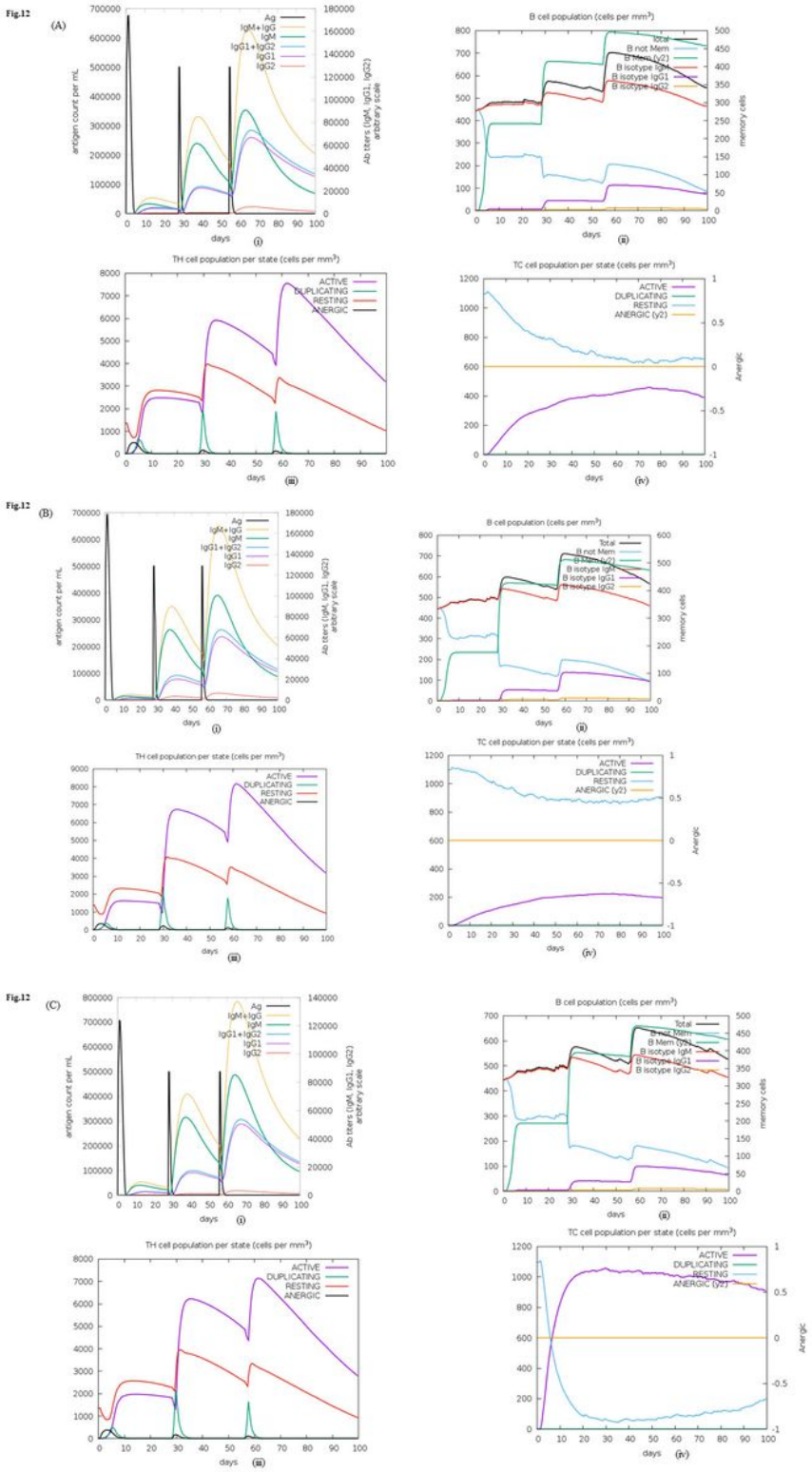


Figure 12

C-ImmSim presentation of an in-silico immune simulation with the multi-epitope protein constructs of SARS-CoV-2. (A) RBD of spike protein, (B) Spike protein, (C) Structural protein construct and (D) Chimeric construct of S and N proteins. (i) nImmunoglobulin production in response to antigen injections (black

vertical lines); specific subclasses are indicated as coloured peaks. (ii) The evolution of B-cell populations after the three injections. (iii) The evolution of T-helper, and (iv) T-cytotoxic cell populations per state after the injections with constructs. The resting state represents cells not presented with the antigen while the anergic state represents tolerance of the T-cells to the antigen due to repeated exposures.

Fig.13

Peptide	Color code	Linear B-cell epitope
P1	Green	331-NITNLCPFGEVFNATRFASVYAWNRK-356
P2	Blue	403-RGDEVRQIAPGQTGKIADYNYKLPD-427
P3	Red	437-NSNNLDSKVGGNYNYLYRLFRKSNL-461
P4	Yellow	483-VEGFNCYFPLQ-493
P5	Purple	543-FNGLTGTGVLTESNKKFL-560
P6	Orange	559-FLPFQQFGRDIADTTDAV-576
P7	Cyan	590-CSFGGVSVITPGTNTSNQ-607
P8	Pink	598-ITPGTNTSNQVAVLYQDVNCTE-619
P9	Pink	607-QVAVLYQDVNCTEVPVAI-624
P10	Black	652- GAEHVNNSYECDIPIGAG-669
P11	Black	653-AEHVNNSYECDIPI-666
P12	Black	660- YECDIPIGAGICASYQTQ-677
P13	Maroon	780-EVFAQVKQIYKTPPIKDF-797
P14	Indigo	798- GGFNFSQILPDPSKPSKR-815
P15	Brown	819-EDLLFNKVTLADAGFIKQ-836
P16	Olive green	1149- KEELDKYFKNHTSPDVDL-1166
P17	Light blue	1236-CKFDEDDSEPVLKGVKLHYT-1255

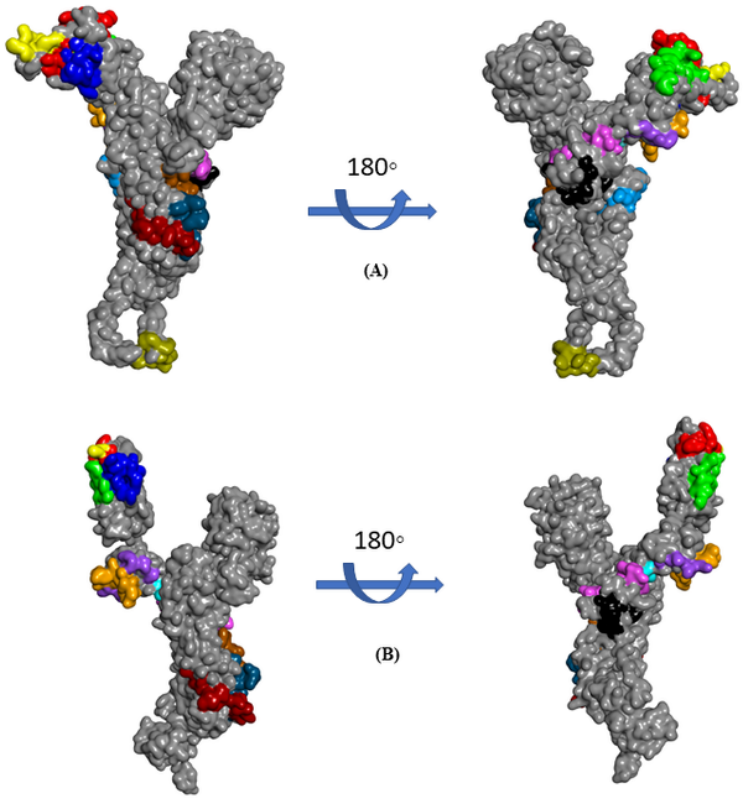


Figure 13

Localization of the linear B-cell epitopes on the modelled and crystal structure of the S protein of SARS-CoV-2. Localization of selected monomeric B-cell epitopes (inset table) on the (A) modelled and (B) crystal structure (PDB:6VSB) of the S protein of SARS-CoV-2 was performed using the BIOVIA discovery studio 2017 R2.

Fig.13

Peptide	Color code	Linear B-cell epitope
P1	Green	331-NITNLCPFGEVFNATRFASVYAWNRK-356
P2	Blue	403-RGDEVRQIAPGQTGKIADYNYKLPD-427
P3	Red	437-NSNNLDSKVGGNYNYLYRLFRKSNL-461
P4	Yellow	483-VEGFNCYFPLQ-493
P5	Purple	543-FNGLTGTGVLTESNKKFL-560
P6	Orange	559-FLPFQQFGRDIADTTDAV-576
P7	Cyan	590-CSFGGVSVITPGTNTSNQ-607
P8	Pink	598-ITPGTNTSNQVAVLYQDVNCTE-619
P9	Pink	607-QVAVLYQDVNCTEVPVAI-624
P10	Black	652- GAEHVNNSYECDIPIGAG-669
P11	Black	653-AEHVNNSYECDIPI-666
P12	Black	660- YECDIPIGAGICASYQTQ-677
P13	Maroon	780-EVFAQVKQIYKTPPIKDF-797
P14	Indigo	798- GGFNFSQILPDPSKPSKR-815
P15	Brown	819-EDLLFNKVTLADAGFIKQ-836
P16	Olive green	1149- KEELDKYFKNHTSPDVDL-1166
P17	Light blue	1236-CKFDEDDSEPVLKGVKLHYT-1255

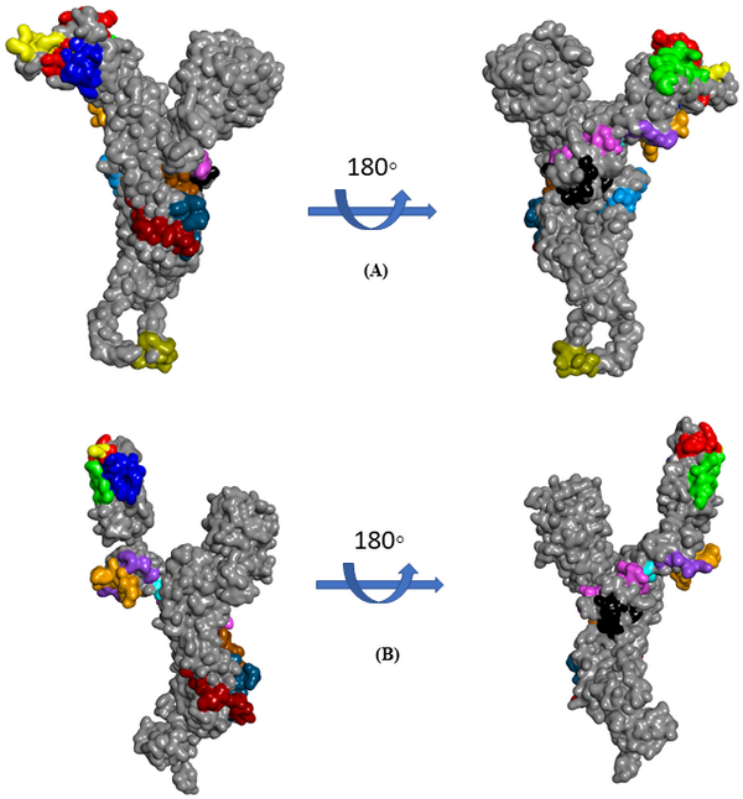


Figure 13

Localization of the linear B-cell epitopes on the modelled and crystal structure of the S protein of SARS-CoV-2. Localization of selected monomeric B-cell epitopes (inset table) on the (A) modelled and (B) crystal structure (PDB:6VSB) of the S protein of SARS-CoV-2 was performed using the BIOVIA discovery studio 2017 R2.

Supplementary Files

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

- supplement1.pdf
- Mergedsuptableandfigure.pdf
- supplement2.pdf
- NimaNamsaTableandfigureNDN1.pdf