Online version at http://www.minervamedica.it

Minerva Biotechnology and Biomolecular Research 2021 March;33(1):29-35
DOI: 10.23736/S2724-542X.20.02659-2

ORIGINAL ARTICLE

# *In silico* identification of linear B-cell epitope in Coronavirus 2019 (SARS-CoV-2) surface glycoprotein: a prospective towards peptide vaccine

Pushpendra SINGH [1], Manish K. TRIPATHI [2], Rahul SHRIVASTAVA [3] *

[1]ICAR- National Institute of High Security Animal Diseases, Bhopal, India; [2]Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology, Banaras Hindu University, Varanasi, India; [3]Department of Biological Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India

*Corresponding author: Rahul Shrivastava, Department of Biological Science and Engineering, Maulana Azad National Institute of Technology, 462001 Bhopal, India. E-mail: shrivastavarm1972@gmail.com

## ABSTRACT

BACKGROUND: The 2020 Coronavirus pandemic continuing spread of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS CoV-2). At the moment, there is no specific antiviral treatment or monoclonal antibodies or vaccines available for COVID-19. SARS-CoV-2 is positive-stranded RNA viruses with a crown-like appearance due to the occurrence of spike (surface) glycoproteins on the envelope. In the present study, the computational method used to predict the significant linear B cell epitopes of SARS-CoV-2 surface glycoprotein.

METHODS: FASTA sequence of SARS-CoV-2 surface glycoprotein was retrieved from the NCBI database, and further its primary and secondary structure was analyzed for its physical and chemicals properties. IEDB server was used to predict the B-cell epitopes.

RESULTS: ABCprep server and IEDB server prediction results for B-cell epitopes showed 16 and 21 linear epitope sequences respectively in the surface glycoprotein of SARS-CoV-2.

CONCLUSIONS: Obtained results conclude that predicted B-cell Epitopes may serve as an immunogen for eliciting monoclonal antibodies which can be used as a potential candidate for the treatment or diagnostic purpose for COVID-19.

(*Cite this article as:* Singh P, Tripathi MK, Shrivastava R. *In silico* identification of linear B-cell epitope in Coronavirus 2019 (SARS-CoV-2) surface glycoprotein: a prospective towards peptide vaccine. Minerva Biotechnol Biomol Res 2021;33:29-35. DOI: 10.23736/S2724-542X.20.02659-2)

KEY WORDS: Coronavirus; COVID-19; Severe acute respiratory syndrome coronavirus 2; Epitopes, B-lymphocyte; Pandemics.

The outbreak of 2019 novel coronavirus disease (COVID-19) was first reported on December 31, 2019, in Wuhan, China.[1] Coronaviruses are a big family of viruses that are common in many different species of animals, such as bats, cats, camels, and cattle.[2] Previous coronavirus outbreak infects and spread in human such as with severe acute respiratory syndrome coronavirus (SARS-CoV) first identified in the Guangdong province of southern China in 2002, Middle East respiratory syndrome coronavirus (MERS-CoV) that was first identified in Saudi Arabia in 2012, and now with this new coronavirus was first identified in Wuhan, China.[3] According to WHO, Worldwide registered 100,264 new COVID-19 cases in the past 24 hours, taking the total to 5,304,772. With 4342 deaths, the tally mounted to 342,029 on May 25, 2020.[4]

The Chinese laboratories isolated and identified a new type of coronavirus (novel coronavirus, nCoV), initially it was called as 2019-nCoV.[5] Afterwards, the International Committee on Taxonomy of Viruses (ICTV) termed it as SARS-CoV-2 virus due to its similarity to the one that caused the SARS outbreak (SARS-CoVs).[6] Now, this virus is known as "coronavirus disease 2019" (COVID-19) which is coined by the world health organization and declared this disease as pandemic on March 11, 2020, due to its global health effect.[7]

The genome size of coronavirus size ranges between approximately 28,000 and 32,000 bases, includes a variable number (from 6 to 11) of open reading frames (ORFs).[3] The first ORF of the entire genome encodes sixteen non-structural proteins (nsps), while others ORFs encode accessory proteins and structural proteins.[8] The four major structural proteins present in the virus are: surface glycoprotein (S), a small envelope protein (E), matrix protein (M), and nucleocapsid protein (N).[9] The structural protein, *i.e.* coronavirus surface protein (S-protein), is responsible for the crown-like shape of the CoV viral particles. Due to this crown-like shape of this viral particle, it is named as "coronavirus." The spike surface glycoprotein of coronavirus plays a vital role of its attachment to the receptors on the host cell, receptor binding, and for mediating host cell membrane and viral membrane fusion during infection.[10] The trimetric surface glycoprotein S-protein is treated at the S1/S2 cleavage site by host cell proteases, during its infection. Subsequent S1/S2 cleavage, protein is divided into an N-terminal S1-ectodomain that recognizes a similar cell surface receptor and a C-terminal S2membrane-anchored protein involved in viral entry into the host cell.[11] The crucial determinant of the host specificity of a coronavirus is the surface trimeric spike (S) glycoprotein, which was believed to be transmitted from a natural host, possibly originating from bats to humans through some intermediate mammalian hosts.[12] The severe acute respiratory syndrome coronavirus (SARS-CoV) S1-protein contains a conserved Receptor Binding Domain (RBD), which recognises the angiotensin-converting enzyme 2 (ACE2).[13]

Inadequate evidence is available, which shows that fragments of the SARS-CoV-2 are recognized by the human immune responses. Based on the current information of other coronaviruses such as SARS-CoV and MERS-CoV, it was concluded that surface glycoprotein (spikes) is the main virulence and pathogenic factor in coronavirus.[2, 14, 15] Surface glycoprotein might cause good antigenicity and could be induced for the production of monoclonal antibodies. These properties of surface glycoprotein make it a suitable candidate for the development of monoclonal antibodies for therapeutic and diagnostic applications. Thus, developing a useful monoclonal antibody is an effective strategy for the diagnostic and treatment of SARS-CoV-2. Hence, in the present study by using the computational approach, we tried to identify a potential B-cell epitope of novel coronavirus 2019 (SARS-CoV-2) surface glycoprotein for the therapeutic and diagnostic application.

## Materials and methods

### Sequence identification and retrieval

The primary amino acid sequence of the surface glycoprotein of SARS-CoV-2 was retrieved from the National Centre for Biotechnological Information (NCBI) database using GenBank number QHS34546.1. The identified surface (spike) glycoprotein has 1272 long amino acid sequences, and its sequence was download in FASTA format for further study.

### Structural analysis of surface glycoprotein

The primary protein sequence of surface glycoprotein on SARS-CoV-2 was analyzed by online tool Protparam to determine its chemical and physical properties.[16] Further, TMpred Server-EMBnet, CCTOP server and TMHMM v2.0 servers were used to predict the transmembrane topology analysis of surface glycoprotein on SARS-CoV-2.[17-19] The Secondary structure analysis of surface glycoprotein on SARS-CoV-2 was carried out by PSIPRED.[20] Finally, antigenicity prediction on the full-length Surface glycoprotein was predicted by vaxijen v2.0.

### Prediction of B-cell epitopes

The web tool IEDB (Immune-Epitope-Database and Analysis-Resource) was used for the prediction of continuous (linear) B-cell epitopes.[21] Antibody epitope prediction was performed by the IEDB analysis resources such as Bepipred 2.0 tool for linear B-cell epitope prediction, Chou and Fasman tool for beta-turn prediction, Emini tool for surface accessibility prediction, Karplus and Schulz tool for flexibility prediction, Kolaskar and Tongaonkar tool for antigenicity and Parker tool for hydrophilicity. ABCprep server was also used to predict 16mer amino acid overlap of the candidate epitopes, and VaxiJen 2.0 server was used to evaluate the antigenicity of the predicted epitopes.[22]

## Results

The occurrence of SARS CoV-2 is a severe health threat to human society. Therefore, there is an urgent need for its diagnostic, drugs, vaccine and preventative measures.[23, 24] According to the information from CDC (centers for disease control and prevention) and WHO (World Health Organization), the COVID-19 symptoms can appear with two days or as long as 14 days after exposure of SARS CoV-2. The disease is transmitted in human through direct

touch, droplets, or little bits of liquid, mostly via the sneezing or coughing from contact with infected air, surfaces and objects.[25, 26] The SARS-CoV-2 infection is characterized by lung infections with symptoms, including fever, cough, and shortness of breath.[27] The First step of SARS CoV-2 viral infection is receptor recognition, and it is a critical determinant of the host cell and tissue tropism.[15] Thus, SARS CoV-2 can be spread easily in humans-to-human transmission events reported to date. Due to its new emerging viral disease, currently, there is not any specific antiviral drugs or vaccines, and no specific monoclonal antibodies are available for its diagnostics and therapeutics purpose. Thus, there is an instantaneous need to rapidly design and develop a computer-based method or study to identify immune epitopes for the development of surface glycoprotein SARS-CoV-2-based monoclonal antibodies. Thus, here in this immunoinformatics study, B-cell epitope was predicted using the strain of severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/29/human/2020/IND using the computational tools and databases.

## Structural analysis of surface glycoprotein

Protparam server was used to analyze the primary structural properties of surface glycoprotein on SARS-CoV-2, obtained results revealed that protein contains 1272 amino acids residues having a molecular weight of 140.97 kDa and theoretical iso-electric point (PI) was found to be 6.16. The total of 19685 atoms was formulated as $C_{6327}H_{9760}N_{1652}O_{1892}S_{54}$ along with GRAVY (grand average of hydropathicity) value of -0.071 and aliphatic-index was found to be 85.05, which predicts this surface glycoprotein consists of a proportional volume of the aliphatic side chain.

Further, PSIPRED 4.0 server was used to predict the secondary structural properties of surface glycoprotein on SARS-CoV-2, and the obtained results showed that proteins secondary structure consists of the beta-strand, helixes, and coil. The transmembrane protein topology of surface glycoprotein was predicted by using the servers TMpred Server-EMBnet, CCTOP server and TMHMM v2.0 respectively. TMpred server predicts the membrane-spanning region and their orientation in the proteins based on the statistical analysis of TMbase database.[19] The obtained results predict two models based on the assumption that all transmembrane helices have been found in the surface glycoproteins of SARS-CoV2. The first identified model revealed that N-terminus site is outside the membrane and four strong transmembrane helices are present.

TABLE I.—*TMpred Server models for transmembrane topology.*

| S. No. | Sequence | From | To | Length | Score |
|---|---|---|---|---|---|
| Model: 1 | | | | | |
| 1. | MFVFLVLLPLVSSQCVNLT | 1 | 19 | 19 | 1140 |
| 2. | VLHSTQDLFLPFFSNVTWFHAI | 47 | 68 | 22 | 543 |
| 3. | AGTITSGWTFGAGAALQIPFAM | 878 | 899 | 22 | 1138 |
| 4. | IWLGFIAGLIAIVMVTIMLC | 1215 | 1234 | 20 | 2682 |
| Model: 2 | | | | | |
| 1. | MFVFLVLLPLVSSQCVNLT | 1 | 19 | 19 | 1140 |
| 2. | AGTITSGWTFGAGAALQIPFAM | 878 | 899 | 22 | 1138 |
| 3. | IWLGFIAGLIAIVMVTIMLC | 1215 | 1234 | 20 | 2682 |

In the case of second model 3 transmembrane helices helix is predicted (Table I).

Further, the results obtained from the CCTOP (Constrained Consensus TOPology) prediction server based on the localization of membrane-spanning region and segments between them, it infers that $_{1215}$IWLGFIAGLIAIVMVTIMLCCMT$_{1237}$ amino acid sequence was found in the α-helical transmembrane of surface glycoprotein. TMHMM server was used to predict the transmembrane region helix in the given protein. The result obtained from the TMHMM servers predicts that amino acid residue from 1 to 1212 was found to exposed on the outside or surface, amino acid residues from 1213 to 1235 ($_{1213}$WYIWLGFIAGLIAIVMVTIMLCC$_{1235}$) were found in the transmembrane region and amino acid residues from 1237 to 1272 were found in the interior side (Figure 1).

## Prediction of B-cell epitopes

The web tool IEDB (Immune-Epitope-Database And Analysis-Resource) was used to predict the continuous (linear) B-cell epitopes. Antigenicity analysis was performed on full-length SARS-CoV-2 surface glycoprotein with the

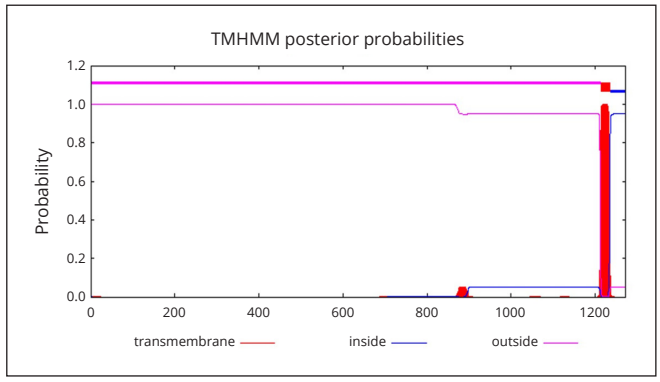


Figure 1.—Prediction of transmembrane helices in SARS-CoV-2 surface glycoprotein (X-axes, shows residue positions in the sequence; Y-axes, shows the probability of transmembrane helices).

TABLE II.—*B-cell epitopes prediction by ABCprep server.*

| S. No. | Sequence | Start position | End Position | Score | Antigenicity |
|---|---|---|---|---|---|
| 1. | VSSQCVNLTTRTQLPP | 11 | 26 | 0.65 | 1.409 |
| 2. | PFLMDLEGKQGNFKNL | 173 | 188 | 0.65 | 1.0418 |
| 3. | FSTFKCYGVSPTKLND | 373 | 388 | 0.76 | 0.965 |
| 4. | PTKLNDLCFTNVYADS | 383 | 398 | 0.62 | 1.2303 |
| 5. | EVIQIAPGQTGKIADY | 405 | 420 | 0.86 | 1.2316 |
| 6. | TGKIADYNYKLPDDFT | 414 | 429 | 0.84 | 0.9642 |
| 7. | NGVGYQPYRVVVLSFE | 500 | 515 | 0.77 | 0.9736 |
| 8. | TGTGVLTESNKKFLPF | 546 | 561 | 0.73 | 0.9925 |
| 9. | TDAVRDPQTLEILDIT | 572 | 587 | 0.67 | 0.9832 |
| 10. | EILDITPCSFGGVSVI | 582 | 597 | 0.8 | 1.3971 |
| 11. | TISVTTEILPVSMTKT | 718 | 733 | 0.77 | 1.2262 |
| 12. | FAMQMAYRFNGIGVTQ | 897 | 912 | 0.88 | 1.3096 |
| 13. | CVLGQSKRVDFCGKGY | 1031 | 1046 | 0.61 | 1.0648 |
| 14. | DSFKEELDKYFKNHTS | 1145 | 1160 | 0.76 | 0.9728 |
| 15. | HTSPDVDLGDISGINA | 1158 | 1173 | 0.66 | 1.0488 |
| 16. | DEDDSEPVLKGVKLHY | 1256 | 1271 | 0.64 | 1.0101 |

help of VaxiJen v2.0 server with the default threshold value 0.4; it predicts that in the surface glycoprotein there is a probable antigen with the antigenicity score of 0.4687. ABCprep server was used to predict the B-cell epitopes in an antigen sequence using an artificial neural network; it predicts the total number of 132 linear 16mer amino acid overlap sequence of the B-cell epitopes. Finally, 16 B-cell epitopes with overlap 16mer amino acids were selected based on its antigenicity, which was evaluated through VaxiJen 2.0 (Table II).

Immune Epitope Database (IEDB) was used for the linear B-cell epitopes prediction, beta-turn prediction, surface accessibility prediction, flexibility prediction, antigenicity prediction and hydrophilicity prediction analysis. These identified several characteristics make them SARS-CoV-2 surface glycoprotein as an attractive immunogenic candidate.

BepiPred 2.0 server was used to predict B cell epitope from the antigen sequence based on the Random Forest algorithm.[28] The amino acid residues of SARS-CoV-2 surface glycoprotein with scores above the threshold (0.5) are predicted as B-cell epitope. However, based on the antigenicity score and 11-40 amino acids residue length, twenty-one epitopes were selected (Table III, Figure 2A).

The BepiPred 1.0 linear epitope prediction generates a good number of B-cell epitopes constructed on sequence characteristics of the antigen using a combination of propensity amino acid scales method and Hidden Markov Model (HMM) method.[29] The amino acid residues of SARS-CoV-2 surface glycoprotein with scores above the threshold (0.35) was predicted to be the part of B-cell epitope (Figure 2B).

Chou and Fasman tool is an empirical technique used to predict the secondary structures in proteins, based on the analyses of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns.[30] Beta turns have a significant effect in inducing antigenicity. Consequently, Chou and Fasman predicted beta-turn result shown as,

TABLE III.—*B-cell epitopes prediction by BepiPred 2.0 tool.*

| S. No. | Sequence | Start position | End position | Length |
|---|---|---|---|---|
| 1. | SQCVNLTTRTQLPPAYTNSFTRGVY | 13 | 37 | 25 |
| 2. | FSNVTWFHAIHVSGTNGTKRFDN | 59 | 81 | 23 |
| 3. | FLGVYHKNNKSWMESEFRVYSSANN | 140 | 164 | 25 |
| 4. | LMDLEGKQGNFKNL | 175 | 188 | 14 |
| 5. | PINLVRDLPQGFS | 208 | 220 | 13 |
| 6. | LTPGDSSSGWTA | 248 | 259 | 12 |
| 7. | KSFTVEKGIYQTSNFRVQP | 303 | 321 | 19 |
| 8. | FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA | 328 | 362 | 35 |
| 9. | YNSASFSTFKCYGVSPTKLNDLCFT | 368 | 392 | 25 |
| 10. | DEVIQIAPGQTGKIADYNYKLP | 404 | 425 | 22 |
| 11. | ELLHAPATVCGPKKSTNLVKN | 515 | 535 | 21 |
| 12. | NCTEVPVAIHADQLTPT | 615 | 631 | 17 |
| 13. | RVYSTGSNVFQ | 633 | 643 | 11 |
| 14. | VNNSYECDIPI | 655 | 665 | 11 |
| 15. | ASYQTQTNSPRRARSVASQ | 671 | 689 | 19 |
| 16. | AYTMSLGAENSVAYSNN | 693 | 709 | 17 |
| 17. | KQIYKTPPIKDFGGF | 785 | 799 | 15 |
| 18. | LADAGFIKQYGDCLG | 827 | 841 | 15 |
| 19. | RNFYEPQIITTD | 1106 | 1117 | 12 |
| 20. | VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI | 1132 | 1171 | 40 |
| 21. | SCCKFDEDDSEPVLKGVK | 1251 | 1268 | 18 |

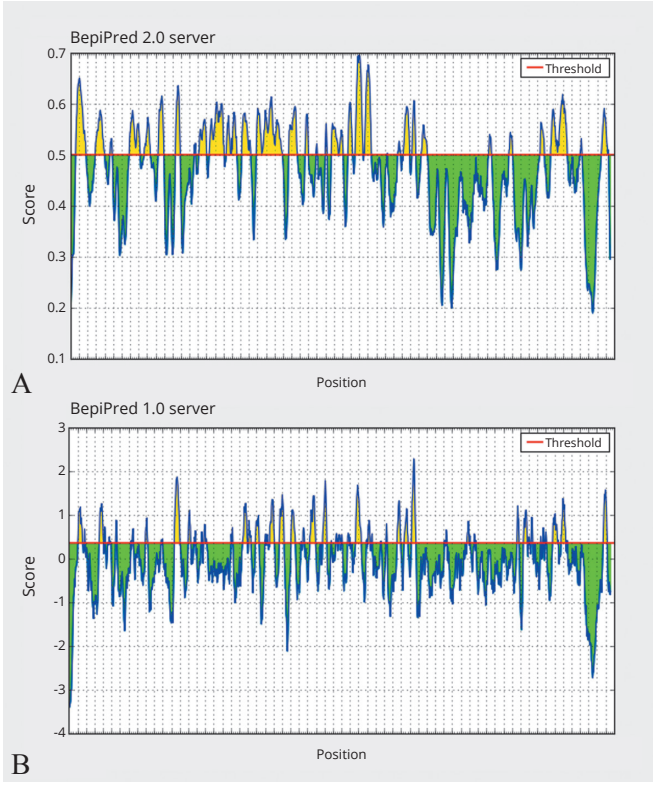


Figure 2.—A) BepiPred 2.0 server predicts B-cell epitopes from a SARS CoV-2 surface glycoprotein (Y-axes, depicts residue scores; X-axes, residue positions in the sequence). The amino acid residues with scores above the threshold (0.5) are predicted as an epitope and are above the line in the graph (colored in yellow in the online version); B) BepiPred 1.0 linear B-cell epitopes prediction from a SARS CoV-2 surface glycoprotein (Y-axes, depicts residue scores; X-axes, residue positions in the sequence). The amino acid residues with scores above the threshold (0.35) are predicted as an epitope (above the line in the graph, colored in yellow in the online version).

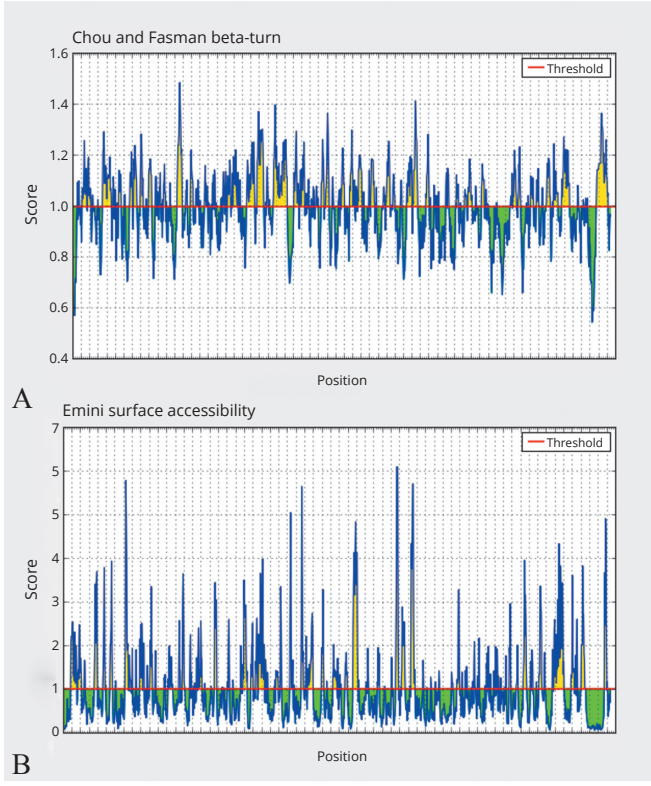


Figure 3.—A) Chou and Fasman beta-turn prediction of the SARS CoV-2 surface glycoprotein where Y-axes and X-axes represent the position and score, respectively and the regions having beta turns in the protein are shown above the line of the threshold value (colored in yellow in the online version); B) Emini surface accessibility prediction of a SARS CoV-2 surface glycoprotein (Y-axes, represents residue scores; X-axes, residue positions in the sequence). The amino acid residues with scores above the threshold (1.0) are predicted as surface accessibility of protein (above the threshold line, colored in yellow in the online version).

minimum (0.541), maximum (1.484) and average (0.997) in SARS-CoV-2 surface glycoprotein (Figure 3A).

Surface accessibility of SARS-CoV-2 surface glycoprotein B cell epitopes prediction is necessary for the hydrophilic regions, which were mostly exposed on the surface and induced the B cell immune response.[31] Thus, surface accessibility prediction was determined by the Emini server, which predicts the surface accessibility score as a minimum of 0.043, and a maximum of 6.091 (Figure 3B).

Karplus and Schulz flexibility prediction tool were used to predict the chain flexibility in protein to select the peptide antigen based on the mobility of protein segments based on the known temperature β factors of the α-carbon of proteins.[32] The obtained results for SARS-CoV-2 surface glycoprotein in linear B-cell epitopes revealed as minimum (0.876), maximum (1.125) and average (0.993) scores. Above peaks from threshold score value, 1.0 represents a positive score for flexibility throughout the primary amino acid sequence in the surface glycoprotein epitopes (Figure 4A).

Kolaskar and Tongaonkar tool were used with threshold value 0.866 to predict the antigenicity prediction in the identified epitopes by analyzing the physicochemical properties of the surface glycoprotein amino acids.[33] The obtained results showed that an average antigenic propensity value of 1.041 was found for the surface glycoprotein with the maximum value of 1.261 and minimum of 0.866 (Figure 4B).

Parker hydrophilicity prediction tool was used with threshold value 1.231 to predict the hydrophilic regions present in the surface of the SARS-CoV-2 surface glycoprotein. The prediction was based on the hydrophilic scale

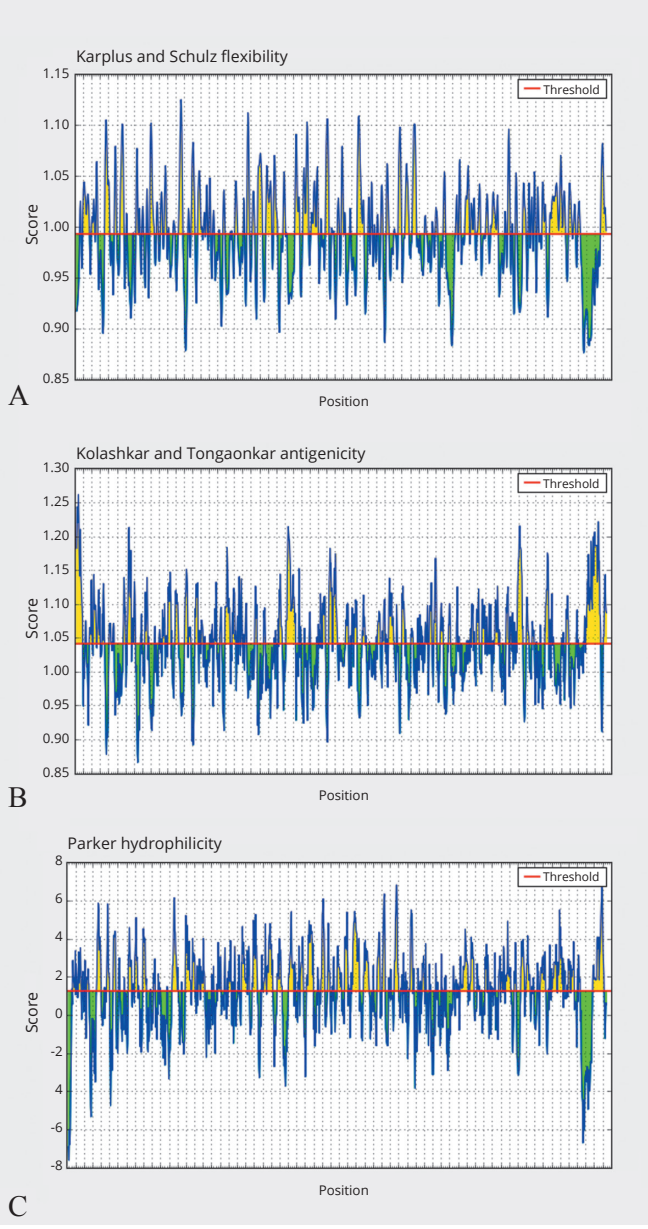


Figure 4.—A) Karplus and Schulz flexibility prediction of a SARS CoV-2 surface glycoprotein (Y-axes, represents residue scores; X-axes, shows residue positions in the sequence). The amino acid residues with scores above the threshold (0.993) are predicted as the flexibility of protein and coloured in yellow on the graph; B) Kolashkar and Tongaonkar antigenicity prediction of a SARS CoV-2 surface glycoprotein (Y-axes, represents residue scores; X-axes, shows residue positions in the sequence). The amino acid residues with scores above the average (1.041) are predicted as antigenicity and are above the line in the graph (colored in yellow in the online version); C) Parker hydrophilicity prediction of a SARS CoV-2 surface glycoprotein (Y-axes, represents residue scores; X-axes shows residue positions in the sequence). The amino acid residues with scores above the average (1.231) are predicted as hydrophobicity of protein and are above the line in the graph (colored in yellow in the online version).

of peptide retention times during the high-performance liquid chromatography (HPLC) on a reversed-phase column.[34] The result showed the average, minimum and maximum value of hydrophilicity was to be 1.231, -7.269 and 7.73, respectively (Figure 4C).

## Conclusions

SARS-COV-2 is the infectious pathogen for severe respiratory diseases, more than five million people are infected from this severe pandemic disease from all over the world. Researchers from across the globe laboratories are struggling to search for a solution of this pandemic outbreak COVID-19. B-cell epitopes are the significant antigenic sites in the virus structural proteins, which are essential in the antibody production and cell-mediated immune response against the viruses. So, there is an urgent need for the identification of B-cell epitopes which were used for developing the epitope-based therapeutics and diagnostic study for COVID-19. In this study, using the computational immunoinformatics approach, we tried to identify the various B-cell epitopes against SARS CoV-2 surface glycoprotein. SARS CoV-2 surface glycoprotein analysis for B-cell epitope prediction in the ABCprep server and IEDB server resulted in the identification of 16 and 21 linear B-cell epitopes in the surface glycoprotein. Subsequently, SARS CoV-2 surface glycoprotein and its B-cell epitopes sequence were also analyzed for their antigenicity, beta-turn, surface accessibility, flexibility and hydrophilicity properties. Fortunately, we identified epitopes which can be used possibly for the monoclonal antibodies production. However, our computational identified B-cell epitopes prediction analyses need numerous *in-vitro* and *in-vivo* validations before formulating the monoclonal antibodies for the therapeutics and diagnostics purpose of COVID-19.

## References

**1.** WHO. Novel coronavirus – China; 2020 [Internet]. Available from: http://Www.Who. Int/Csr/Don/12-January-2020-Novel-Coronavirus-China/En/ [cited 2020, Jun 15].

**2.** Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 2015;1282:1–23.

**3.** Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, *et al.* A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–9.

**4.** WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 126; 2020 [Internet]. Available from: https://Www.Who.Int/Docs/Default-Source/Coronaviruse/Situation-Reports/20200525-Covid-19-Sitrep-126.Pdf?Sfvrsn=887dbd66_2 [cited 2020, Jun 15].

**5.** Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.

**6.** Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, *et al.*; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020;5:536–44.

**7.** WHO. WHO Director-General's opening remarks at the media briefing on COVID-19; 2020 [Internet]. Available from: https://www.Who.Int/Dg/Speeches/Detail/Who-Director-General-s-Opening-Remarks-at-the-Media-Briefing-on-Covid-19---11-March-2020 [cited 2020, Jun 2020].

**8.** Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019;17:181–92.

**9.** Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, *et al.* Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 2020;27:325–8.

**10.** Zhu Z, Zhang Z, Chen W, Cai Z, Ge X, Zhu H, *et al.* Predicting the receptor-binding domain usage of the coronavirus based on kmer frequency on spike protein. Infect Genet Evol 2018;61:183–4.

**11.** Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res 2020;176:104742.

**12.** Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol 2015;23:468–78.

**13.** Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 2003;426:450–4.

**14.** Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. Virology 2001;279:371–4.

**15.** Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J Virol 2020;94:e00127–20.

**16.** Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al.* Protein Identification and Analysis Tools on the ExPASy Server BT - The Proteomics Protocols Handbook. In: Walker JM. Totowa, NJ: Humana Press; 2005. p. 571–607.

**17.** Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–80.

**18.** Dobson L, Reményi I, Tusnády GE. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res 2015;43(W1):W408-12.

**19.** Hofmann K, Stoffel W. TMbase: A Database of Membrane Spanning Protein Segments, Biological Chemistry Hoppe-Seyler; 1993 [Internet]. Available from: https://books.google.co.in/books?id=7EX8jwEACAAJ [cited 2020, Jun 15].

**20.** Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.

**21.** Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, *et al.* IEDB-AR: immune epitope database-analysis resource in 2019. Nucleic Acids Res 2019;47(W1):W502–6.

**22.** Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4.

**23.** Pang J, Wang MX, Ang IY, Tan SH, Lewis RF, Chen JI, *et al.* Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review. J Clin Med 2020;9:9.

**24.** Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, *et al.* Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Cent Sci 2020;6:315–31.

**25.** Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents 2020;55:105924.

**26.** Ong SW, Tan YK, Chia PY, Lee TH, Ng OT, Wong MS, *et al.* Air, Surface Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient. JAMA 2020;323:1610–2.

**27.** Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, *et al.* Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ 2020;368:m606.

**28.** Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 2017;45(W1):W24–9.

**29.** Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res 2006;2:2.

**30.** Chou PY, Fasman GD. Prediction of beta-turns. Biophys J 1979;26:367–83.

**31.** Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 1985;55:836–9.

**32.** Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften 1985;72:212–3.

**33.** Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 1990;276:172–4.

**34.** Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry 1986;25:5425–32.

---

*Conflicts of interest.*—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

*Authors' contributions*.—Rahul Shrivastava conceived of the presented idea and supervised the studies; Pushpendra Singh and Manish K. Tripathi contributed equally, developed the theory and performed the computations. All authors read and approved the final version of the manuscript.

*History.*—Manuscript accepted: June 12, 2020. - Manuscript received: May 28, 2020.