



Exploration of antigenic determinants in spike glycoprotein of SARS-CoV2 and identification of five salient potential epitopes

Aditya Agrawal¹ · Rajat Varshney² · Mamta Pathak¹ · Shailesh Kumar Patel¹ · Vishal Rai¹ · Sourabh Sulabh³ · Rohini Gupta⁴ · Khushal Singh Solanki¹ · Ritu Varshney⁵ · Ramadevi Nimmanapalli²

Received: 5 March 2021 / Accepted: 7 August 2021 / Published online: 7 September 2021
© Indian Virological Society 2021

Abstract Emerging pathogens have been an eternal threat to mankind. In a series of pandemics caused by notorious coronaviruses, a newly emerged SARS-CoV2 virus is creating panic among the world population. The unavailability of reliable theranostics insists the exploration of antigenic determinants in spike glycoprotein of SARS-CoV2. The four novel inserts (⁷⁰VSGTNGT⁷⁶, ¹⁵⁰KSWM¹⁵³, ²⁴⁷SYLTPG²⁵² and ⁶⁷⁴QTQTNSPRR⁶⁸²) in SARS-CoV2 spike protein were unraveled via multiple sequence alignment of spike proteins of SARS-CoV2, SARS-CoV, and MERS-CoV. The three-dimension (3D) modeling of the spike protein of the SARS-CoV2 and their interaction with the ACE2 receptor was delineated with the help of SWISS-MODEL and 3DLigandSite web servers. The predicted 3D model of SARS-CoV2 was further verified by SAVES, RAMPAGE, and ProSA-web tools. The potential B-cell immunogenic epitopes of SARS-CoV2 were predicted out by

using various software viz. IEDB B-cell epitopes prediction tool, BepiPred linear epitope prediction tool, Emini Surface Accessibility Prediction tool, and Kolaskar-Tongaonkar antigenicity web tool. The five epitopes (*i.e.* ⁷¹SGTNGTKRFDN⁸¹, ²⁴⁷SYLTPG²⁵², ⁶³⁴RVYST⁶³⁸, ⁶⁷⁵QTQTNSPRRARSV⁶⁸⁷, and ¹⁰⁵⁴QSAPH¹⁰⁵⁸) were selected as potent antigenic determinants. The quantum of information generated by this study will prove beneficial for the development of effective therapeutics, diagnostics, and multi-epitopic vaccines to combat this ongoing menace.

Keywords SARS-CoV2 · Spike protein · ACE2 receptor · Vaccines · Epitopes

Introduction

The first case of SARS-CoV2 had been documented on 31 December 2019 by the health commission of Hubei province, China [34]. This has been the third straight decade in which a coronavirus has traversed species barriers to cause disease in the human population [26]. The symptoms include pneumonia in affected individuals along with fever, sneezing, dyspnea, myalgia, and chest pain. The novel virus has been associated with pneumonia in affected individuals and clinically exhibited by fever, sneezing, dyspnea, myalgia, and chest pain [2, 15, 37].

The SARS-CoV2 is a member of *Coronaviridae*, a family of single-stranded positive-sense RNA viruses [9]. It belongs to the subgenus *Sarbecovirus*, under the genus betacoronavirus of this family [41]. The genome of the virus was reported to be 29.891 kb in size and arranged in the order: 5'-orf1/ab (replicase genes)-Structural genes (S, E, M, N)-3' [7]. The S gene encodes the spike glycoprotein of coronaviruses, which is responsible for binding to the

✉ Ritu Varshney
rituvarshney.2009@gmail.com

✉ Ramadevi Nimmanapalli
ramadevi.nimmanapalli@gmail.com

¹ ICAR-Indian Veterinary Research Institute Izatnagar, Bareilly, Uttar Pradesh 243122, India

² Department of Veterinary Microbiology, Faculty of Veterinary and Animal Sciences, IAS, RGSC, Banaras Hindu University, Mirzapur, Uttar Pradesh 231001, India

³ Department of Animal Science, Kazi Nazrul University, Asansol, West Bengal 713340, India

⁴ Department of Veterinary Medicine, Nanaji Deshmukh Veterinary Science University, Jabalpur, Madhya Pradesh 482001, India

⁵ Department of Biological Engineering, Indian Institute of Technology Gandhinagar, Palaj, Gujarat 382355, India

host cell receptor to initiate infection [21]. The spike glycoprotein is made up of two subunits, viz. S1 and S2. N-terminal domain (NTD) and receptor-binding domain (RBD) makes up the S1 subunit whereas fusion peptide (FP), heptad repeat (HR) 1&2, transmembrane domain (TM), and a cytoplasmic domain (CP) constitutes the S2 subunit [7]. The role of the S1 subunit as receptor binding and the S2 subunit role in facilitating membrane fusion have been witnessed [6]. The SARS-CoV2 utilizes the same receptor [*i.e.* the angiotensin-converting enzyme 2 (ACE2)] as that of severe acute respiratory syndrome coronavirus (SARS-CoV) [24, 31] but bind with higher affinity in comparison to SARS-CoV. Being the major structural protein, neutralizing antibodies are directed against the spike glycoprotein and thus they have been described as potent targets for the development of vaccines against both Middle East respiratory syndrome coronavirus (MERS-CoV) and SARS-CoV [12, 40]. Similarly, they can be targeted for the development of novel vaccines against SARS-CoV2 as well.

Bioinformatics tools for the prediction of epitopes in a viral protein can have versatile applications such as the development of immunoassays, synthetic antiviral peptides, and vaccines [28]. Prediction of antigenic determinants for the development of peptide-based vaccines for other diseases has been done earlier [27, 34]. The same can be applied in the case of SARS-CoV2, especially for the improvement in the existing vaccines which is the need of the hour in the wake of this global emergency. Therefore, this study was undertaken to predict and analyze antigenic determinants present on the spike glycoprotein of SARS-CoV2 using different bioinformatics tools.

Material and methods

Retrieval, alignment and conservation analysis of sequences of coronaviruses and ACE2 receptor

Using NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein/>), the spike glycoprotein amino acid sequences of SARS-CoV2 (1273 AA; NCBI reference sequence number: NC_045512.2), SARS-CoV (1255 AA; GenBank accession number: BAC81404.1) and MERS-CoV (1353 AA; GenBank accession number: AMO03401.1) and along with ACE2 receptor sequences of *Homo sapiens* (805 AA; NCBI reference sequence number: NP_001358344.1) were retrieved in FASTA format. To identify the similarity between the spike proteins of these coronaviruses, Clustal Omega multiple sequence alignment tool (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [23] was employed.

The 3D modeling of SARS-CoV2 Spike protein and delineation of its interaction with ACE2 receptor

The 3D structure assessments from retrieved amino acid sequences of the spike protein of SARS-CoV2 and ACE2 receptor were performed using protein structure homology modelling namely SWISS-MODEL web server (<https://swissmodel.expasy.org/interactive>) [5]. The ligand-binding position in the 3D structure of the ACE-2 receptor was determined by the 3DLigandSite online web server (<http://www.sbg.bio.ic.ac.uk/3dligandsite>) [32].

Analysis and validation of model quality

The predicted 3D model of SARS-CoV2 was analysed by ERRAT and VERIFY3D tools of SAVES (Structure Analysis and Verification Server) (<http://services.mbi.ucla.edu/SAVES/>) and ProSA-web tool (<https://prosa.services.came.sbg.ac.at>) based on their physio-chemical properties. Inappropriate amino acid residues in models were determined by ERRAT based on atomic interaction [10] whereas the quality and compatibility of models were evaluated using electron density of atomic interaction [8]. Also, the estimation of model quality was done by ProSA-web tool (<https://prosa.services.came.sbg.ac.at>) which utilizes the energy plot generated by atomic interaction to design the residues z-score [33]. Model excellence was verified by determining the amino acid residues of different regions by RAMPAGE web tool (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) based on their phi-psi torsion angles [22].

Predictions of continuous B cell epitopes

To predict the surface antigenic epitopes in the spike protein of SARS-CoV2, the Immune Epitope Database (IEDB) webserver (<http://tools.iedb.org/bcell/>) [11] possessing a variety of immunoinformatics prediction tools which recognize the immunogenic amino acid sequences based on their physio-chemical properties were utilized. BepiPred linear epitope prediction tool (<http://tools.iedb.org/bcell/>) [20] was applied to find out the list of continuous B cell epitopes from the retrieved sequence of SARS-CoV2 spike protein. The surface assessable nature and antigenicity of the predicted peptide sequences were determined by Emini Surface Accessibility Prediction tool (<http://tools.iedb.org/bcell/results>) [13] and Kolaskar-Tongaonkar antigenicity web tool (<http://tools.immuneepitope.org/bcell/>) [19] respectively. To check the maximum retention period during HPLC, the hydrophilic or hydrophobic behaviour of the predicted peptide sequences of SARS-CoV2 spike protein was measured on the hydrophilicity scale using

Parker Hydrophilicity Prediction method (<http://tools.immuneepitope.org/bcell/>) [25]. The predicted surface immunogenic sequences of the SARS-CoV2 spike protein were docked with the 3D model of ACE2 receptor by HPEPDOCK web server (<http://huanglab.phys.hust.edu.cn/hpepdock/>) [39], which utilizes hierarchical algorithm.

Results

Comparison of the spike protein of SARS-CoV2 with SARS-CoV and MERS-CoV

The multiple sequence alignment (MSA) between the spike protein of SARS-CoV2, SARS-CoV, and MERS-CoV revealed 77.38% and 31.98% similarity of the virus with SARS-CoV and MERS-CoV respectively. The spike protein of SARS-CoV2 was found to be highly similar to SARS-CoV, however, four regions of novel inserts were found. Furthermore, these novel inserts viz. ⁷⁰VSGTNGT⁷⁶, ¹⁵⁰KSWM¹⁵³, ²⁴⁷SYLTPG²⁵² and ⁶⁷⁴QTQTNSPRR⁶⁸² were found unique for SARS-CoV2 (Fig. 1).

The 3D modeling of SARS-CoV2 Spike protein and delineation of its interaction with ACE2 receptor

The 3D homology structure of SARS-CoV2 spike protein created by the Swiss model revealed 76.47% sequence identity with SARS-CoV using 6acc.1A [PDB title: SARS-CoV spike glycoprotein and ACE2 complex] as a template. Besides, the final model was prepared with the overall coverage of 0.93, the sequence similarity of 0.54 along with global quality estimation values of -3.63 , -1.99 , and -2.69 for QMEAN, solvation, and torsion respectively (Fig. 2). Similarly, ACE2 receptor homology model was designed with an overall coverage of 0.74 and a sequence similarity of 0.62 using 2ajf.1 [PDB title: Structure of SARS coronavirus spike receptor-binding domain complexed with its receptor] as a template (Fig. 3). The ligands binding position in the ACE2 receptor was observed using the 3DLigandSite web server and found that the probability of spike protein (ligand) of SARS-CoV2 to bind/interact with ACE2 receptor at these “374HIS, 378HIS, 402GLU, and 518ARG” positions are quite high (Fig. 3).

Model Quality analysis and validation of SARS-CoV2 spike protein

The quality factor values of 46.94 and 46.57 were obtained on the analysis of 3D model of SARS-CoV2 spike protein in ERRAT (Fig. 4). The VERIFY3D model analysis tool

revealed 57.03% residues had an average of $> = 0.2$ 3D-1D score in comparison to the accepted value of 80% (Fig. 4). The value of z-score analysed by the ProSA-web server was found to be -11.85 , which indicates a fair model quality along with largely negative energy of the residues (Fig. 5). Ramachandran plots generated by RAMPAGE web server revealed 83.2% amino acid residues lie in the most favoured regions while 16.1% and 0.7% of the residues fit in allowed and disallowed regions respectively (Fig. 6).

Continuous B cell epitope prediction

The surface antigenic epitopes of SARS-CoV2 spike protein were analysed by numerous prediction tools and a total of 13 continuous epitopes of different lengths were selected through the IEDB server and concluding scores of the predicted sequences are mentioned in Table 1. In BepiPred tool, the peptide sequence ⁸⁰⁵ILPDPSKPSKRS⁸¹⁶ achieved the highest score of 2.291 against the threshold value of 0.350 among all the predicted amino acid sequences (Fig. 7). The most accessible surface peptide was ⁶⁷⁵QTQTNSPRRARSV⁶⁸⁷ with a value of 8.837 in Emini Surface Accessibility scale (Fig. 7) whereas the most antigenic peptide region was ⁶³⁴RVYST⁶³⁸ with the highest value of 1.068 in Kolaskar-Tongaonkar antigenicity scale in comparison to their threshold value of 1.000 and 1.041 respectively (Fig. 7). On Parker hydrophilic scale, the epitope ⁷⁷²VEQDKNTQE⁷⁸⁰ achieved a score of 5.756 against the average score of 1.238 (Fig. 7). However, among all the predicted sequences, only two epitopes (*i.e.* ⁶³⁴RVYST⁶³⁸ and ¹⁰⁵⁴QSAPH¹⁰⁵⁸) passed all the threshold levels at each antigenic scale while three epitopic sequences (*i.e.* ⁷¹SGTNGTKRFDN⁸¹, ²⁴⁷SYLTPG²⁵² and ⁶⁷⁵QTQTNSPRRARSV⁶⁸⁷) were selected that contain novel inserts.

Interaction of ACE2 receptor with SARS-CoV2

The interaction of the ACE2 receptor with the domain of spiked protein of SARS-CoV2 was assessed via the HPEPDOCK server by using the 3D homology model of the ACE2 receptor. The exposed peptide sequences (*i.e.* ⁷¹SGTNGTKRFDN⁸¹, ²⁴⁷SYLTPG²⁵², ⁶³⁴RVYST⁶³⁸, ⁶⁷⁵QTQTNSPRRARSV⁶⁸⁷, and ¹⁰⁵⁴QSAPH¹⁰⁵⁸) of the domain of spike protein served as ligands to interact with the ACE2 receptor. During interaction (*i.e.* docking), numerous ligand binding sites of the ACE2 receptor were unearthed by using the aforesaid server; amongst which the docking site with the highest binding energy was chosen. The peptide epitope, ⁶³⁴RVYST⁶³⁸ of the domain of spike protein of SARS-CoV2 binds more efficiently with the highest docking score of -209.636 in comparison to



Fig. 1 Multiple sequence alignment between spike proteins of SARS-CoV2 (1273 AA), SARS-CoV (1255AA) and MERS-CoV (1353 AA) with highlighted four novel inserts in 2019-nCoV

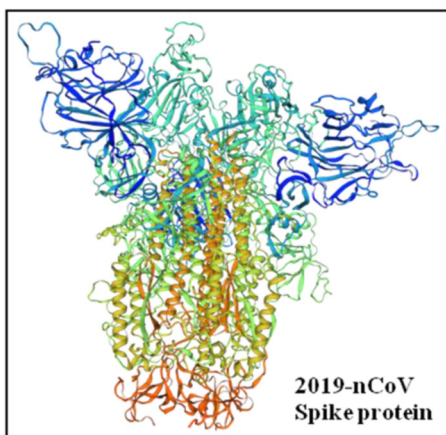


Fig. 2 Prediction of three-dimensional (3D) structure of SARS-CoV2 spike protein using reference sequence (1273 amino acids) by homology modelling software Swiss Model

epitope ⁷¹SGTNGTKRFDN⁸¹ whose docking score was -194.568 . The epitopic peptides viz. ⁶⁷⁵QTQTNSPRRRSV⁶⁸⁷, ²⁴⁷SYLTPG²⁵², and ¹⁰⁵⁴QSAPH¹⁰⁵⁸ of the domain of SARS-CoV2 were also found to dock successfully with a score of -190.391 , -189.671 , and -159.815 respectively (Fig. 8).

Discussion

Three well-known trans-species switches of coronavirus (SARS-CoV1 in 2002, MERS in 2012 and SARS-CoV2 in 2019) affected human population [26]. The subtle genetic and spike protein changes account for alteration in host range, tissue tropism and pathogenicity. Hence, the comparative study of spike protein of SARS-CoV2, SARS-CoV and MERS with MSA has been conducted. The MSA analysis revealed only 77.38% similarity in the spike

Fig. 3 Representation of 3D structure **a** and ligand binding position **b** in ACE2 receptor (*Homo Sapiens*) using NCBI reference sequence (805 amino acids) by homology modelling server (Swiss Model) and 3DligandSite server respectively

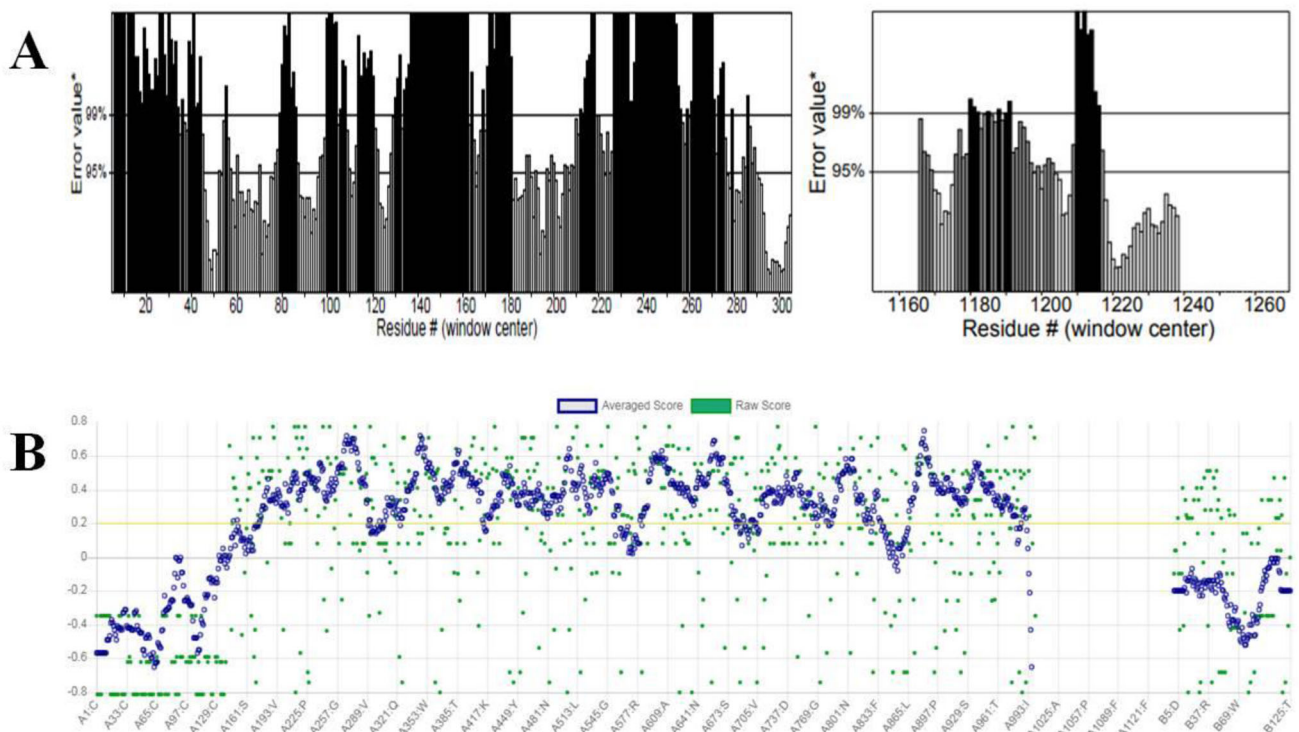
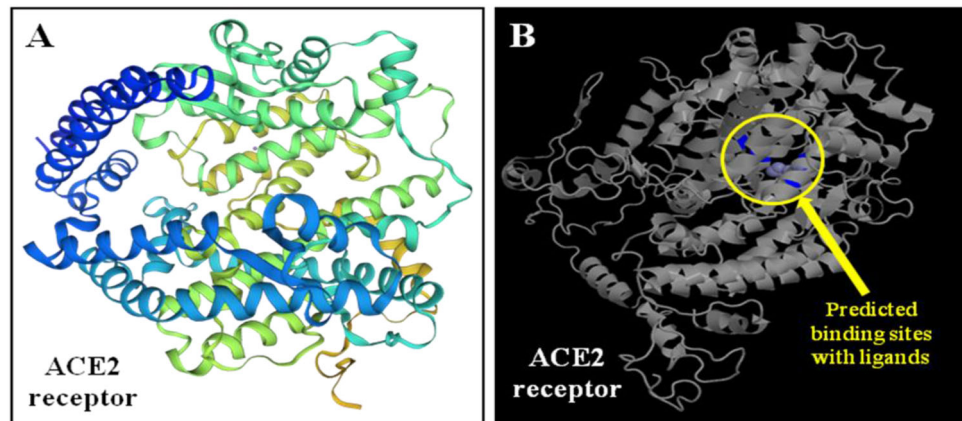


Fig. 4 Quality analysis and verification of predicted 3D model of SARS-CoV2 spike protein through ERRAT **a** and VERIFY3D **b** (The gap difference between the raw score and average score was due to atomic interaction)

protein with SARS-CoV, which gives a clue about its probable emergence and change in host preference. The change in surface proteins like spike protein might be crucial for interspecies receptor recognition on probable intermediate hosts and reservoirs facilitating the viral multiplication and assembly [4]. As per reports, SARS-CoV2 has been associated with several cross-species transmission based on homologous recombination in spike protein between SARS-CoV2 and bat coronavirus [17]. The presence of novel inserts in the spike protein of SARS-CoV2 may render the vaccine of SARS-CoV and MERS-CoV futile for combating the current SARS-CoV2

pandemic. Hence, the immunogenic profile of SARS-CoV2 needs to be explored.

The bio-informative tools implicated in the previous epitopic prediction of the spike protein of SARS-CoV and MERS-CoV [16, 28] were also included in our study to unravel immunogenic surface epitopes of SARS-CoV 2 spike protein which in turn may facilitate the development of decoy ligands/neutralizing antibodies to limit viral infection. In this study, we found four novel inserts in the SARS-CoV2 spike protein after comparing it with SARS-CoV. Consequently, the comparative analysis pointed towards the divergent immunogenic profile of SARS-CoV2

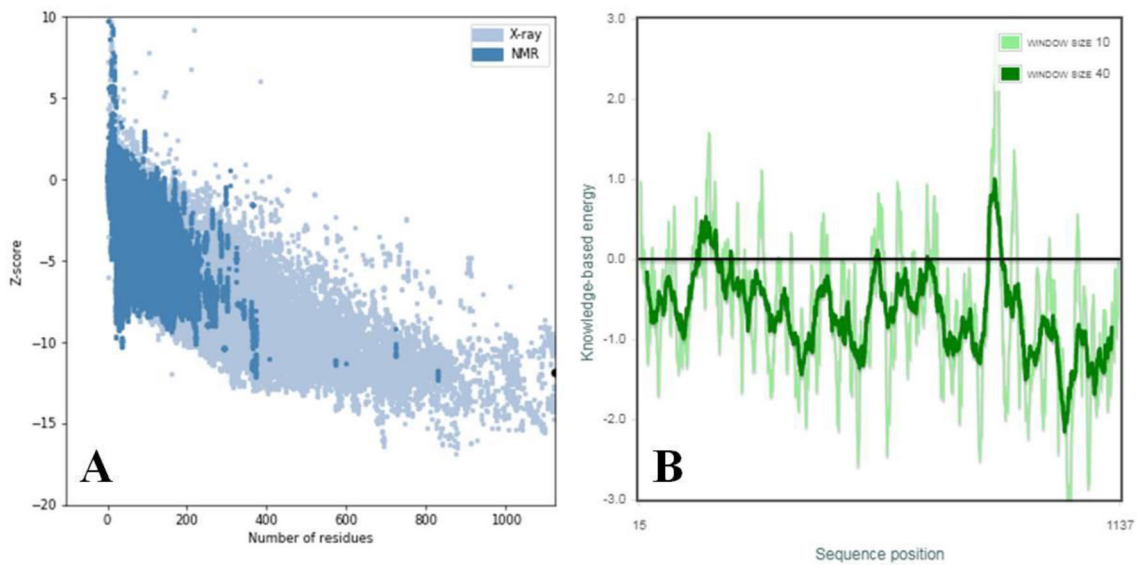


Fig. 5 Model quality analysis for the predicted 3D Model of SARS-CoV2 spike protein through ProSA-web server **a** z-scores of all protein chains in PDB determined by X-ray crystallography (light

blue) or NMR spectroscopy (dark blue) with respect to their length. The Z-scores are represented as large dots. **b** Energy plot of 2019-nCoV spike protein (color figure online)

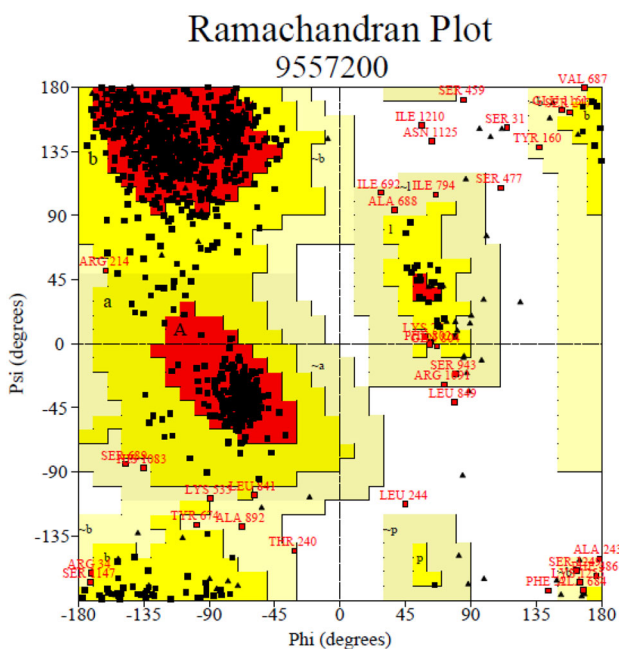


Fig. 6 Ramachandran plots of the predicted 3D model of SARS-CoV2 spike protein

with SARS-CoV A total of 13 surface epitopes of SARS-CoV2 spike protein were predicted out using different bioinformatic tools among which 5 were found potent epitopes which well supported the previous study [30].

To neutralize the virus, the presence of surface accessible potent immunogenic epitopes of virus spike protein is the prerequisite for the fabrication of neutralizing antibodies. Moreover, the binding of neutralizing antibodies

with homologous epitopes will disrupt virus spike protein interaction with the ACE2 receptor. Hence, the Emini Surface Accessibility Prediction tool and Kolaskar-Tongonkar antigenicity web tool were taken into account in the present study for assessment of the surface assessable nature and antigenicity of the predicted peptide sequences apart from merely ruling out B-cell epitopes from the retrieved sequence of SARS-CoV2 spike protein via BepiPred linear epitope prediction tool. Various bio-informative tools viz. MODPEP [36], HPEPDOCK, PEP-FOLD, etc. can be used to assess protein-peptide interaction. However, the HPEPDOCK software has been used in the present study as it was proven to be significantly better than the other peptide docking algorithms for the prediction of near-native conformations [38].

As reported recently, the ACE2 receptor is the probable target for the SARS-CoV2 facilitating its entry inside the host cells [24, 31]. In this context, successful docking between the novel inserts of RBD of SARS-CoV2 and ACE2 receptors of humans were delineated using 3D homology modelling. The peptide epitope (with the highest docking score) present on spike protein of SARS-CoV2 binds more efficiently with ACE-2 receptor in comparison to low docking score epitope. The more binding energy (negative) value consequence the compact binding. Although the structure of full-length ACE2 in complex with SARS-CoV2 has been elucidated recently [35], however, these novel predicted peptides may be explored in rationale vaccine design, immunotherapy, and various diagnostic platforms like ELISA, LFA, Dot-blot assay, etc. to detect antibodies against SARS-CoV2. The

Table 1 List of predicted immunogenic epitopes of B cell for 2019-nCoV spike protein, their position, number of residues and scores on different IEDB scales (BepiPred, Emini accessibility

prediction scale, Kolaskar and Tongaonkar antigenicity prediction scale and Parker hydrophilicity plot scale)

S. No	Peptide sequence (Epitope)	Position	No. of residues	BepiPred score (0.350)	Emini accessibility Score (1.000)	Kolaskar and Tongaonkar antigenicity Score (1.041)	Parker hydrophilicity plot Score (1.238)
1	RTQLPPAYTNS	21–32	12	1.185	2.868	1.016	1.675
2	SGTNGTKRFDN [#]	71–81	11	1.262	3.464	0.899	4.818
3	NKSWME	149–154	6	0.687	2.022	0.881	2.133
4	GKQGNF	181–186	6	0.934	1.229	0.927	3.483
5	SYLTPG [#]	247–252	6	1.309	0.995	1.045	1.400
6	YQAGSTPCNGV	473–483	11	1.358	0.444	1.049	3.282
7	TVCGPKKSTN	523–532	10	1.121	1.170	1.020	4.080
8	RVYST [*]	634–638	5	0.554	1.426	1.068	2.060
9	QTQTNSPRRARSV [#]	675–687	13	1.685	8.837	0.983	4.269
10	VEQDKNTQE	772–780	9	1.323	6.923	0.995	5.756
11	ILPDPSKPSKRS	805–816	12	2.291	4.690	1.019	2.850
12	QSAPH [*]	1054–1058	5	1.213	1.597	1.052	3.760
13	KNHTSPDVDLG	1157–1167	11	1.378	1.913	1.003	3.764

^{*}Peptides revealed a higher score than threshold in all tools.

[#]Peptide epitopes possess novel inserts in spike protein.

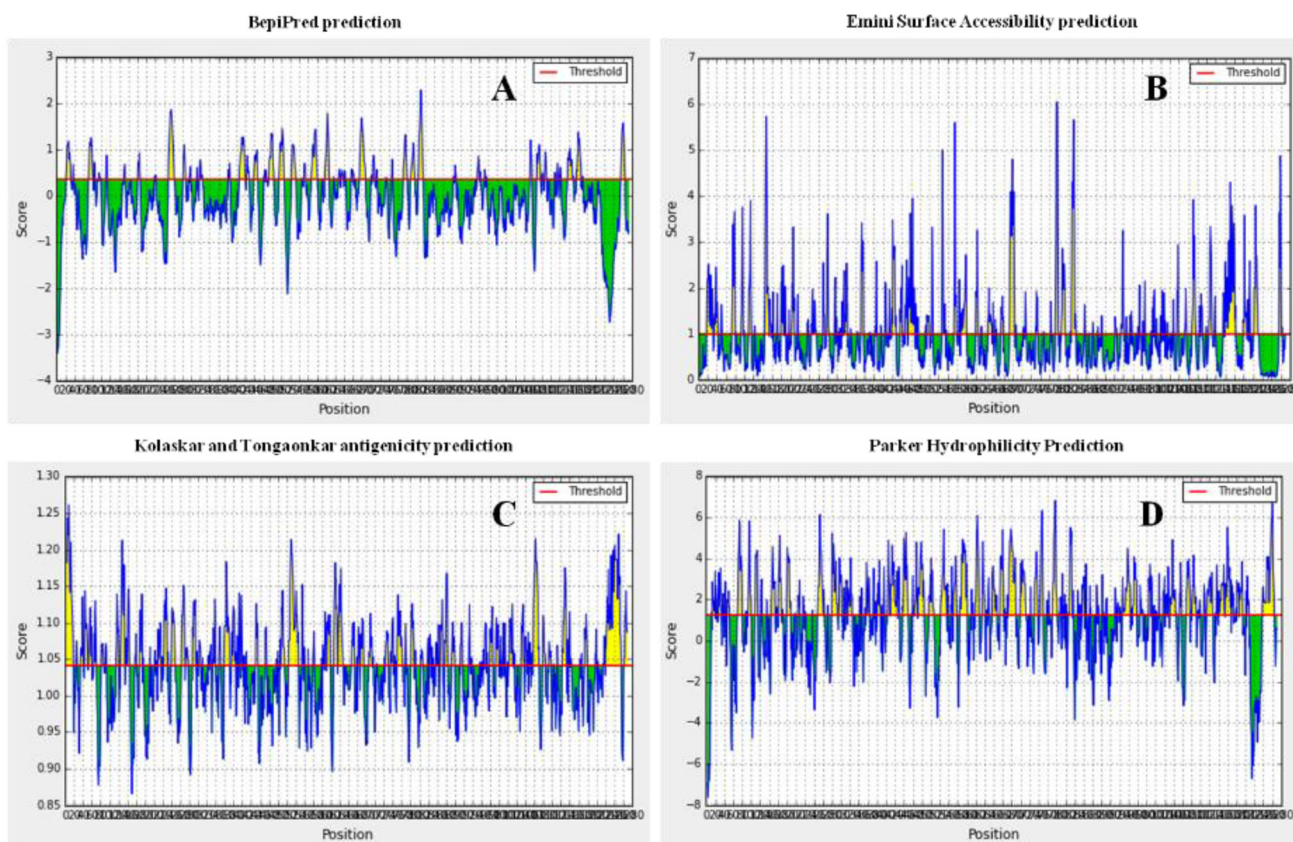


Fig. 7 Immunogenic B cell epitopes prediction for SARS-CoV2 spike protein based on physiochemical properties of residues on different IEDB scales (BepiPred, Emini accessibility prediction, Kolaskar and Tongaonkar antigenicity prediction method and Parker

hydrophilicity plot). Immunogenic sequences showing peaks above the threshold (red line) will be expected as a fraction of B cell epitope (Color figure online)

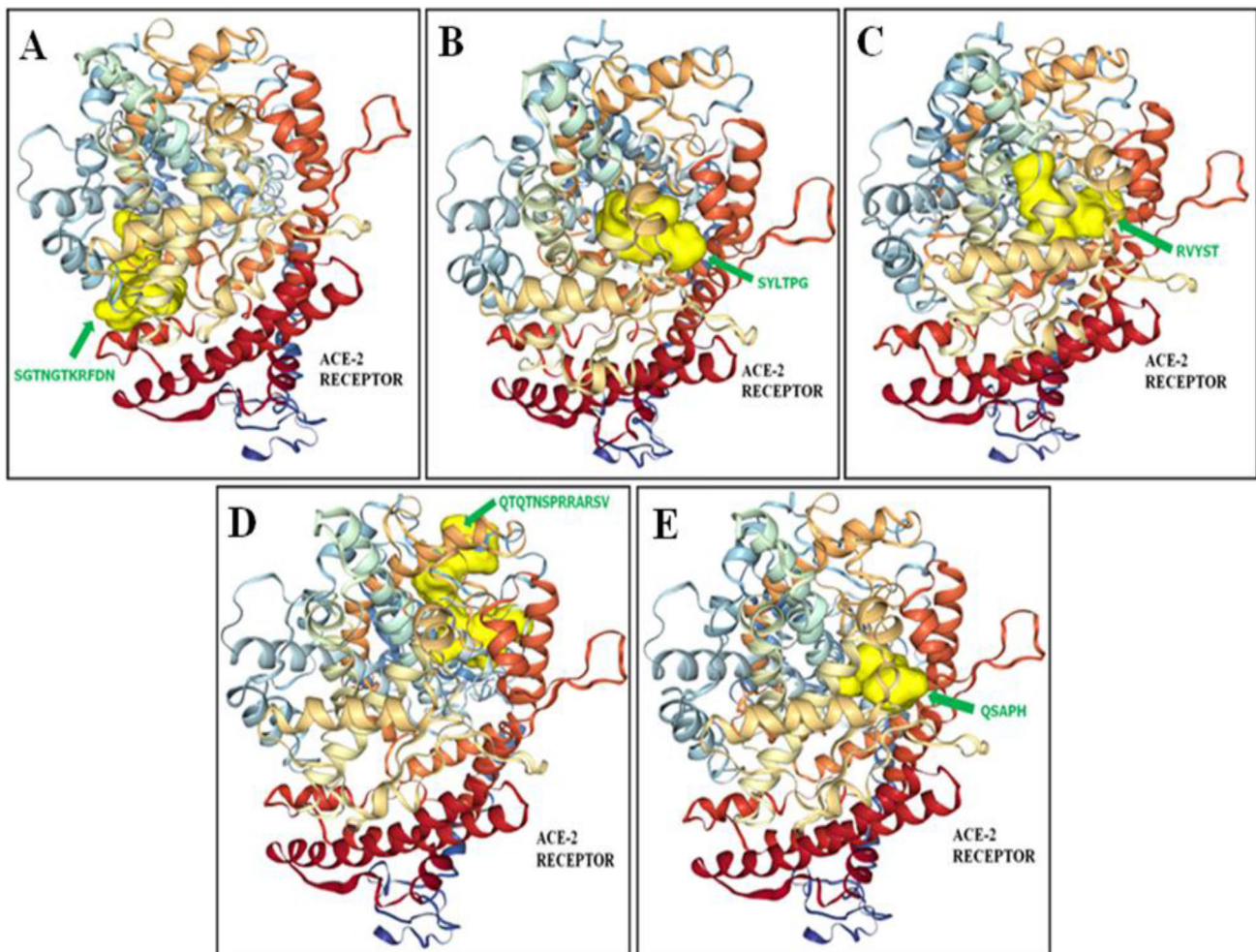


Fig. 8 Depiction of docked predicted immunogenic peptides on 3D model of ACE2 receptor using HPEPDOCK Server

Immunoinformatics approach for designing multi-epitopic vaccines based on surface immunogens has been used earlier for various diseases such as Ebola, Zika [1, 3, 27]. Previous studies reported the development of SARS-CoV vaccines using spike protein as a target [12, 18, 29]. Blocking of the receptors by an antiviral peptide specifically prevents the entry of SARS-CoV in Vero cells [14] has been reported earlier. Therefore, the same can be achieved for SARS-CoV2 using the predicted peptides; however, *in-vitro* and *in-vivo* studies need to be performed for further validation.

Conclusion

Keeping into consideration the threat caused by the virus to humankind and its massive spread throughout the world, it is essential to develop highly specific prophylactic and therapeutic measures viz. vaccines, antiviral peptides, and small biological molecules to curb this ongoing menace. As

the spike protein of SARS-CoV2 has a role in binding with cellular receptors, it gives a clue about the possible usage of the same as a target. This study reveals the prediction of five potent epitopes (*i.e.* ⁷¹SGTNGTKRFDN⁸¹, ²⁴⁷SYLTPG²⁵², ⁶³⁴RVYST⁶³⁸, ⁶⁷⁵QTQNSPRRARSV⁶⁸⁷, and ¹⁰⁵⁴QSAPH¹⁰⁵⁸) homologous to spike protein of the virus and their successful docking with the 3D model of ACE2 receptor. Experimental validations of these *in silico* predicted epitopes would be required for its consideration as an antigen-based diagnostic tool and as antiviral peptides. The use of virus antigen-based peptides along with their combination cocktails may serve as a boon in the wake of this current havoc.

Funding No funding was received from any organization.

Declarations

Conflict of interest There is no conflict of interest between the authors.

References

- Ahmad B, Ashfaq UA, Rahman MU, Masoud MS, Yousaf MZ. Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach. *Microb pathog.* 2019;132:243–53. <https://doi.org/10.1016/j.micpath.2019.05.010>.
- Bassetti M, Vena A, Roberto GD. The novel chinese coronavirus (2019-nCoV) infections: challenges for fighting the storm. *Eur J Clin Invest.* 2020. <https://doi.org/10.1111/eci.13209>.
- Bazhan SI, Antonets DV, Karpenko LI, Oreshkova SF, Kaplina ON, Starostina EV, Dudko SG, Fedotova SA, Ilyichev AA. In silico designed ebola virus T-cell multi-epitope DNA vaccine constructions are immunogenic in mice. *Vaccines.* 2019;7:34. <https://doi.org/10.3390/vaccines7020034>.
- Beniac DR, Andonov A, Grudski E, Booth TF. Architecture of the SARS coronavirus refusion spike. *Nat Struct Mol Biol.* 2006;13:751–2. <https://doi.org/10.1038/nsmb1123>.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014;8:W252. <https://doi.org/10.1093/nar/gku340>.
- Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol.* 2003;77:8801–11. <https://doi.org/10.1128/jvi.77.16.8801-8811.2003>.
- Bosch JFW, Kok KH, Zhu Z, Chu H, Kai-Wang TK, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 2020;9:221–36. <https://doi.org/10.1080/22221751.2020.1719902>.
- Bowie J, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991;253:164–70. <https://doi.org/10.1126/science.1853201>.
- Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J med virol.* 2020;92:418–23. <https://doi.org/10.1002/jmv.25681>.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993;2:1511–9. <https://doi.org/10.1002/pro.5560020916>.
- Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, Sette A. IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Res.* 2019;47:W502–6. <https://doi.org/10.1093/nar/gkz452>.
- Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV—A target for vaccine and therapeutic development. *Nat Rev Microbiol.* 2009;7:226–36. <https://doi.org/10.1038/nrmicro2090>.
- Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985;55:836–9. <https://doi.org/10.1128/JVI.55.3.836-839.1985>.
- Hu H, Li L, Kao RY, Kou B, Wang Z, Zhang L, Zhang H, Hao Z, Tsui WH, Ni A, Cui L. 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 Combin Chem.* 2005;7:648–56. <https://doi.org/10.1021/cc0500607>.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. *China Lancet.* 2020;S0140–6736(20):30183–5. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Huang J, Cao Y, Bu X, Wu C. Residue analysis of a CTL epitope of SARS-CoV spike protein by IFN-gamma production and bioinformatics prediction. *BMC Immunol.* 2012;13:50. <https://doi.org/10.1186/1471-2172-13-50>.
- Ji W, Wang W, Zhao X, Zai J, Li X. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. *J Med Virol.* 2020. <https://doi.org/10.1002/jmv.25682>.
- Jiang S, He Y, Liu S. SARS vaccine development. *Emerg Infect Dis.* 2005;11:1016–20. <https://doi.org/10.3201/1107.050219>.
- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 1990;276:172–4. [https://doi.org/10.1016/0014-5793\(90\)80535-q](https://doi.org/10.1016/0014-5793(90)80535-q).
- Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* 2006;2:2. <https://doi.org/10.1186/1745-7580-2-2>.
- Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol.* 2016;3:237–61. <https://doi.org/10.1146/annurev-virology-110615-042301>.
- Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by α geometry: ϕ , ψ and $C\beta$ deviation. *Proteins.* 2003;50:437–50. <https://doi.org/10.1002/prot.10286>.
- McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 2013;41:W597–600. <https://doi.org/10.1093/nar/gkt376>.
- Morse JS, Lalonde T, Xu S, Liu W. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *ChemBioChem.* 2020;21:730–8. <https://doi.org/10.1002/cbic.202000047>.
- 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;19:5425–32. <https://doi.org/10.1021/bi00367a013>.
- Perlman S. Another decade, another coronavirus. *N Engl J Med.* 2020;20:760–2. <https://doi.org/10.1056/NEJMe2001126>.
- Prasasty VD, Grazzolie K, Rosmalena R, Yazid F, Ivan FX, Sinaga E. Peptide-based subunit vaccine design of T- and B-cells multi-epitopes against zika virus using immunoinformatics approaches. *Microorganisms.* 2019;7:226. <https://doi.org/10.3390/microorganisms7080226>.
- Tahir M, Ul Qamar, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J Transl Med.* 2019;17:362. <https://doi.org/10.1186/s12967-019-2116-8>.
- Taylor DR. Obstacles and advances in SARS vaccine development. *Vaccine.* 2006;24:863–71. <https://doi.org/10.1016/j.vaccine.2005.08.102>.
- Vashi Y, Jagrit V, Kumar S. Understanding the B and T cell epitopes of spike protein of severe acute respiratory syndrome coronavirus-2: a computational way to predict the immunogens. *Infect Genet Evol.* 2020;84:104382. <https://doi.org/10.1016/j.meegid.2020.104382>.
- Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS. *J Virol.* 2020;17:e00127–e220. <https://doi.org/10.1128/JVI.00127-20>.

32. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* 2010;38:W469–73. <https://doi.org/10.1093/nar/gkq406>.
33. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35:W407–10. <https://doi.org/10.1093/nar/gkm290>.
34. World Health Organization. Pneumonia of unknown cause—China. 2020a, <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/>.
35. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science.* 2020;367:1444–8. <https://doi.org/10.1126/science.abb2762>.
36. Yan Y, Zhang D, Huang SY. Efficient conformational ensemble generation of protein-bound peptides. *J Cheminform.* 2017;9:59. <https://doi.org/10.1186/s13321-017-0246-7>.
37. Zhou L, Liu HG. Early detection and disease assessment of patients with novel coronavirus pneumonia. *Chinese J Tuberc Respir Dis.* 2020;43:E003. <https://doi.org/10.3760/cma.j.issn.1001-0939.2020.0003>.
38. Zhou P, Jin B, Li H, Huang SY. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* 2018;46:W443–50. <https://doi.org/10.1093/nar/gky357>.
39. Zhou P, Li B, Yan Y, Jin B, Wang L, Huang SY. Hierarchical flexible peptide docking by conformer generation and ensemble docking of peptides. *J Chem Inf Model.* 2018;58:1292–302. <https://doi.org/10.1021/acs.jcim.8b00142>.
40. Zhou Y, Jiang S, Du L. Prospects for a MERS-CoV spike vaccine. *Expert Rev Vaccines.* 2018;17:677–86. <https://doi.org/10.1080/14760584.2018.1506702>.
41. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A novel coronavirus from patients with pneumonia in China 2019. *N Engl J Med.* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.