

Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach

M. Shaminur Rahman^{1,*}, M. Nazmul Hoque^{1,2,*}, M. Rafiul Islam¹, Salma Akter^{1,3}, A. S. M. Rubayet Ul Alam⁴, Mohammad Anwar Siddique¹, Otun Saha¹, Md. Mizanur Rahaman¹, Munawar Sultana¹, Keith A. Crandall⁵ and M. Anwar Hossain^{1,6}

¹ Department of Microbiology, University of Dhaka, Dhaka, Bangladesh

² Department of Gynecology, Obstetrics and Reproductive Health, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Gazipur, Bangladesh

³ Department of Microbiology, Jahangirnagar University, Savar, Bangladesh

⁴ Department of Microbiology, Jashore University of Science and Technology, Jashore, Bangladesh

⁵ Computational Biology Institute, Milken Institute School of Public Health, George Washington University, Washington, Washington D.C., United States of America

⁶ Vice-Chancellor, Jashore University of Science and Technology, Jashore, Bangladesh

* These authors contributed equally to this work.

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiologic agent of the ongoing pandemic of coronavirus disease 2019 (COVID-19), a public health emergency of international concerns declared by the World Health Organization (WHO). An immuno-informatics approach along with comparative genomics was applied to design a multi-epitope-based peptide vaccine against SARS-CoV-2 combining the antigenic epitopes of the S, M, and E proteins. The tertiary structure was predicted, refined and validated using advanced bioinformatics tools. The candidate vaccine showed an average of $\geq 90.0\%$ world population coverage for different ethnic groups. Molecular docking and dynamics simulation of the chimeric vaccine with the immune receptors (TLR3 and TLR4) predicted efficient binding. Immune simulation predicted significant primary immune response with increased IgM and secondary immune response with high levels of both IgG1 and IgG2. It also increased the proliferation of T-helper cells and cytotoxic T-cells along with the increased IFN- γ and IL-2 cytokines. The codon optimization and mRNA secondary structure prediction revealed that the chimera is suitable for high-level expression and cloning. Overall, the constructed recombinant chimeric vaccine candidate demonstrated significant potential and can be considered for clinical validation to fight against this global threat, COVID-19.

Submitted 16 April 2020
Accepted 29 June 2020
Published 27 July 2020

Corresponding authors
Md. Mizanur Rahaman,
razu002@du.ac.bd
M. Anwar Hossain,
hossaina@du.ac.bd

Academic editor
Vasco Azevedo

Additional Information and
Declarations can be found on
page 23

DOI 10.7717/peerj.9572

© Copyright
2020 Rahman et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Computational Biology, Virology, Immunology, Infectious Diseases

Keywords SARS-CoV-2, Muti-epitope, Chimeric Peptide Vaccine, B-cell Epitope, T-cell Epitope

INTRODUCTION

Emergence of the SARS-CoV-2, which was first reported in Hubei Province of Wuhan, China in December 2019, is responsible for the ongoing global pandemic of coronavirus disease 2019 (COVID-19) spreading across 216 countries, areas, or territories with 6,992,274 active infection cases and 403,128 deaths until June 9, 2020 ([WHO, 2020](#)). This SARS-CoV-2 is the third coronovirus (CoV) belonging to Genus *Betacoronavirus* that can infect human after the two previously reported coronavirus- severe acute respiratory syndrome (SARS-CoV) ([Almofti et al., 2018](#); [Wu et al., 2020](#)), and Middle East respiratory syndrome (MERS-CoV) ([Badawi et al., 2016](#); [Pallesen et al., 2017](#); [Ul Qamar et al., 2019](#)). The SARS-CoV-2 has a positive-sense, single-stranded, and ~30 kilobase long RNA genome showing 79.0% and 50.0% identity to the genomes of SARS-CoV and MERS-CoV, respectively ([Abdelmageed et al., 2020](#); [Tang et al., 2020](#); [Lu et al., 2020](#); [Hoque et al., 2020a](#)). Among multiple encoded proteins (structural, non-structural, and accessory), four major structural proteins are the spike (S) glycoprotein, small envelope protein (E), membrane protein (M), and nucleocapsid protein (N) ([Ahmed, Quadeer & McKay, 2020](#)).

The S glycoprotein, because of its higher antigenicity and surface exposure ([Almofti et al., 2018](#); [Zhou et al., 2019](#); [Shang et al., 2020](#)), plays the most crucial role for the attachment and entry of viral particles into the host cells through the host angiotensin-converting enzyme 2 (ACE2) receptor ([Gralinski & Menachery, 2020](#); [Shang et al., 2020](#)). It is noteworthy that E and M proteins also have important functions in the viral assembly, budding and replication of virus particles, as well as play role in augmenting the immune response against SARS-CoV ([Shi et al., 2006](#); [Schoeman & Fielding, 2019](#); [Shang et al., 2020](#)). Monoclonal antibodies (mAbs) primarily target the trimeric S glycoprotein of the virion consisting of three homologous chains (A, B, and C), and this protein is composed of two major domains, the receptor-binding domain (RBD) and the N-terminal domain (NTD) ([Pallesen et al., 2017](#); [Song et al., 2018](#); [Zhou et al., 2019](#); [Wrapp et al., 2020](#)). The NTD is located on the side of the spike trimer and has not been observed to undergo any dynamic conformational changes ([Shang et al., 2020](#)). Thus, this specific region might play role in viral attachment, inducing neutralizing antibody responses and stimulating a protective cellular immunity ([Almofti et al., 2018](#); [Ul Qamar et al., 2019](#); [Shang et al., 2020](#)).

Most of the recent vaccine candidates induce neutralizing antibodies against the different forms and/or variants of the spike protein of the SARS-CoV-2 ([Le et al., 2020](#)). However, the immune responses generated from using single protein have generally been inadequate to warrant their use in the development of an effective prophylactic tool ([Shi et al., 2015](#); [Shey et al., 2019](#)). On this note, multi-epitope vaccine candidates have already been designed against several viruses, including MERS-CoV and SARS-CoV, and their efficacies have been further reported ([Almofti et al., 2018](#); [Ul Qamar et al., 2019](#); [Yong et al., 2019](#)). Two related studies have reported the *in-silico* design of epitope based chimeric vaccine candidates targeting E, M, S and N proteins of SARS-CoV-2, albeit not peer-reviewed ([Yazdani et al., 2020](#); [Akhand et al., 2020](#)). Besides, [Kibria, Ullah & Miah \(2020\)](#) performed an immunoinformatic approach to design a 70 aa long multi-epitope vaccine focusing on the the virion outer surface proteins (E, M, and S) ([Chan et al., 2020](#)).

Scientists are racing over the clock to develop effective vaccine for controlling and preventing COVID-19 based on the genomics, functional structures, and host-pathogen interactions; nevertheless, the ultimate results of these efforts is yet uncertain. Currently, 10 candidate vaccines against SARS-CoV-2 are in the clinical trial, and 121 more under preclinical evaluation ([WHO, 2020](#)). Strikingly, researchers are trialing different technologies, albeit targeting spike protein mainly, some of which have not been used in a licensed vaccine before ([Le et al., 2020](#)). Their vaccine appears to be effective and safe based on a limited data and application of the vaccine within a relatively tiny group of individuals. However, there are several uncertainties, for example, whether it will relate to antibody responses in the general population, be safe within a specific sub-population (children, pregnant women, and elder people), as well as lack of a standardized virus neutralization assay and accurate vaccine titer are complicating data interpretation. Moreover, only half of the medium dose receiver developed neutralizing antibody and the T-cell response is not particularly impressive ([Sheridan, 2020](#)).

Thus, mimicking of a more natural state of the virus where surface exposed proteins, or the immunodominant epitopes of those proteins influencing the immune response might be a solution, if those candidates ultimately cannot meet the final goal. Furthermore, excluding the nucleocapsid (N) protein, which is embedded within the structure and attached to the viral genome, will amplify the chance of developing a more pseudo-virus state for the expressed chimera that may produce specific antibody as well as T-cell responses. Furthermore, peptide-based chimeric vaccines are biologically safe as they do not need in vitro virus culture, and their selectivity might ensure accurate activation of specific immune responses ([Dudek et al., 2010](#); [Wang et al., 2019](#)).

Considering the facts, we have proposed the development of a multi-epitope vaccine candidate, which differs from all the previous studies in the aspect of containing whole RBD and NTD regions of the spike protein alongwith specific epitopes of M and E proteins, giving an excellent chimeric conformation and might lead to the generation of a more potent protective immune responses since smaller epitopes have less ability to give better immune protection. Hence, we can assume that chimeric vaccine targeting multiple epitopes on the RBD and NTD segments of the S protein, M and E proteins would be a potentially effective vaccine candidate in combatting COVID-19 pandemic, and therefore, could be used against the could be used against the highly pathogenic SARS-CoV-2.

RESULTS

Comparative structural analysis of SARS-CoV, MERS-CoV and SARS-CoV-2

Multiple sequence alignment revealed that the S protein of SARS-CoV-2 shares 77.38% and 31.93% sequence identity with the S proteins of the SARS-CoV and MERS-CoV, respectively ([Fig. S1](#)). The structural (validated using the Ramachandran plot as in [Fig. S2](#)) alignment of the coronavirus S proteins reflects high degree of structural heterogeneity in the receptor-binding domain (RBD) and N-terminal domain (NTD) of the chain A and chain C compared to that of chain B ([Fig. S3](#)). Divergence of individual structural domains,

NTD and RBD of 2019-nCoV spike protein from both of the SARS-CoV and MERS-CoV warrants the domains for epitope-based chimeric vaccine development, particularly against SARS-CoV-2.

Screening for B-cell epitopes

Linear epitopes prediction (ElliPro) based on solvent-accessibility and flexibility revealed 15, 18, and 19 epitopes within the chain A, B and C of S protein, respectively wherein score >0.8 was the threshold for the highly antigenic epitopes ([Table 1](#)). The amino acid (aa) residues in 56-194 and 395-514 position of the detected epitopes belonged to RBD and NTD regions of the S protein, respectively. However, the epitopes with aa position of 1067-1146 were not selected as the potential epitope candidate because of their presence in viral transmembrane domains ([Fig. S4](#)). The tertiary structures of the RBD and NTD illustrate their surface-exposed position on the S protein ([Fig. 1](#)). Using IEDB analysis resource and Bepipred linear epitope prediction 2.0 tools, we predicted eight and six B-cell epitopes in RBD and NTD regions out of total 22 epitopes in S protein, while the E and M proteins had 2 and 6 epitopes, respectively ([Fig. 2](#), [Table S1](#)). However, only 5 epitopes were exposed on the surface of the virion, and had a high antigenicity score (>0.4), indicating their potentials in initiating immune responses ([Table 2](#)).

Among the five annotated epitopes having antigenicity score of ≥ 0.5 (VaxiJen 2.0 tool), RBD and NTD regions each possessed two highly antigenic epitopes while the envelope (E) protein contained only one highly antigenic epitope and membrane (M) protein has none ([Table 2](#)). Furthermore, the Kolaskar and Tongaonkar antigenicity profiling found five highly antigenic epitopes in RBD region with an average (antigenicity) score of 1.042 (minimum = 0.907, maximum = 1.214), and seven highly antigenic epitopes in NTD with an average (antigenicity) score of 1.023 (minimum = 0.866, maximum = 1.213) ([Fig. S5](#), [Table S2](#)). The average Kolaskar scores for envelope protein B-cell epitope (EBE) and membrane protein B-cell epitope (MBE) were 0.980 and 1.032, respectively ([Table S2](#)). However, through ABCPred analysis, we further verified 18 and 11 B-cell epitopes in RBD and NTD regions with average antigenicity score of 0.775 and 0.773 in the associated domains, respectively ([Table S3](#)).

Selection of T-cell and IFN- γ inducing epitopes

The IEDB MHC-I prediction tool retrieved 77 T-cell epitopes in RBD that interacted with 21 possible MHC-I alleles whereas the NTD domain possessed 35 T-cell epitopes with 17 possible MHC-I alleles ([Data S1](#)). Similarly, the IEDB MHC-II prediction tool generated 13-mer 124 peptides from the RBD, and 10-mer 73 peptides in the NTD segments of the S protein that showed interaction with many different and/or common MHC-II alleles with an IC₅₀ value ranging from 1.4 to 49.9 nM ([Data S1](#)). Furthermore, the analysis tool of the IEDB generated an overall scores for proteasomal processing, TAP transport, and MHC-binding efficiency indicating the intrinsic potential of the epitopes to be recognized by immunoreactive T-cells ([Data S1](#)).

The findings of IFNepitope program suggests that, both the target RBD and NTD regions of S protein, and membrane protein B-cell linear epitope (MBE) had great probability to

Table 1 Linear epitopes present on spike (S) glycoprotein surface predicted through ElliPro in IEDB-analysis resource based upon solvent-accessibility and flexibility are shown with their antigenicity scores. The highlighted green coloured regions were the potential antigenic domains while the yellow coloured region represents the trans-membrane domain of the S protein.

No.	Chain	Start	End	Peptide	Residues	Score
1		395	514	VYADSFVIRGDEVRQIAPGQTGKIADNYKLP DDFTGCIAWNSNNLDSKVGGNNYLYRLFRK SNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG FQPTNGVGYQPYRVVVLS	120	0.837
2		58	194	FFSNVTWFHAIHVSGTNGTKRFNDNPVLPFNDGVYFAS TEKSNIIRGWIFGTTLDSKTQSLIVNNATNVVIKCEF QFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYV SQPFLMDLEGKQGNFKNLREFVF	137	0.835
3		1067	1146	YVPAQEKNFTTAPAICHHDGKAHFREGVFSNG THWFVTQRNFYEPQIITTDNTF VSGNCDVVIGIVNNNTVYDPLQPELD	80	0.83
4		201	270	FIYSKHTPINLVRDLPQGFSALEPLVDLPIG INITRFQTLLAHLRSYLTGPDSSSGWTAGAAAYVGYL	70	0.76
5		331	381	NITNLCPFGEVFVNATRFASVYAWNKRISN CVADYSVLYNSASFSTFKCYG	51	0.706
6		700	720	GAENSVAYSNNNSIAPTNFTI	21	0.668
7		27	35	AYTNSFTRG	9	0.66
8	A	909	936	IGVTQNVLYENQKLIANQFNNSAIGKIQD	28	0.633
9		789	813	YKTPPIKDFGGFNFSQLPDPSKPS	25	0.6
10		623	642	AIHADQLTPTWRVYSTGSNV	20	0.598
11		891	907	GAALQIPFAMQMAYRFN	17	0.591
12		579	583	PQTLE	5	0.551
13		687	692	VASQSI	6	0.55
14		653	659	AEHVNNS	7	0.539
15		679	684	NSPRRA	6	0.521
16		1067	1146	YVPAQEKNFTTAPAICHHDGKAHFREGVFSNGTH WFVTQRNFYEPQIITTDNTFVS GNCDVVIGIVNNNTVYDPLQPELD	80	0.826
17		89	194	GVYFASTEKSNIIRGWIFGTTLDSKTQSLIVNNATN VVIKCEFQFCNDPFLGVYYHKNNKSWMESEFRVY SANNCTFEYVSQPFLMDLEGKQGNFKNLREFVF	106	0.816
18		58	87	FFSNVTWFHAIHVSGTNGTKRFNDNPVLPFN	30	0.81
19		203	270	IYSKHTPINLVRDLPQGFSALEPLVDLPIGINIT RFQTLLAHLRSYLTGPDSSSGWTAGAAAYVGYL	68	0.748
20		465	509	ERDISTEIYQAGSTPCNGVEGF NCYFPLQSYGFQPTNGVGYQPYR	45	0.727
21		436	458	WNSNNLDSKVGGNNYLYRLFRK	23	0.672
22		700	720	GAENSVAYSNNNSIAPTNFTI	21	0.671
23		27	35	AYTNSFTRG	9	0.666
24		909	9036	IGVTQNVLYENQKLIANQFNNSAIGKIQD	28	0.641
25		624	643	IHADQLTPTWRVYSTGSNVF	20	0.617
26	B	328	365	RFPNITNLCPFGEVFVNATRFASVYAWNKRISNCVADY	38	0.608

(continued on next page)

Table 1 (continued)

No.	Chain	Start	End	Peptide	Residues	Score
27		891	907	GAALQIPFAMQMAYRFN	17	0.602
28		577	583	RDQPQTL	7	0.598
29		790	817	KTPPIKDFGGFNFSQILPDPSKPSKRSF	28	0.595
30		673	693	SYQTQTNSPRRARSVASQSII	21	0.567
31		526	537	GPKKSTNLVKNK	12	0.553
32		653	661	AEHVNNSYE	9	0.548
33		554	563	ESNKKFLPFQ	10	0.52
34		56	194	LPFFSNVTWPHAIHVSGTNGTKRFDNPVLPFNDGVYFA STEKSNIIRGWIFGTTLDSTQSLIVNNATNVVIKV CEFAQCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFE YVSQPFLMDLEGKQGNFKNLREFVFT	139	0.84
35		1067	1146	YVPAQEKNFTTAPAIChDGKAHFPREGVFV SNGTHWFVTQRNFYEPQIITTDNTFVSGN CDVVIGVNNTVYDPLQPELD	80	0.822
36		201	270	FKIYSKHTPINLVRDLPQGFS LEPLVDLPIGINITRFQTLALHR SYLTPGDSSSGWTAGAAAYVGYL	70	0.77
37		27	35	AYTNSFTRG	9	0.676
38		465	509	ERDISTEIYQAGSTPCNGVEGFNCY FPLQSYGFQPTNGVGYQPYR	45	0.675
39		700	720	GAENSVAYSNNNSIAPTNFTI	21	0.658
40		909	936	IGVTQNVLYENQKLIANQFNSAIGKIQD	28	0.633
41		437	458	NSNNLDSKVGGNNYLYRLFRK	22	0.629
42		673	684	SYQTQTNSPRRA	12	0.619
43	C	790	817	KTPPIKDFGGFNFSQILPDPSKPSKRSF	28	0.602
44		891	907	GAALQIPFAMQMAYRFN	17	0.597
45		578	583	DPQTL	6	0.589
46		620	631	VPVAIHADQLTP	12	0.579
47		329	362	FPNITNLCPFGEVFNATRFASVYAWNRKRISNCV	34	0.571
48		687	692	VASQSI	6	0.566
49		835	845	KQYGDCLGDI	11	0.567
50		653	659	AEHVNN	7	0.559
51		527	536	PKKSTNLVKN	10	0.546
52		635	642	VYSTGSNV	8	0.51

release of IFN- γ with a positive score. A total of 56 potential positive IFN- γ inducing epitopes (15-mer) were predicted for the RBD domain with an average epitope prediction score of 0.255 and the maximum SVM score of 0.625. On the other hand, a total of 33 potential positive epitopes were predicted for the NTD domain with an average epitope prediction score of 0.312 and the maximum SVM score of 0.811. Moreover, the M protein also possessed several IFN- γ inducing epitopes having an average epitope prediction score of 0.980 (Table S4).

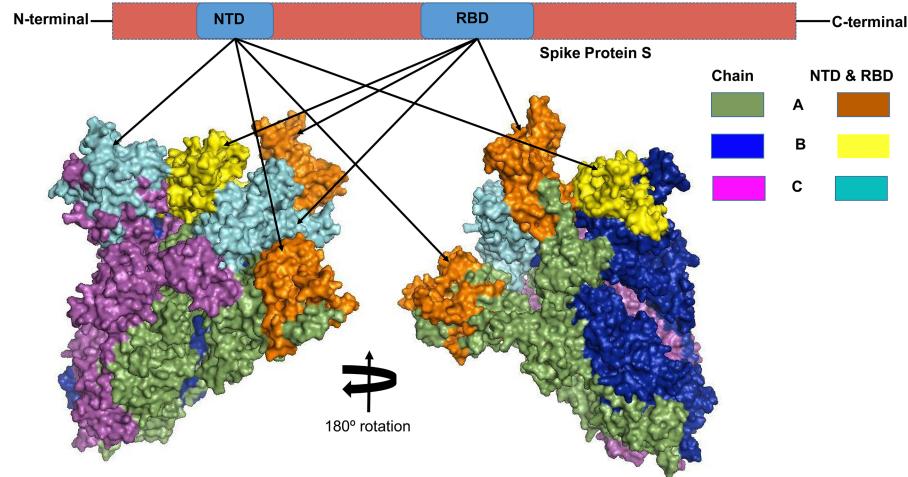


Figure 1 The three-dimensional (3D) structure of the N-terminal domains (NTDs) and receptor binding domains (RBDs) of the spike (S) proteins of SARS-CoV-2 (surface view). The orange, cyan, and yellow colored regions represent the potential antigenic domains predicted by the IEDB analysis resource ElPro analysis.

[Full-size](#) DOI: [10.7717/peerj.9572/fig-1](https://doi.org/10.7717/peerj.9572/fig-1)

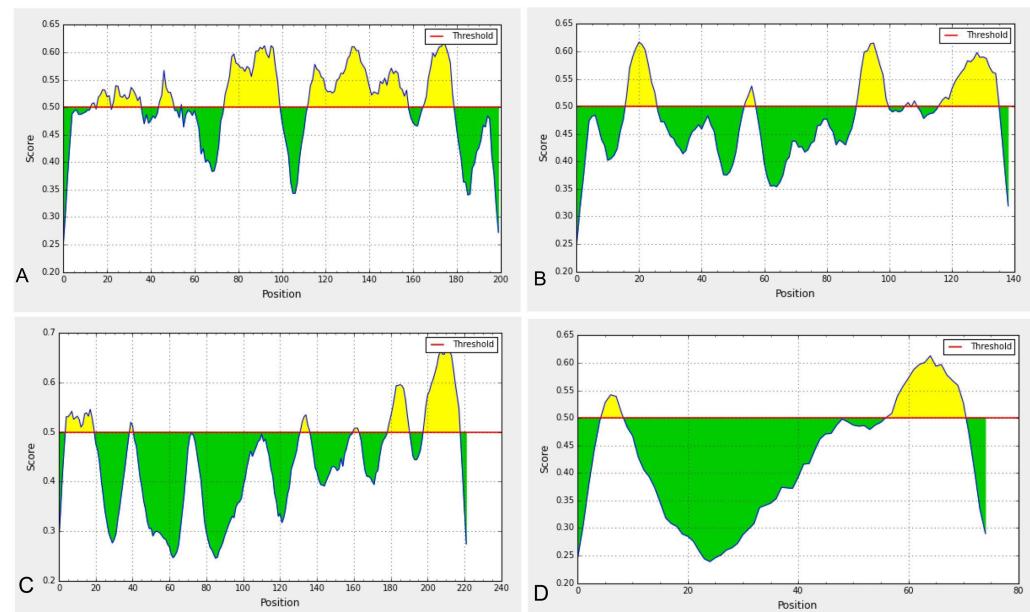


Figure 2 Predicted B-cell epitopes using BepiPred-2.0 epitope predictor in IEDB-analysis resource web-based repository. Yellow areas above threshold (red line) are proposed to be a part of B cell epitopes in (a) RBD and (b) NTD regions of S protein, (c) envelop (E) and (d) membrane (M) proteins of SARS-CoV-2.

[Full-size](#) DOI: [10.7717/peerj.9572/fig-2](https://doi.org/10.7717/peerj.9572/fig-2)

Table 2 B-cell epitopes predicted using Bepipred linear epitope prediction 2.0 in IEDB analysis resource web-server along with their start and end positions, average score, and VaxiJen 2.0 determined antigenicity scores.

Domain/ proteins	Position	Sequences	Average Score	Antigenicity
RBD	341-342	VF	0.502	—
	344-349	ATRFAS	0.520	-0.151
	351-363	YAWNRKRISNCVA	0.522	0.394
	372-378	ASFSTFK	0.527	0.087
	382	V	0.464	—
	402-427	IRGDEVVRQIAPGQTGKIADYNKYKLPD	0.575	0.932
	440-485	NLDSKVGGNNYNYLYRLFRKSN LKPFERDISTEIYQAGSTPCNGVEG	0.554	0.210
NTD	493-516	QSYGFQPTNGVGYQ	0.535	0.670
	72-81	GTNGTKRFDN	0.573	0.667
	110-113	LDSK	0.511	—
	146-155	HKNNKSWMES	0.573	0.174
MBE	161-162	SS	0.503	—
	164	N	0.499	—
EBE	172-191	SQPFLMDLEGKQGNFKNLRE	0.553	0.749
	199-218	YRIGNYKLNTDHSSSDNIA	0.614	0.222
EBE	57-71	YVYSRVKNLNSSRVP	0.565	0.449

Design-construction, antigenicity and physicochemical properties of the chimeric vaccine candidate

The selected epitope-sequences for designing of chimeric construct were PADRE (13 aa), MBE (20 aa), NTD (139 aa), RBD (200 aa), EBE (15 aa), and Invasin (16 aa), and the construct was named as CoV-RMEN (417 aa) as shown in Fig. 3A. These segments were connected with a repeat of hydrophobic (glycine; G) and acidic aa (glutamic acid; E) linkers for making the final vaccine construct more flexible with balanced ratio of acidic and basic amino acids. The molecular weight (MW) of the CoV-RMEN was 46.8 kDa with a predicted isoelectric point (pI) of 8.71. The projected half-life was 4.4 h in mammalian reticulocytes in vitro, and >20 h in yeast and >10 hours in *E. coli* in vivo. The protein was predicted to be less soluble upon expression with a solubility score of 0.330. An instability index (II) value 29.74 predicted the protein as stable (II of >40 indicates instability). The estimated aliphatic index was 66.59, indicating thermostability of the final chimera. The predicted grand average of hydropathicity (GRAVY) was -0.300. The antigenicity score of 0.450 was predicted by the VaxiJen 2.0 server with a virus model at a threshold of 0.4, and further verified by ANTIGENpro showing score of 0.875 (maximum expected score ranking is 1.0) indicating the high antigenic nature of the designed vaccine, CoV-RMEN. Moreover, the vaccine was also predicted to be non-allergenic on both the AllerTOP v.2 and AllergenFP servers.

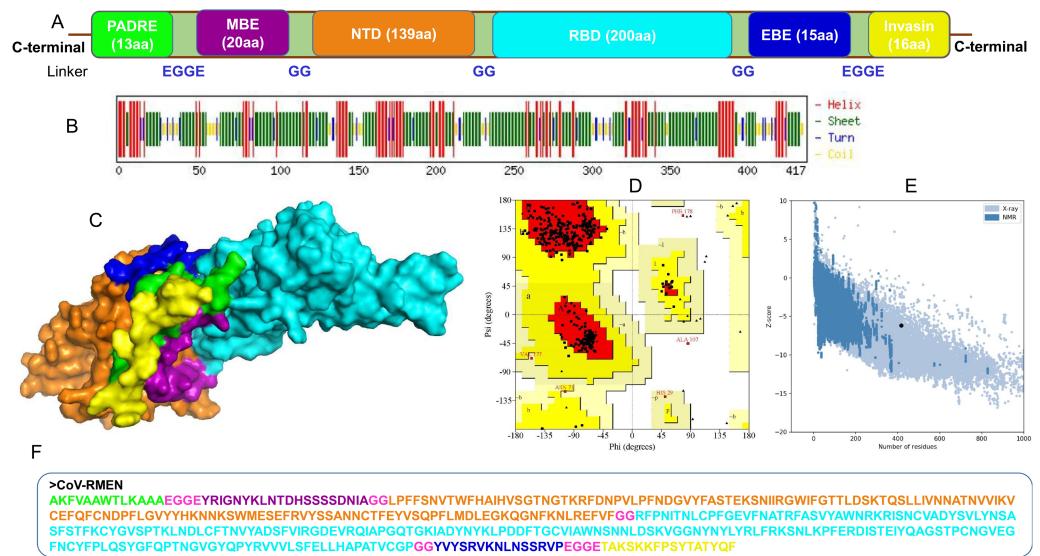


Figure 3 Design, construction and structural validation of multi-epitope vaccine candidate (CoV-RMEN) for SARS-CoV-2. (A) Structural domains and epitopes rearrangement of CoV-RMEN, (B) secondary structure of CoV-RMEN as analyzed through CFSSP: Chou and Fasman secondary structure prediction server , (C) final tertiary structure of CoV-RMEN (surface view) obtained from homology modelling on Phyre2 in which domains and epitopes are represented in different colors (PADRE-smudge; membrane B-cell epitope, MBE-magenta; N-terminal domain, NTD-orange; receptor-binding domain, RBD-cyan; envelop B-cell epitope, EBE-blue; invasin-yellow), (D) validation of the refined model with Ramachandran plot analysis showing 94.7%, 4.8% and 0.5% of protein residues in favored, allowed, and disallowed (outlier) regions respectively, (e) ProSA-web, giving a Z-score of -6.17, and (f) the finally predicted primary structure of the CoV-RMEN.

Full-size DOI: 10.7717/peerj.9572/fig-3

Structural characterization of the CoV-RMEN

The CoV-RMEN peptide was predicted to contain 43.2% alpha helix, 67.4% beta sheet, and 12% turns (Fig. 3B, Fig. S6) using CFSSP:Chou and Fasman secondary structure prediction server. Additionally, regarding the solvent accessibility of aa residues, 34%, 30% and 34% were predicted to be exposed, medium exposed and buried respectively (Fig. S7). The RaptorX Property server predicted only two aa residues in the disordered domains. The Phyre2 server predicted the tertiary structure model of the designed chimeric protein in 5 templates (c5x5bB, c2mm4A, c6vsbB, c5x29B and c6vybB) based upon heuristics to maximize confidence, alignment coverage and percent identity. The final 3D structure of the CoV-RMEN peptide modelled at 82% with more than 90% confidence (Fig. 3C). Moreover, 65 residues were modelled by *ab initio*. The selected structural model has parameters of RMSD (0.414), GDT-HA (0.9538), and MolProbity (2.035). The Ramachandran plot analysis of the finally modelled protein exhibited 94.7% of the aa residues in favored regions (Fig. 3D), consistent with the 94.0% score predicted by the GalaxyRefine analysis. Additionally, 4.8% of the residues located in allowed regions, and only 0.5% in disallowed regions (Fig. 3D). The chosen model after refinement had an overall quality factor of 74.45% with ERRAT (Fig. S8) and a ProSA-web based Z-score of -6.17 (Fig. 3E).

Molecular docking and dynamics simulation analysis

Among the selected epitopes from the RBD and NTD segments, top five based on IC₅₀ score ([Data S1](#)) revealed highly favorable molecular interaction for stable binding with their respective HLA alleles. Docking complexes thus formed have significantly negative binding affinity (ΔG always remained $\leq -8.2 \text{ kcal mol}^{-1}$, average = $-9.94 \text{ kcal mol}^{-1}$), and most of the amino acid (aa) residues of the epitopes were involved in molecular interactions with their respective HLA alleles ([Fig. 4](#), [Data S1](#)). The immune responses of TLR2, TLR3 and TLR4 against vaccine construct (CoV-RMEN) were estimated by analyzing the overall conformational stability of vaccine protein-TLRs docked complexes. The active interface aa residues of refined complexes of CoV-RMEN and TLRs were predicted ([Fig. 5](#), [Table 3](#)).

The relative binding free energies (ΔG) of the protein-TLRs complexes were significantly negative ([Table 3](#)) which suggest that the interaction of the chimeric protein might favor stimulation of the TLR receptors. Consistently, the number of contacts made at the interface (IC) per property (ICs charged-charged: 5, ICs charged-polar: 2, ICs charged-apolar: 17, polar-polar: 1, ICs polar-apolar: 7 and apolar-apolar: 16) for the vaccine protein-TLR2 complex. Interface contacts (IC) per property (ICs charged-charged: 16, ICs charged-polar: 22, ICs charged-apolar: 26, polar-polar: 6, ICs polar-apolar: 25 and apolar-apolar: 29) were for the vaccine protein-TLR3 complex. Also, vaccine protein-TLR4 complex showed similar (ICs) per property (ICs charged-charged: 5, ICs charged-polar: 11, ICs charged-apolar: 30, polar-polar: 4, ICs polar-apolar: 31 and apolar-apolar: 39). Furthermore, the molecular dynamics (MD) simulation analysis of the docked CoV-RMEN-TLR3 and CoV-RMEN-TLR4 complexes showed soundly stable RMSD values between ~ 4.35 and $\sim 5.4 \text{ nm}$ for a specified time frame of 100 ps at the reasonably consistent temperature ($\sim 300 \text{ K}$) and pressure (1bar), whereas CoV-RMEN-TLR2 complex showed RMSD value between 5.5 and 6.2 with same cut-off parameters. These data validated that the docked complexes (CoV-RMEN-TLR3 and CoV-RMEN -TLR4) are more stable than CoV-RMEN-TLR2 ([Fig. 5](#)).

Immune simulation

The cumulative results of immune responses after three times antigen exposure with four-week interval each time revealed that the primary immune response against the antigenic fragments was elevated indicated by gradual increase of IgM level after each antigen exposure ([Fig. 6A](#)). Besides, the secondary immune response, crucial for immune stability, have been shown as increased with adequate generation of both IgG1 and IgG2. Also, the elevated level of all circulating immunoglobulins indicates the accuracy of relevant clonal proliferation of B-cell and T-cell population. The level of cytokines after antigen exposure increased concomitantly reflected by escalation of IFN- γ and IL-2, which are most significant cytokines for anti-viral immune response and clonal selection ([Fig. 6B](#)). The abundance of different types of B-cells and T-cells, like antigen processing B-cells, resting memory B- and T-cells, B-cells with IgM and IgG remains significantly higher indicating development of immune memory and consequently increased clearance of antigen after exposure ([Figs. 6C and 6D](#)). Additionally, T-helper cells and cytotoxic T-cells were found with a drastic up-regulation of Th1 concentration enhancing the B-cell proliferation and

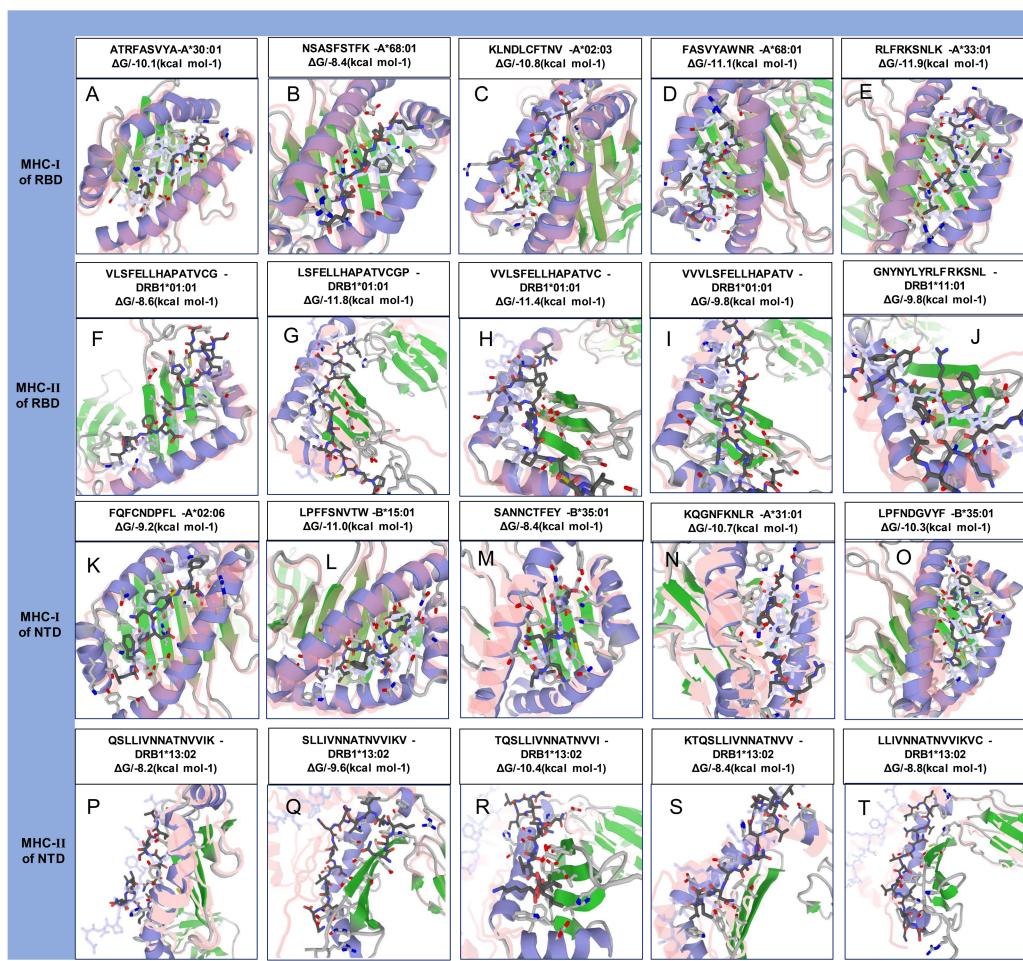


Figure 4 Molecular docking of top five MHC-I and MHC-II epitopes of RBD and NTD domains with respect to HLA allele binders. (A–E) and (K–O) represent the top five MHC-I epitopes of RBD and NTD domains, respectively. (F–J) and (P–T) represent the top five MHC-II epitopes binds of the same domains binds to their respective HLA alleles. The protein-peptide docking was performed in GalaxyWEB-GalaxyPepDock-server followed by the refinement using GalaxyRefineComplex and free energy (ΔG) of each complex was determined in PRODIGY server. Ribbon structures represent HLA alleles and stick structures represent the respective epitopes. Light color represents the templates to which the alleles and epitopes structures were built. Further information on molecular docking analysis is also available in [Data S1](#).

Full-size DOI: [10.7717/peerj.9572/fig-4](https://doi.org/10.7717/peerj.9572/fig-4)

immune memory development (Figs. 6E and 6F). The high level of immunoglobulin IgG1 + IgG2, active B-cell and T-helper cell population reflected the development of strong immune response reinforcing the indelible and peerless antigenicity of the CoV-RMEN vaccine candidate.

Population coverage analysis

The selected CTL and HTL epitopes covered 94.9% and 73.11% of the world population, respectively. Importantly, CTL and HTL epitopes showed 98.63% population coverage worldwide when considered in combination. The highest population coverage was found

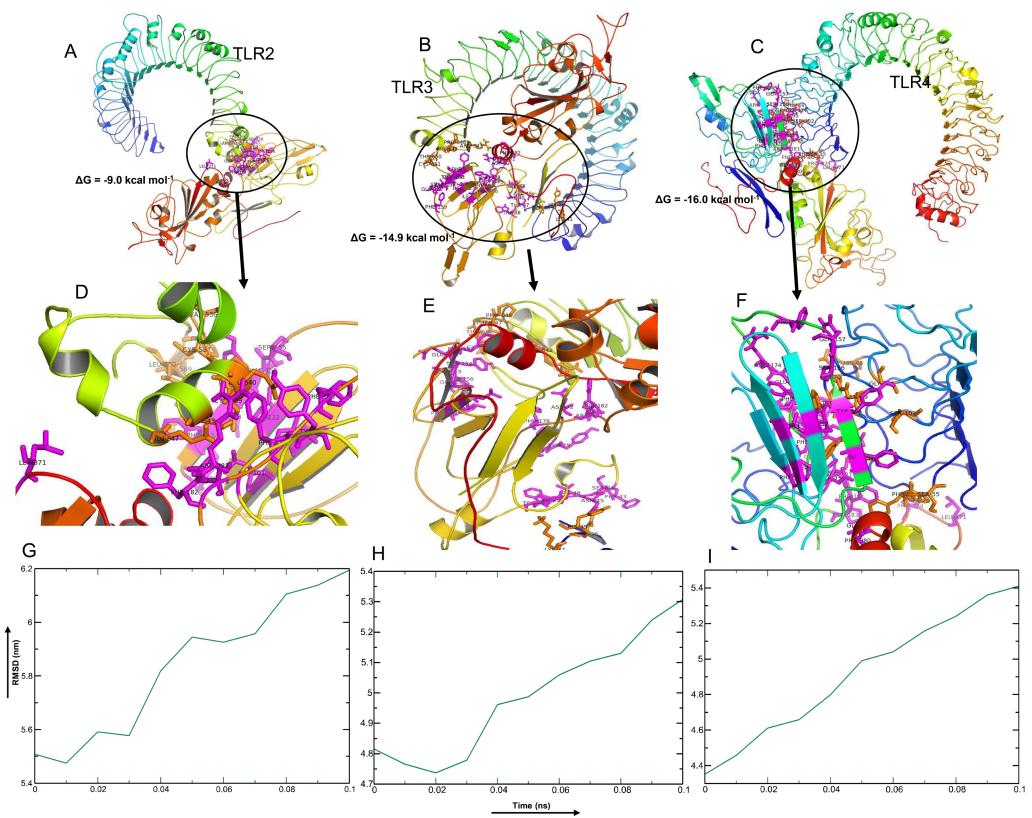


Figure 5 Molecular docking and dynamics of CoV-RMEN vaccine with immune receptors (TLR2, TLR3 and TLR4). Docked complexes for (A) CoV-RMEN and TLR2, (B) CoV-RMEN and TLR3, and (C) CoV-RMEN and TLR4. Magnified interfaces of the respective complexes are figured to (D), (E) and (F) respectively. Active residues of CoV-RMEN colored magenta, and of TLRs colored orange with stick view. ΔG represents the binding affinity of the complexes. Molecular dynamics simulation study of (G) CoV-RMEN and TLR2, (H) CoV-RMEN and TLR3, and (I) CoV-RMEN and TLR4 complexes across the time window of 100 ps. The reasonably invariable RMSD value indicates a stable complex formation.

Full-size DOI: 10.7717/peerj.9572/fig-5

Table 3 Active interface amino acid residues and binding scores among Toll Like Receptors (TLRs) and the constructed vaccine CoV-RMEN.

	Active residues of TLRs	Active residues of CoV-RMEN	HADDOCK score	ΔG (kcal mol^{-1})
TLR-2	V536, C537, S538, C539, E540, S543, E547, P567, R569, L570	D72, Y75, L101, I103, I112, C150, F152, E153, Y154, V155, S156, F176, F178, R181, F182, L371	-30.4	-9.0
TLR-3	D36, H39, K41, R643, F644, P646, F647, T650, C651, E652, S653, I654, W656, F657, V658, N659, W660, I661, N662, E663	F43, S44, N45, V46, T47, W48, D72, Y75, F76, L101, I103, I112, F152, E153, Y154, V155, S156, Q157, F159, F178, R181, F182, L371	-47.2	-14.9
TLR-4	P53, F54, S55, H68, G70, Y72, S73, F75, S76, Q99, S102, G124	G39, L40, D72, Y75, F76, F90, L101, I103, I112, F152, Y154, S156, Q157, F159, R174, E175, F176, F178, R181, F182, P183, L371, P374, P380, G381	-52.1	-16.0

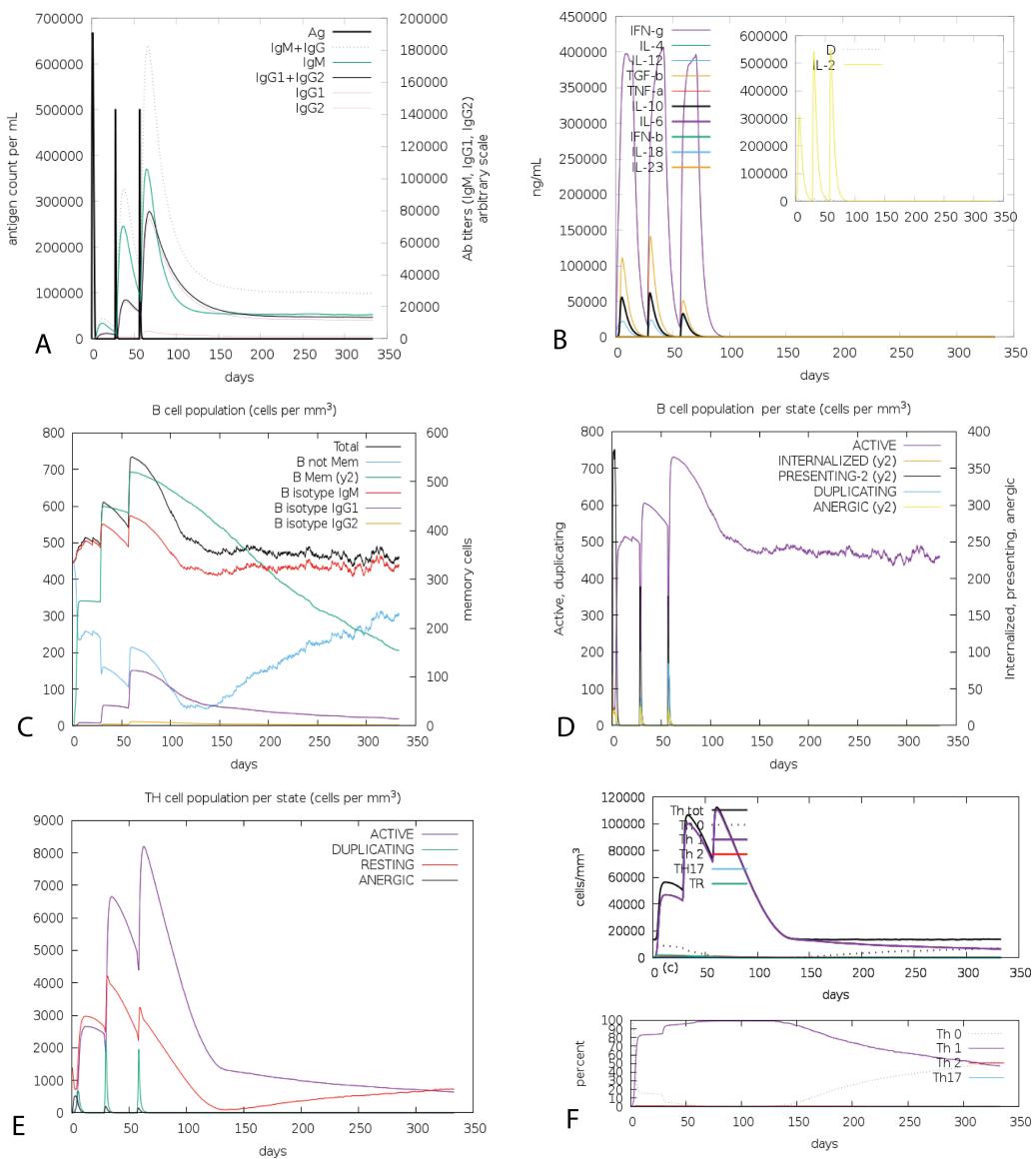


Figure 6 C-ImmSim presentation of an *in silico* immune simulation with the chimeric peptide. (A) The immunoglobulins and the immunocomplex response to antigen (CoV-RMEN) inoculations (black vertical lines); specific subclasses are indicated as colored peaks, (B) concentration of cytokines and interleukins, and inset plot shows danger signal together with leukocyte growth factor IL-2, (C) B-cell populations after three injections, (D) evolution of B cell, (E) T-helper cell populations per state after injections, and (F) evolution of T-helper cell classes with the course of vaccination.

Full-size DOI: 10.7717/peerj.9572/fig-6

to be 99.99% in the Latin American country, Peru (Fig. 7, Data S2). In China, where the viral strain (SARS-CoV-2) first appeared and had more devastating outbreaks, the population coverage for CTL and HTL epitopes was 92.67% and 53.44%, respectively with a combined coverage of 96.59%. SARS-CoV-2 is currently causing serious pandemics in different continents of the globe including Italy, England, Spain, Iran, South Korea and United States of America where the combined population coverage was found to be

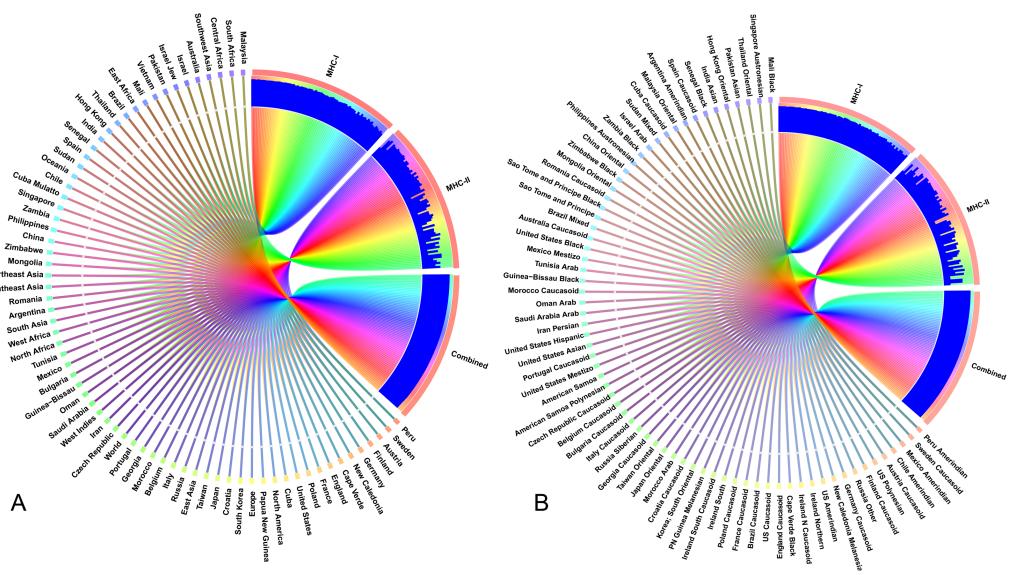


Figure 7 Population coverage of the selected T-cell epitopes and their respective HLA alleles. The circular plot illustrates the relative abundance of the top 70 geographic regions and ethnic groups for selected CTL and HTL epitopes, which were used to construct the vaccine and their corresponding MHC HLA alleles were obtained for population coverage analysis both individually (either MHC-I or MHC-II) and in combination (MHC-I and MHC-II). (A) Population coverage of top seventy geographical regions out of 123 regions. (B) Population coverage of top seventy ethnic groups selected from 146 ethnic groups. Regions and ethnic groups in the respective MHC-I and MHC-II epitopes are represented by different colored ribbons, and the inner blue bars indicate their respective relative coverages. Further information on population coverage analysis is also available in Data S1.

Full-size DOI: 10.7717/peerj.9572/fig-7

98.8%, 99.44%, 95.35%, 98.48%, 99.19% and 99.35%, respectively (Fig. 7A, Data S2). In addition to geographical distribution, the ethnic groups also found to be an important determinant for good coverage of the CTL and HTL epitopes (Fig. 7B). Of the studied 147 ethnic groups, the Peru Amerindian had highest population coverage for CTL (99.98%) while the HTL epitopes had highest population coverage for Austria Caucasoid (88.44%) (Fig. 7B, Data S2). Furthermore, 53.06% of the ethnic groups had a combined population coverage of more than 90.0% for both CTL and HTL epitopes.

Expression prediction of the CoV-RMEN

The length of the optimized codon sequence of the vaccine construct CoV-RMEN in *E. coli* (strain K12) was 1,251 nucleotides. The optimized nucleotide sequence had a Codon Adaptation Index (CAI) of 0.87, and the average GC content of 50.26% showing the possibility of good expression of the vaccine candidate in the *E. coli* host (Figs. 8A–8C). Moreover, the evaluation of minimum free energy for 25 structures of chimeric mRNA through Mfold'server showed that ΔG of the best predicted structure for the optimized construct was $\Delta G = -386.50$ kcal/mol. The first nucleotides at 5' did not have a long stable hairpin or pseudoknot. Therefore, the binding of ribosomes to the translation initiation site, and the following translation process can be readily accomplished in the target host.

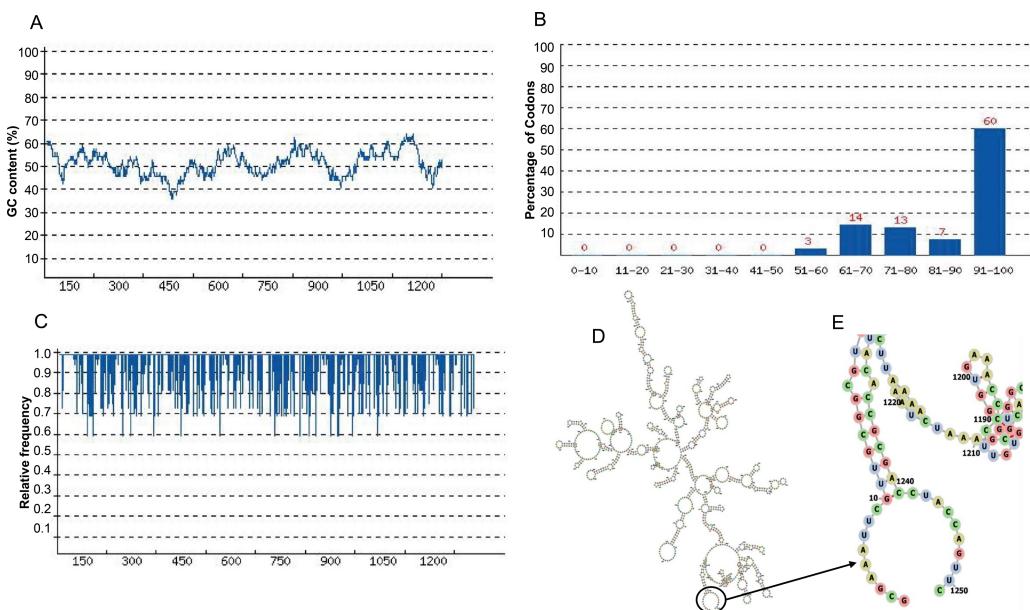


Figure 8 Codon optimization and mRNA structure of CoV-RMEN gene for expression in *E. coli*. (A) GC curve (average GC content: 50.26%) of the optimized CoV-RMEN gene, (B) percentage distribution of codons in computed codon quality groups, (C) relative distribution of codon usage frequency along the gene sequence to be expressed in *E. coli*, and codon adaptation index (CAI) was found to be 0.87 for the desired gene, (D) secondary structure and stability of corresponding mRNA, and (E) resolved view of the start region in the mRNA structure of CoV-RMEN.

Full-size DOI: 10.7717/peerj.9572/fig-8

These outcomes were in the agreement with data obtained from the ‘RNAfold’ web server (Figs. 8D and 8E) where the free energy was -391.37 kcal/mol .

After codon optimization and mRNA secondary structure analysis, the sequence of the recombinant plasmid was designed by inserting the adapted codon sequences into pETite vector (Lucigen, USA), which contains SUMO (Small Ubiquitin-like Modifier) tag and 6x-His tag facilitating both the solubilization and affinity purification of the recombinant protein using SnapGene software (Fig. 9). As alternative to *E. coli* for the expression system, HEK-293 eukaryotic cell line found promising for CoV-RMEN expression. The codon adaption index (CAI), GC content for this system were 1.0 and 61.60 respectively, which indicate high level of expression of the vaccine construct in the HEK-293 cell line as well.

DISCUSSION

SARS-CoV-2, the virus with high zoonotic importance and transmission rate, has spread rapidly around the world and causes life-threatening COVID-19 (*Gorbalenya et al., 2020*). The number of SARS-CoV-2 infections, and subsequent deaths are increasing day by day (*Tai et al., 2020*), and thus, COVID-19 outbreak was declared as a public health emergency by the International Concerns (*Hui et al., 2020; Zhou et al., 2020a*). Scientific community across the world is trying to develop an effective and safe vaccine against this rapidly emerging SARS-CoV-2 (*Abdelmaged et al., 2020*). Although a good number of vaccine

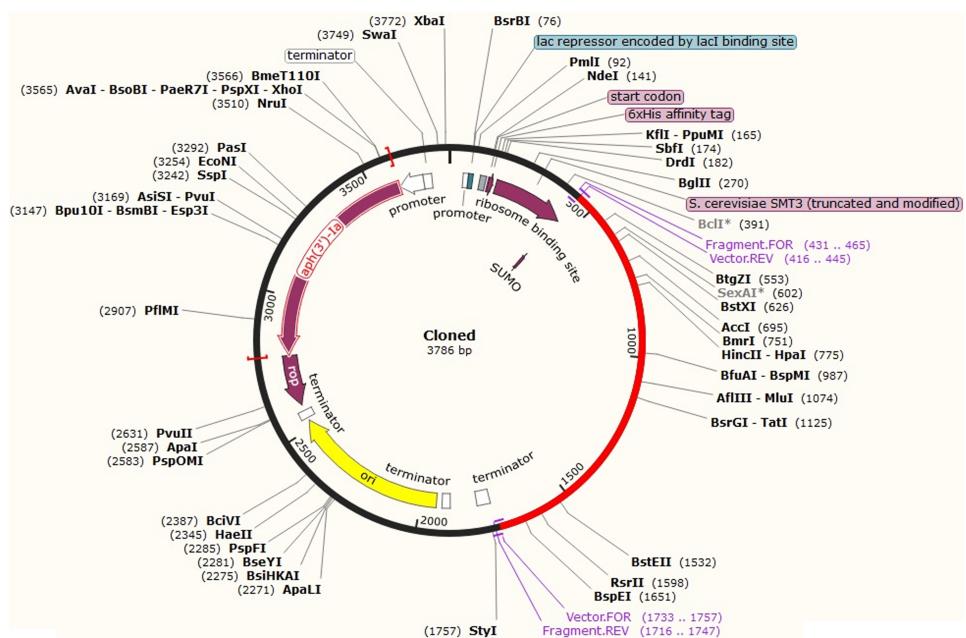


Figure 9 In silico fusion cloning of the CoV-RMEN. The final vaccine candidate sequence was inserted into the pETite expression vector where the red part represents the gene coding for the predicted vaccine, and the black circle represents the vector backbone. The six His-tag and SUMU-tag are located at the Carboxy-terminal end.

Full-size DOI: 10.7717/peerj.9572/fig-9

candidates for COVID-19 are now under trials, some of them are advanced to human trials (Lane, 2020), none has yet been declared to be effective and safe for prevention of SARS-CoV-2 infections. Strikingly, no effective therapeutic drugs or vaccines are yet to be available for the treatment of SARS-CoV-2 patients (Hoque et al., 2020a). Through a comprehensive genomic and proteomic study, we endeavor to design an antigenic multi-epitope (immunodominant) chimeric vaccine for SARS-CoV-2, named as CoV-RMEN (417 aa), which will nullify the involvement of lab-escape viral transmission, reduce the cost, and may elicit immunity by selectively stimulating antigen-specific B- and T-cells.

The novel approach of multi-epitope based (includes conserved multiple epitopes) vaccines designing represents inducing specific cellular immunity, and highly potent neutralizing antibodies against infections (Dawood et al., 2019; Yong et al., 2019; Gralinski & Menachery, 2020; Kibria, Ullah & Miah, 2020). These epitope-based vaccines also provide increased safety and have the ability to focus on sustainable immune responses because of including conserved multiple epitopes. Unlike the full-length S protein, the RBD and NTD segments possess critical neutralizing domains without any non-neutralizing immunodominant region (Ul Qamar et al., 2019; Gralinski & Menachery, 2020; Shang et al., 2020; Wrapp et al., 2020). Mutations on the RBD may enable the new strains to escape neutralization by established RBD-targeting antibodies, hence other functional regions, especially the NTD, should be considered for developing an effective vaccine as well (Wang et al., 2019; Zhou et al., 2019). Besides, combined administration of RBD and NTD proteins induced highly potent neutralizing antibodies and long-term protective immunity in animal

models ([Song et al., 2018](#)). Considering the safety and effectiveness perspectives, the RBD and NTD are more promising candidates in the development of SARS-CoV-2 vaccines over the full-length S protein. The presence of E and M proteins on the envelope can augment the immune response against SARS-CoV ([Millet & Whittaker, 2015](#); [Almofti et al., 2018](#)) and thus, considered for suitable candidate for vaccine development ([Yong et al., 2020](#); [Ahmed, Quadeer & McKay, 2020](#); [Gralinski & Menachery, 2020](#)). Thus, antibodies against the immunologically substantial epitopes of S, M and E proteins of SARS-CoV-2 would provide protective immunity to the infection ([Yong et al., 2020](#); [Ahmed, Quadeer & McKay, 2020](#); [Gralinski & Menachery, 2020](#); [Shang et al., 2020](#)). Therefore, the immune response targeting the RBD and/or NTD of the S glycoprotein, M and E proteins of SARS-CoV-2 would be an important prophylactic and therapeutic interventions, which can be tested further in suitable models before clinical trials ([Chan et al., 2020](#)).

Effective immunity to viral infections is significantly dependent on activation of both B- and T-cells ([Shi et al., 2015](#); [Shey et al., 2019](#)). Therefore, inducing specific humoral or cellular immunity against pathogens, an ideal vaccine should contain both B-cell and T-cell epitopes. Our analyses revealed that selected RBD and NTD regions of the CoV-RMEN contain ample amount of high-affinity B-cell, MHC Class I, MHC Class II and interferon- γ (IFN- γ) epitopes with high antigenicity scores. Moreover, membrane B-cell epitope (MBE) and envelope B-cell epitope (EBE) enhanced the overall stability, immunogenicity and antigenicity of the CoV-RMEN. The development of memory B-cells and T-cells was evident, with memory in B-cells lasting for several months. These finding opposed to several earlier reports where T-cell mediated immune response was considered a long-lasting response compared to B-cells ([Abdelmageed et al., 2020](#); [Wrapp et al., 2020](#)). Another engrossing finding of this study was the development of Th1 response which enhances the growth and proliferation of B- cells augmenting the adaptive immunity ([Carvalho et al., 2002](#)). If a strong B-cell response occurred in animal trials (mice or rabbit), these antibodies could be used in diagnostic purposes, as they should recognize the prominent antigens on the viral surface ([Kibria, Ullah & Miah, 2020](#)). Moreover, CD8+ and CD4+ T-cell responses play major role in antiviral immunity ([Abdelmageed et al., 2020](#)). Another crucial fact is that Toll-Like Receptors (TLRs) can effectively bind with spike protein of the CoV ([Totura et al., 2015](#); [Zander et al., 2017](#)), and might play an important role in the innate immune response to SARS-CoV-2 infection ([Shahabi et al., 2020](#)).

The physicochemical properties also revealed the chimera as a basic or alkaline protein ($pI = 8.71$) and would be thermostable upon expression, and thus, our proposed vaccine CoV-RMEN would be best suited for worldwide use in different endemic areas ([Shey et al., 2019](#); [Ul Qamar et al., 2019](#)). The structural forms (secondary and tertiary) of the CoV-RMEN, when tested as the synthetic peptides, showed the ability to fold into their native structure, hence could mimic the natural infection by SARS-CoV-2 ([Almofti et al., 2018](#)). The refined tertiary (3D) structure of the final vaccine construct markedly presented the desirable structural features based on Ramachandran plot predictions ([Shey et al., 2019](#); [Srivastava et al., 2019](#)). Molecular docking analysis showed that predicted chimeric protein can establish stable protein-protein interactions with TLRs (TLR-2, TLR-3, TLR-4)

(*Totura et al., 2015*). An efficient activation of surface molecules of the CoV-RMEN is very crucial for immune activation of dendritic cells, and subsequent antigen processing and presentation to CD4+ and CD8+ T-cells via MHC-II and MHC-1, respectively (*Shi et al., 2015; Shey et al., 2019; Shang et al., 2020*). The molecular dynamics simulation also revealed that the docked CoV-RMEN-TLRs complexes were stable, and had more binding affinity TLR-3 and TLR-4 (*Abraham et al., 2015*). Furthermore, the CoV-RMEN showed good antigenicity scores on Vaxigen v2.0 and ANTIGENpro indicating that these peptide sequences are supposed to be highly antigenic in nature (*Shey et al., 2019*). The non-allergenic properties of the CoV-RMEN further strengthens its potential as a vaccine candidate (*Shey et al., 2019; Ul Qamar et al., 2019*).

Immune simulation of the CoV-RMEN exhibited expected results consistent with typical immune responses, and there was a growing immune responses after the recurrent antigen exposures (Fig. 6). The antiviral cytokine IFN- γ and cell stimulatory IL-2 level significantly increased, which also contribute to the subsequent immune response after vaccination in host (*Almofti et al., 2018*). This indicates high levels of helper T-cells and consequently efficient antibody production, supporting a humoral response (*Shey et al., 2019*). A lower IC₅₀ value indicates higher binding affinity of the epitopes with the MHC class I and II molecules. While most of the previous studies (*Sakib et al., 2014; Adhikari, Tayebi & Rahman, 2018*) reported that a binding affinity (IC₅₀) threshold of 250 nM identifies peptide binders recognized by T-cells, and this threshold can be used to select peptides, we kept binding affinity within 50 nM to get better confidence level in predicting epitopes for MHC-I and MHC-II alleles. The Simpson index estimated clonal specificity suggested a possible diverse immune response and this is plausible considering the generated chimeric peptide is composed of sufficient B- and T-cell epitopes (Fig. 6).

The interaction between T-cell epitopes, and their respective HLA alleles revealed significant binding affinity reflecting the immune activation of B- and T-cells as supported by other reports (*Srivastava et al., 2019; Jaimes et al., 2020*). T-cell epitopes from RBD and NTD regions showing high interaction with HLA alleles covered more than 98% of the world population with different ethnic groups, and these findings corroborated with many of earlier studies (*Huang et al., 2007; Jaimes et al., 2020; Ul Qamar et al., 2019*). The incorporation of GG and EGGE linkers between the predicted epitopes of the CoV-RMEN produced sequences with minimized junctional immunogenicity, and allowed the rational design construction of a potent multi-epitope vaccine (*Huang et al., 2007; Badawi et al., 2016; Shey et al., 2019; Srivastava et al., 2019*). The glycosylation of the surface antigens helps the enveloped viruses evade recognition by the host immune system, and can influence the ability of the host to raise an effective adaptive immune response (*Pereira et al., 2018*) or even be exploited by the virus to enhance infectivity (*Wolfert & Boons, 2013*). Moreover, some antibodies such as mAb 8ANC195 have evolved to recognize peptide epitope with no dependence on glycan binding (*Kong et al., 2015*). However, there is no data available for antibodies specific to spike glycoproteins of SARS-CoV-2, whether their recognition is interfered by the glycosylation of spike or may either be strengthened by sugars close to the peptide epitope, or not interfered by sugar modification (*Zhou et al., 2020b*). Furthermore, most of the epitopes of the CoV-RMEN harbor no glycoside apart

from the NTD region (Watanabe *et al.*, 2020). Furthermore, most of the glycans of the NTD epitopes present at the terminii of the putative HLA antigens, and may not interfere with the antigen presentation in an HLA complex (Grant *et al.*, 2020). Integrity of the ACE2 receptor of the RBD, envelope protein B-cell epitope (EBE) and membrane protein B-cell epitope (MBE) in the CoV-RMEN suggests that the vaccine may maintain efficacy despite antigenic drift and glycosylation phenomena, as long as the virus continues to target the same host receptor (Grant *et al.*, 2020).

One of the first steps in validating a candidate vaccine is to screen for immunoreactivity through serological analysis. This requires the expression of the recombinant protein in a suitable host. As we focused on the epitopes without the glycosylation or non-significant glycosylation, high-level expression of the vaccine was optimized into well-established and cost-effective prokaryotic expression system *E. coli* K-12 strain as the first choice using the plasmid pETite containing SUMO (Small Ubiquitin-like Modifier) tag and 6x-His tag facilitated both the solubilization and affinity purification of the recombinant protein (Biswal *et al.*, 2015). Codon optimization of the CoV-RMEN revealed its high-level expression in *E. coli* (strain K12). Stable mRNA structure, codon adaptability index (0.87), and the GC content (50.26%) were favourable for high-level expression of the protein in the bacterium. After successful cloning of the gene, the recombinant plasmid can be propagated efficiently using *E. coli* cells, and subsequent protein expression can be performed in *E. coli* K-12 strain using IPTG (Isopropyl β -D-1-thiogalactopyranoside) induction, and cultivation at 28 °C as also reported earlier (Biswal *et al.*, 2015).

As alternative to *E. coli*, eukaryotic cell line, HEK-293 was considered for CoV-RMEN expression. The codon adaption index (CAI), GC content were 1.0 and 61.60 respectively, which indicate high level of expression of vaccine construct in the HEK-293 cell line. In this case, pSec Tag2 mammalian expression vector could be used, which have secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of the recombinant protein, C-terminal poly-histidine (6xHis) tag for rapid purification with C-terminal c-myc epitope for detection with an anti-myc antibody. To remove the affinity and detection tags (His tag and c-myc epitope) after purification, Factor Xa cleavage site (LVPR \downarrow GS) could be added to the C-terminal of the CoV-RMEN (Waugh, 2011).

METHODS

Sequence retrieval and structural analysis

A total of 250 partial and complete genome sequences of SARS-CoV-2 were retrieved from NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/protein>) (Table S5). We aligned these sequences through MAFFT online server (<https://mafft.cbrc.jp/alignment/server/>) using default parameters, and Wu-Kabat protein variability was analyzed (Fig. S9) in protein variability server (<http://imed.med.ucm.es/PVS/>) for SARS-CoV-2 NCBI reference genome (Accession no: NC_045512.2). We retrieved the S protein sequences of the SARS-CoV and MERS-CoV from the whole genome reference sequences of the respective three viruses from the NCBI database. Moreover, the S proteins of SARS-CoV (GenBank accession no: NC_004718.3),

MERS-CoV (GenBank accession no: NC_019843.3) and SARS-CoV-2 were structurally compared using SWISS homology modeling (*Waterhouse et al., 2018*) based on the protein databank (PDB) templates 6acd, 5w9 h and 6vsb, respectively aligned using PyMOL (*Faure et al., 2019*), and observed for heterogeneous domains in the conformations. The models were optimized by energy minimization using the GROMOS96 program (*Van Gunsteren et al., 1996*) implemented within the Swiss-PdbViewer, version 4.1.0 (*Guex & Peitsch, 1997*). The Ramachandran plots of the derived models were evaluated using a PROCHECK (version 3.5)-server to check the stereochemical properties of the modeled structures (*Laskowski et al., 1993*). All further analyses including epitopes selection, antigenicity and allergenicity profiles, molecular docking and final vaccine construct were performed based on the NCBI reference genome of SARS-CoV-2.

Screening for B and T cell epitopes

Conformational B-cell epitopes on the S protein were predicted by ElliPro (<http://tools.iedb.org/ellipro/>) with the minimum score value set at 0.4 while the maximum distance selected as 6 Å (*Kringelum et al., 2012*). The linear B-cell epitopes of RBD and NTD regions of S protein, full length E and M proteins were predicted by “BepiPred Linear Epitope Prediction” (*Larsen, Lund & Nielsen, 2006*), and ABCPred with default parameters (*Saha & Raghava, 2006*). To find out the most probable peptide-based epitopes with better confidence level, antigenicity of the predicted peptides were further verified using VaxiJen antigenicity scores (*Kringelum et al., 2012*). The Kolaskar and Tongaonkar antigenicity profiling from IEDB analysis resource was also used for the RBD and NTD segments, E and M proteins (*Kolaskar & Tongaonkar, 1990*).

CTL epitopes, proteasomal cleavage and transporter associated with antigen processing (TAP), and HTL epitopes of the SARS-CoV-2 S, E and M proteins were predicted using IEDB resource tool Proteasomal cleavage/TAP transport/MHC class I combined predictor (<http://tools.iedb.org/main/tcell/>) with all default parameters. Moreover, the HTL epitopes of the proteins were screened using the IEDB tool “MHC-II Binding Predictions” (<http://tools.iedb.org/mhcii/>). Threshold for strong binding peptides (IC₅₀) was set at 50 nM to determine the binding and interaction potentials of helper T-cell epitopes and both major histocompatibility complex (MHC) class I and alleles (*Shi et al., 2015*). Top five HLA epitopes for each RBD and NTD segments were docked against the respective HLA (MHC-I and MHC-II) allele binders by interaction similarity-based protein-peptide docking system GalaxyPepDock of the GalaxyWeb, docked HLA-epitope complexes were refined in GalaxyRefineComplex and binding affinity (ΔG) was determined PROtein binDIng enerGY prediction (PRODIGY) tool (*Xue et al., 2016*).

IFN-γ-inducing epitope prediction

Potential IFN-γ epitopes of all the selected antigenic sites of RBD, NTD, envelope protein B-cell epitope (EBE), and membrane protein B-cell epitope (MBE) were predicted by “IFNepitope” server (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>). To identify the set of epitopes associated with MHC alleles that would maximize the population coverage,

we adopted the “Motif and SVM hybrid” (MERCI: Motif-Emerging and with Classes-Identification, and SVM) approach. The prediction is based on a dataset of IFN- γ -inducing and IFN- γ -noninducing MHC allele binders ([Dhanda, Vir & Raghava, 2013](#)).

Design and construction of multi-epitope vaccine candidate (CoV-RMEN)

The candidate vaccine (denoted as ‘CoV-RMEN’) design and construction method follows previously published peptide vaccine development protocol for different emerging infectious diseases like SARS and MERS ([Shi et al., 2015](#); [Badawi et al., 2016](#); [Almofti et al., 2018](#); [Shey et al., 2019](#); [Srivastava et al., 2019](#); [Ul Qamar et al., 2019](#)). The multi-epitope protein was constructed by positioning the selected RBD, NTD, MBE and EBE aa sequences linked with short, rigid and flexible linkers GG. To develop highly immunogenic recombinant proteins, two universal T-cell epitopes were used, namely, a pan-human leukocyte antigen DR-binding peptide (PADRE) ([Agadjanyan et al., 2005](#)), and an invasin immunostimulatory sequence taken from *Yersinia* (Invasin) ([Li et al., 2015](#)) were used to the N and C terminal of the vaccine construct respectively, linked by EGGE.

Secondary and tertiary structure of the CoV-RMEN

Chou and Fasman secondary structure prediction server (CFSSP: <https://www.biogem.org/tool/chou-fasman/>), and RaptorX Property (<http://raptorg.uchicago.edu/StructurePropertyPred/predict/>) web-servers were used for secondary structure predictions ([Källberg et al., 2014](#)). The tertiary structure of the CoV-RMEN was built in homology/analogy recognition engine (Phyre2) ([http://www.sbg.bio.ic.ac.uk/\protect\\\$\\relax\svsim \\\$phyre2/html/page.cgi?id=index](http://www.sbg.bio.ic.ac.uk/\protect\$\\relax\svsim \$phyre2/html/page.cgi?id=index)) web-server. The 3D model was refined in a three-step process, initially energy minimization using the GROMOS96 program implemented within the Swiss-PdbViewer (version 4.1.0). After energy minimization, the model was refined using ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) and then GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>).

Furthermore, the local structural quality of the CoV-RMEN was refined with GalaxyRefine server, and ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) was used to calculate overall quality score for the refined structure. The ERRAT server (<http://services.mbi.ucla.edu/ERRAT/>) was also used to analyze non-bonded atom-atom interactions compared to reliable high-resolution crystallography structures. A Ramachandran plot was obtained through the RAMPAGE server ([Lovell et al., 2003](#)).

Physicochemical properties prediction of CoV-RMEN

Moreover, the online web-server ProtParam ([Gasteiger et al., 2005](#)) was used to assess various Physicochemical parameters of the CoV-RMEN including aa residue composition, molecular weight, theoretical pI, instability index, in vitro and in vivo half-life, aliphatic index, and grand average of hydropathicity (GRAVY). The solubility of the multi-epitope vaccine peptide was evaluated using the Protein-Sol server (<https://protein-sol.manchester.ac.uk/>).

Antigenicity, allergenicity and Immune simulation of the CoV-RMEN

VaxiJen 2.0 and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) web-servers were used to predict the antigenicity of the CoV-RMEN while the AllerTOP 2.0 (<http://www.ddg-pharmfac.net/AllerTOP>) and AllergenFP (<http://ddg-pharmfac.net/AllergenFP/>) web-servers were used to predict vaccine allergenicity. Further, the immunogenicity and immune response profile of the CoV-RMEN were characterized by *in silico* immune simulations using the C-ImmSim server (<http://150.146.2.1/C-IMMSIM/index.php>) under default parameters with time steps set at 1, 84, and 170 (each time step is 8 h and time step 1 is injection at time = 0). Therefore, three shots were given at four weeks apart.

Molecular docking and dynamics of the CoV-RMEN with TLRs

Molecular docking of the CoV-RMEN with the TLR2 (PDB ID:3D3I), TLR3 (PDB ID: 1ZIW) and TLR4 (PDB ID: 4G8A) receptors was performed using the High Ambiguity Driven DOCKing (HADDOCK, version 2.4) (*De Vries, Van Dijk & Bonvin, 2010*) web-server to evaluate the interaction between ligand (CoV-RMEN) and receptors (TLRs) leading to an enhanced immune response. CPoRT (<https://milou.science.uu.nl/services/CPoRT/>) was used to predict active interface residues between the CoV-RMEN and TLRs. For the analysis of the stable complex formation, all the complexes (TLRs with CoV-RMEN) were subjected to molecular dynamics (MD) simulation by Gromacs 2020.2 using OPLS-AA (*Abraham et al., 2015; Jorgensen et al., 1996*). Finally, the binding affinities of the docked chimeric protein-TLRs complexes were predicted using the PRODIGY (PROtein binDIng enerGY prediction) (<https://nestor.science.uu.nl/prodigy/>) web-server.

Analysis of cDNA and mRNA for cloning and expression of CoV-RMEN

Reverse translation and codon optimization were performed using the GenScript Rare Codon Analysis Tool (<https://www.genscript.com/tools/rare-codon-analysis>) in order to express the CoV-RMEN in the *E. coli* (strain K12). Stability of the mRNA was verified using two different tools, namely RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>) and the mfold (<http://unafold.rna.albany.edu/?q=mfold>) web-servers. The optimized gene sequence of CoV-RMEN will be artificially synthesized having N-terminal recombinant human rhinovirus (HRV 3C) protease site (LEVLFQ \downarrow GP) and cloned the final construct into pETite vector (Lucigen, USA) through enzyme-free method (*Waugh, 2011*). Finally, the sequence of the recombinant plasmid was designed by inserting the adapted codon sequences into pETite vector using SnapGene software (from Insightful Science; available at snapgene.com). As an alternative to *E. coli*, eukaryotic expression system HEK-293 was optimized using similar analysis for the vaccine production.

Population coverage by CTL and HTL epitopes

The predicted T-cell epitopes were shortlisted based on the aligned Artificial Neural Network (ANN) with half-maximal inhibitory concentration (annIC₅₀) values up to 50 nM. The IEDB “Population Coverage” tool (<http://tools.iedb.org/population/>) was used to determine the world human population coverage by the shortlisted CTL and HTL

epitopes ([Bui et al., 2006](#)). We used OmicsCircos to visualize the association between world population and different ethnic groups ([Hoque et al., 2020b](#)).

CONCLUSIONS

This multi-epitope peptide vaccine candidate, CoV-RMEN possesses potential epitopes from the RBD and NTD segments of spike (S), M and E proteins retaining potential antigenicity and non-allergenicity properties. This chimera, suitable for high-level expression and cloning, includes potential CTL, HTL and B-cell epitopes ensuring humoral and cell mediated immunity, as well as predicted immune-simulation refers to increased production of immunoglobulins and cytokines. Molecular docking and dynamic simulation of the CoV-RMEN with the immune receptors (TLRs) predicted strong binding affinity, in particular with TLR3 and TLR4. Remarkably, the CoV-RMEN had more than 90.0% world population coverage for different ethnic groups. The limitations posed by fewer number of SARS-CoV-2 genome sequence data which tends to mutate frequently may not affect our analysis since we included four peptides of high conservancy from three major proteins of SARS-CoV-2 genome in a multi-epitope vaccine with the high conservancy. However, future in vitro and in vivo studies are required to assess the potentiality of the designed vaccine candidate to induce a positive immune response against SARS-CoV-2 infections, and also to validate the results obtained herein through immuno-informatics analyses.

ACKNOWLEDGEMENTS

The authors thank Joynob Akter Puspo, Masuda Akter and Israt Islam (MS student), and DR. Kazi Alamgir Hossain (PhD Fellow) of the Microbial Genetics and Bioinformatics Laboratory, Department of Microbiology, University of Dhaka for their cooperation, suggestions and overall encouragement during the preparation of the manuscript. The authors also extend their thanks to those who made their sequence data available in NCBI.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

Keith A. Crandall is an Academic Editor for PeerJ. The authors declare there are no competing interests.

Author Contributions

- M. Shaminur Rahman, M. Nazmul Hoque, M. Rafiul Islam, Salma Akter and ASM Rubayet-Ul-Alam performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Mohammad Anwar Siddique analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

- Otun Saha analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Md. Mizanur Rahaman, Munawar Sultana, Keith A. Crandall and M. Anwar Hossain conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Accession numbers for the third-party sequences are available in [Table S1](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.9572#supplemental-information>.

REFERENCES

- Abdelmageed MI, Abdelmoneim AH, Mustafa MI, Elfadol NM, Murshed NS, Shantier SW, Makhawi AM.** 2020. Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach. *BioMed Research International* 2020:2683286 DOI [10.1155/2020/2683286](https://doi.org/10.1155/2020/2683286).
- Abraham MJ, Murtola T, Schulz R, Pál S, Smith JC, Hess B, Lindahl E.** 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2(2):19–25.
- Adhikari UK, Tayebi M, Rahman MM.** 2018. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *Journal of Immunology Research* 2018:6718083 DOI [10.1155/2018/6718083](https://doi.org/10.1155/2018/6718083).
- Agadjanyan MG, Ghochikyan A, Petrushina I, Vasilevko V, Movsesyan N, Mkrtchyan M, Tommy S, Cribbs DH.** 2005. Prototype Alzheimer's disease vaccine using the immunodominant B cell epitope from β -amyloid and promiscuous T cell epitope pan HLA DR-binding peptide. *The Journal of Immunology* 174(3):1580–1586 DOI [10.4049/jimmunol.174.3.1580](https://doi.org/10.4049/jimmunol.174.3.1580).
- Ahmed SF, Quadeer AA, McKay MR.** 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 12(3):254 DOI [10.3390/v12030254](https://doi.org/10.3390/v12030254).
- Akhand MRN, Azim KF, Hoque SF, Moli MA, Joy BD, Akter H, Afif IK, Ahmed N, Hasan M.** 2020. Genome based evolutionary study of SARS-CoV-2 towards the prediction of epitope based chimeric vaccine. *BioRxiv* DOI [10.1101/2020.04.15.036285](https://doi.org/10.1101/2020.04.15.036285).
- Almofti YA, Abd-elrahman KA, Gassmallah SAE, Salih MA.** 2018. Multi epitopes vaccine prediction against Severe Acute Respiratory Syndrome (SARS) coronavirus using immunoinformatics approaches. *American Journal of Microbiological Research* 6(3):94–114 DOI [10.12691/ajmr-6-3-5](https://doi.org/10.12691/ajmr-6-3-5).
- Badawi MM, SalahEldin MA, Suliman MM, AbduRahim SA, Aelghafoor Mohammed A, SidAhmed ASA, Marwa MO, Salih MA.** 2016. *In silico* prediction of a novel

- universal multi-epitope peptide vaccine in the whole spike glycoprotein of MERS CoV. *American Journal of Microbiological Research* 4(4):101–121.
- Biswal JK, Bisht P, Mohapatra JK, Ranjan R, Sanyal A, Pattnaik B.** 2015. Application of a recombinant capsid polyprotein (P1) expressed in a prokaryotic system to detect antibodies against foot-and-mouth disease virus serotype O. *Journal of Virological Methods* 215:45–51.
- Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A.** 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7(1):153 DOI [10.1186/1471-2105-7-153](https://doi.org/10.1186/1471-2105-7-153).
- Carvalho LH, Sano GI, Hafalla JC, Morrot A, De Lafaille MAC, Zavala F.** 2002. IL-4-secreting CD4+ T cells are crucial to the development of CD8+ T-cell responses against malaria liver stages. *Nature Medicine* 8:166–170 DOI [10.1038/nm0202-166](https://doi.org/10.1038/nm0202-166).
- Chan JFW, Kok KH, Zhu Z, Chu H, To KK, Yuen S, Yuen KY.** 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes and Infections* 9:221–236 DOI [10.1080/22221751.2020.1719902](https://doi.org/10.1080/22221751.2020.1719902).
- Dawood RM, Moustafa RI, Abdelhafez TH, El-Shenawy R, El-Abd Y, El Din NGB, Jean D, El Awady MK.** 2019. A multiepitope peptide vaccine against HCV stimulates neutralizing humoral and persistent cellular responses in mice. *BMC Infectious Diseases* 19:932 DOI [10.1186/s12879-019-4571-5](https://doi.org/10.1186/s12879-019-4571-5).
- De Vries SJ, Van Dijk M, Bonvin AM.** 2010. The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols* 5:883–897 DOI [10.1038/nprot.2010.32](https://doi.org/10.1038/nprot.2010.32).
- Dhanda SK, Vir P, Raghava GP.** 2013. Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct* 8(1):30 DOI [10.1186/1745-6150-8-30](https://doi.org/10.1186/1745-6150-8-30).
- Dudek LN, Perlmutter P, Aguilar I, Croft PN, Purcell WA.** 2010. Epitope discovery and their use in peptide based vaccines. *Current Pharmaceutical Design* 16(28):3149–3157 DOI [10.2174/138161210793292447](https://doi.org/10.2174/138161210793292447).
- Faure G, Joseph AP, Craveur P, Narwani TJ, Srinivasan N, Gelly J-C, Rebehmed J, de Brevern AG.** 2019. iPBAvizu: a PyMOL plugin for an efficient 3D protein structure superimposition approach. *Source Code for Biology and Medicine* 14:1–5 DOI [10.1186/s13029-019-0075-3](https://doi.org/10.1186/s13029-019-0075-3).
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A.** 2005. Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook*. Cham: Springer Nature Switzerland AG, 571–607.
- Gorbalenya AE, Baker SC, Baric RS, De Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW, Penzar D.** 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5(4):536–544 DOI [10.1038/s41564-020-0695-z](https://doi.org/10.1038/s41564-020-0695-z).
- Gralinski LE, Menachery VD.** 2020. Return of the coronavirus: 2019-nCoV. *Viruses* 12:1–8 DOI [10.3390/v12020135](https://doi.org/10.3390/v12020135).
- Grant OC, Montgomery D, Ito K, Woods RJ.** 2020. Analysis of the SARS-CoV-2 spike protein glycan shield: implications for immune recognition. *BioRxiv* DOI [10.1101/2020.04.07.030445](https://doi.org/10.1101/2020.04.07.030445).

- Guex N, Peitsch MC.** 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15):2714–2723 DOI [10.1002/elps.1150181505](https://doi.org/10.1002/elps.1150181505).
- Hoque MN, Chaudhury A, Akanda MAM, Hossain MA, Islam MT.** 2020a. Genomic diversity and evolution, diagnosis, prevention, and therapeutics of the pandemic COVID-19 disease. *Preprints* **2020**:2020040359 DOI [10.20944/preprints202004.0359.v1](https://doi.org/10.20944/preprints202004.0359.v1).
- Hoque MN, Istiaq A, Clement RA, Gibson KM, Saha O, Islam OK, Abir RA, Sultana M, Siddiki AMAM, Crandall KA, Hossain MA.** 2020b. Insights into the resistome of bovine clinical mastitis microbiome, a key factor in disease complication. *Frontiers in Microbiology* **11**:860 DOI [10.3389/fmicb.2020.00860](https://doi.org/10.3389/fmicb.2020.00860).
- Huang H, Hao S, Li F, Ye Z, Yang J, Xiang J.** 2007. CD4(+) Th1 cells promote CD8(+) Tc1 cell survival, memory response, tumor localization and therapy by targeted delivery of interleukin 2 via acquired pMHC I complexes. *Immunology* **120**:148–159 DOI [10.1111/j.1365-2567.2006.02452.x](https://doi.org/10.1111/j.1365-2567.2006.02452.x).
- Hui DSIAE, Madani TA, Ntoumi F, Koch R, Dar O.** 2020. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases* **91**:264–266 DOI [10.1016/j.ijid.2020.01.009](https://doi.org/10.1016/j.ijid.2020.01.009).
- Jaimes JA, André NM, Millet JK, Whittaker GR.** 2020. Structural modeling of 2019-novel coronavirus (nCoV) spike protein reveals a proteolytically-sensitive activation loop as a distinguishing feature compared to SARS-CoV and related SARS-like coronaviruses. *arXiv Preprint ArXiv preprint*. arXiv:2002.06196.
- Jorgensen WL, Maxwell DS, Tirado-Rives J.** 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*. 11225–11236.
- Källberg M, Margaryan G, Wang S, Ma J, Xu J.** 2014. RaptorX server: a resource for template-based protein structure modeling. In: *Protein structure prediction*. New York: Humana Press, 17–27.
- Kibria KMK, Ullah H, Miah M.** 2020. The multi-epitope vaccine prediction to combat Pandemic SARS-CoV-2, an immunoinformatic approach. *Preprints* DOI [10.21203/rs.3.rs-21853/v1](https://doi.org/10.21203/rs.3.rs-21853/v1).
- Kolaskar A, Tongaonkar PC.** 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters* **276**(1-2):172–174 DOI [10.1016/0014-5793\(90\)80535-Q](https://doi.org/10.1016/0014-5793(90)80535-Q).
- Kong L, Torrents de la Peña A, Deller MC, Garces F, Sliepen K, Hua Y, Stanfield RL, Sanders RW, Wilson IA.** 2015. Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. *Acta Crystallography D Biology Crystallography* **71**(Pt 10):2099–2108 DOI [10.1107/S1399004715013917](https://doi.org/10.1107/S1399004715013917).
- Kringelum JV, Lundsgaard C, Lund O, Nielsen M.** 2012. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLOS Computational Biology* **8**(12):e1002829 DOI [10.1371/journal.pcbi.1002829](https://doi.org/10.1371/journal.pcbi.1002829).

- Lane R.** 2020. Sarah Gilbert: carving a path towards a COVID-19 vaccine. *Lancet* **395**(10232):1247 DOI [10.1016/S0140-6736\(20\)30796-0](https://doi.org/10.1016/S0140-6736(20)30796-0).
- Larsen JEP, Lund O, Nielsen M.** 2006. Improved method for predicting linear B-cell epitopes. *Immunome Research* **2**(1):2 DOI [10.1186/1745-7580-2-2](https://doi.org/10.1186/1745-7580-2-2).
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM.** 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**(2):283–291 DOI [10.1107/S0021889892009944](https://doi.org/10.1107/S0021889892009944).
- Le TT, Andreadakis Z, Kumar A, Roman RG, Tollefsen S, Saville M, Mayhew S.** 2020. The COVID-19 vaccine development landscape. *Nature Reviews Drug Discovery* **19**(5):305–306 DOI [10.1038/d41573-020-00073-5](https://doi.org/10.1038/d41573-020-00073-5).
- Li H, Ning P, Lin Z, Liang W, Kang K, He L, Zhang Y.** 2015. Co-expression of the C-terminal domain of *Yersinia enterocolitica* invasin enhances the efficacy of classical swine-fever-vectored vaccine based on human adenovirus. *Journal of Biosciences* **40**:79–90 DOI [10.1007/s12038-014-9495-z](https://doi.org/10.1007/s12038-014-9495-z).
- Lovell SC, Davis IW, Arendall III WB, De Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC.** 2003. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, Bioinformatics* **50**(3):437–450 DOI [10.1002/prot.10286](https://doi.org/10.1002/prot.10286).
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y.** 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**(10224):565–574 DOI [10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- Millet JK, Whittaker GR.** 2015. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Research* **202**:120–134 DOI [10.1016/j.virusres.2014.11.021](https://doi.org/10.1016/j.virusres.2014.11.021).
- Pallesen J, Wang N, Corbett KS, Wrapp D, Kirchdoerfer RN, Turner HL, Cottrell CA, Becker MM, Wang L, Shi W, Kong WP.** 2017. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences of the United States of America* **114**(35):E7348–E7357 DOI [10.1073/pnas.1707304114](https://doi.org/10.1073/pnas.1707304114).
- Pereira MS, Alves I, Vicente M, Campar A, Silva MC, Padrão NA, Pinto V, Fernandes Â, Dias AM, Pinho SS.** 2018. Glycans as key checkpoints of T cell activity and function. *Frontiers in Immunology* **9**:2754 DOI [10.3389/fimmu.2018.02754](https://doi.org/10.3389/fimmu.2018.02754).
- Saha S, Raghava GPS.** 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, Bioinformatics* **65**(1):40–48 DOI [10.1002/prot.21078](https://doi.org/10.1002/prot.21078).
- Sakib MS, Islam M, Hasan AKM, Nabi AHM.** 2014. Prediction of epitope-based peptides for the utility of vaccine development from fusion and glycoprotein of nipah virus using *in silico* approach. *Advances in Bioinformatics* **2014**:402492 DOI [10.1155/2014/402492](https://doi.org/10.1155/2014/402492).
- Schoeman D, Fielding BC.** 2019. Coronavirus envelope protein: current knowledge. *Virology Journal* **16**:69 DOI [10.1186/s12985-019-1182-0](https://doi.org/10.1186/s12985-019-1182-0).

- Shahabi nezhad F, Mosaddeghi P, Negahdaripour M, Dehghani Z, Farahmandnejad M, Moghadami M, Nezafat N, Masoompour SM. 2020.** Therapeutic Approaches for COVID-19 based on the dynamics of interferon-mediated immune responses. *Preprints* 2020030206 DOI 10.20944/preprints202003.0206.v2.
- Shang W, Yang Y, Rao Y, Rao X. 2020.** The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. *npj Vaccines* 5:18 DOI 10.1038/s41541-020-0170-0.
- Sheridan C. 2020.** Questions remain following first COVID-19 vaccine results. *Nature Biotechnology* Available at <https://www.nature.com/articles/d41587-020-00015-x>.
- Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, Asa BF, Ngale FN, Vanhamme L, Souopgui J. 2019.** *In-silico* design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports* 9(1):1–18 DOI 10.1038/s41598-018-37186-2.
- Shi J, Zhang J, Li S, Sun J, Teng Y, Wu M, Li J, Li Y, Hu N, Wang H, Hu Y. 2015.** Epitope-based vaccine target screening against highly pathogenic MERS-CoV: an *in silico* approach applied to emerging infectious diseases. *PLOS ONE* 10(12):e0144475 DOI 10.1371/journal.pone.0144475.
- Shi SQ, Peng JP, Li YC, Qin C, Liang GD, Xu L, Yang Y, Wang JL, Sun QH. 2006.** The expression of membrane protein augments the specific responses induced by SARS-CoV nucleocapsid DNA immunization. *Molecular Immunology* 43:1791–1798 DOI 10.1016/j.molimm.2005.11.005.
- Song W, Gui M, Wang X, Xiang Y. 2018.** Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLOS Pathogens* 14(8):e1007236 DOI 10.1371/journal.ppat.1007236.
- Srivastava S, Kamthania M, Kumar Pandey R, Kumar Saxena A, Saxena V, Kumar Singh S, Kumar Sharma R, Sharma N. 2019.** Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. *Journal of Biomolecular Structure and Dynamics* 37(16):4345–4360 DOI 10.1080/07391102.2018.1548977.
- Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. 2020.** Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology* 19:1–8 DOI 10.1038/s41423-020-0400-4.
- Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J. 2020.** On the origin and continuing evolution of SARS-CoV-2. *National Science Review* 2020:nwaa036 DOI 10.1093/nsr/nwaa036.
- Totura AL, Whitmore A, Agnihothram S, Schäfer A, Katze MG, Heise MT, Baric RS. 2015.** Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *MBio* 6(3):e00638-15.
- Ul Qamar MT, Saleem S, Ashfaq UA, Bari A, Anwar F, Alqahtani S. 2019.** Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *Journal of Translational Medicine* 17(1):362 DOI 10.1186/s12967-019-2116-8.

- Van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark A, Scott W, Tironi I.** 1996. Biomolecular simulation: the GROMOS96 manual and user guide. In: *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*. 86. 1–1044.
- Wang N, Rosen O, Wang L, Turner HL, Stevens LJ, Corbett KS, Bowman CA, Pallesen J, Shi W, Zhang Y, Leung K.** 2019. Structural definition of a neutralization-sensitive epitope on the MERS-CoV S1-NTD. *Cell Reports* 28(13):3395–3405 DOI [10.1016/j.celrep.2019.08.052](https://doi.org/10.1016/j.celrep.2019.08.052).
- Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M.** 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* 2020:eabb9983 DOI [10.1126/science.abb9983](https://doi.org/10.1126/science.abb9983).
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R.** 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46(W1):W296–W303 DOI [10.1093/nar/gky427](https://doi.org/10.1093/nar/gky427).
- Waugh DS.** 2011. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification* 80(2):283–293 DOI [10.1016/j.pep.2011.08.005](https://doi.org/10.1016/j.pep.2011.08.005).
- World Health Organization (WHO).** 2020. Novel Coronavirus (2019-nCoV) situation reports—World Health Organization (WHO).
- Wolfert MA, Boons GJ.** 2013. Adaptive immune activation: glycosylation does matter. *Nature Chemical Biology* 9:776–784 DOI [10.1038/nchembio.1403](https://doi.org/10.1038/nchembio.1403).
- Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS.** 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367(6483):1260–1263 DOI [10.1126/science.abb2507](https://doi.org/10.1126/science.abb2507).
- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML.** 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269 DOI [10.1038/s41586-020-2008-3](https://doi.org/10.1038/s41586-020-2008-3).
- Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A.** 2016. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 32(23):3676–3678.
- Yazdani Z, Rafiei A, Yazdani M, Valadan R.** 2020. Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: an *in silico* analysis. *bioRxiv* DOI [10.1101/2020.04.20.2051557](https://doi.org/10.1101/2020.04.20.2051557).
- Yong CY, Ong HK, Yeap SK, Ho KL, Tan WS.** 2019. Recent advances in the vaccine development against Middle East respiratory syndrome–Coronavirus. *Frontiers in Microbiology* 10:1–18 DOI [10.3389/fmicb.2019.01781](https://doi.org/10.3389/fmicb.2019.01781).
- Yong SEF, Anderson DE, Wei WE, Pang J, Chia WN, Tan CW, Teoh YL, Rajendram P, Toh MPHS, Poh C.** 2020. Connecting clusters of COVID-19: an epidemiological and serological investigation. *The Lancet Infectious Diseases* 20:809–815 DOI [10.1016/S1473-3099\(20\)30273-5](https://doi.org/10.1016/S1473-3099(20)30273-5).
- Zander RA, Vijay R, Pack AD, Guthmiller JJ, Graham AC, Lindner SE, Vaughan AM, Kappe SH, Butler NS.** 2017. Th1-like plasmodium-specific memory CD4+ T cells support humoral immunity. *Cell Reports* 21(7):1839–1852 DOI [10.1016/j.celrep.2017.10.077](https://doi.org/10.1016/j.celrep.2017.10.077).

- Zhou D, Qi R, Zhang W, Tian X, Peng C.** 2020b. Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface glycopeptide motifs: implications on vaccination and antibody therapeutics. *Preprints* 2020020381
[DOI 10.20944/preprints202002.0381.v2](https://doi.org/10.20944/preprints202002.0381.v2).
- Zhou H, Chen Y, Zhang S, Niu P, Qin K, Jia W, Huang B, Zhang S, Lan J, Zhang L.** 2019. Structural definition of a neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein. *Nature Communications* 10:1–13
[DOI 10.1038/s41467-019-10897-4](https://doi.org/10.1038/s41467-019-10897-4).
- Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G, Shi Z-L.** 2020a. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *Nature* 579:270–273
[DOI 10.1038/s41586-020-2012-7](https://doi.org/10.1038/s41586-020-2012-7).