



OPEN

Immunoinformatic design of a COVID-19 subunit vaccine using entire structural immunogenic epitopes of SARS-CoV-2

Esmail Behmard^{1,2}, Bijan Soleymani¹, Ali Najafi³✉ & Ebrahim Barzegari¹✉

Coronavirus disease 2019 (COVID-19) is an acute pneumonic disease, with no prophylactic or specific therapeutical solution. Effective and rapid countermeasure against the spread of the disease's associated virus, SARS-CoV-2, requires to incorporate the computational approach. In this study, we employed various immunoinformatics tools to design a multi-epitope vaccine polypeptide with the highest potential for activating the human immune system against SARS-CoV-2. The initial epitope set was extracted from the whole set of viral structural proteins. Potential non-toxic and non-allergenic T-cell and B-cell binding and cytokine inducing epitopes were then identified through a priori prediction. Selected epitopes were bound to each other with appropriate linkers, followed by appending a suitable adjuvant to increase the immunogenicity of the vaccine polypeptide. Molecular modelling of the 3D structure of the vaccine construct, docking, molecular dynamics simulations and free energy calculations confirmed that the vaccine peptide had high affinity for Toll-like receptor 3 binding, and that the vaccine-receptor complex was highly stable. As our vaccine polypeptide design captures the advantages of structural epitopes and simultaneously integrates precautions to avoid relevant side effects, it is suggested to be promising for elicitation of an effective and safe immune response against SARS-CoV-2 in vivo.

Coronaviruses (CoVs) named for the crown-like spikes on their surface are grouped into four main types, known as α , β , γ and δ ¹. In recent decades, new human β -CoVs have evolved from animal-infecting types, leading to intercontinental pneumonic outbreaks. Severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are two highlight examples, which engendered epidemics in numerous countries worldwide in 2002 and 2012, respectively. SARS-CoV-2 was a return of coronaviruses in December 2019, and is the etiologic agent for the ongoing coronavirus disease 2019 (COVID-19). The viral spill-over originated from a Wuhan seafood market, the place where horseshoe bat SARS-like coronavirus was believed to transmit to humans. By 21st September 2020, over 30.6 million cases and 950,000 deaths have been reported to World Health Organization².

Identifying an appropriate preventive, i.e. a COVID-19 vaccine, is critical for apt response to the SARS-CoV-2 mass contagion. There are diverse candidate types for immunogen development, including but not limited to live-attenuated or inactivated whole virus, DNA vaccines, vectored vaccines, subunit vaccines, and self-assembling virus-like particle vaccines, each with particular advantages and disadvantages. Vaccines based on chemically inactivated CoV virions have led to production of neutralizing antibodies with different levels of protection, but they raise potential side effects and biosafety concerns³. DNA vaccines expressing full-length spike protein or its fragments, as well as DNA priming coupled with protein boosting, have also been effective against MERS-CoV infection^{4,5}. By intracellular expression of immunogenic antigens, vectored vaccines allow activation of cellular immune response in addition to humoral immunity; they have a proven safety record, however these vectors are limited to presenting one or a reduced number of CoV antigens to the immune system⁶. Subunit vaccines

¹Medical Biology Research Center, Health Technology Institute, Kermanshah University of Medical Sciences, Zakariya Razi Blvd., Kermanshah, Iran. ²Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ³Molecular Biology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran. ✉email: najafi74@bmsu.ac.ir; e.barzegari@kums.ac.ir

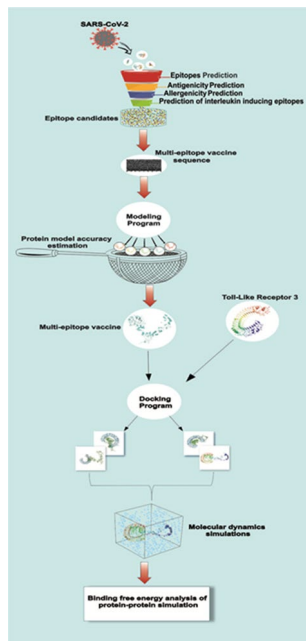


Figure 1. Systemic flowchart of the multi-epitope subunit vaccine building against COVID-19.

based on recombinant spike protein have been shown to be very well suited for creating immunogenicity against SARS and MERS^{7,8}. While DNA, vectored and subunit candidates presenting viral epitopes can elicit a focused antibody response, their subviral components may not portray the full antigenic complexity of the virus, resulting in limited protective efficacy or immunopathology due to unbalanced immune responses⁶. One promising option to confront this issue is the subunit vaccines harbouring numerous and diverse antigenic elements, allowing to produce an inclusive spectrum of native viral antigens.

The process of vaccine development takes at least ten years from bench research to approved vaccine use⁹. Identification of apt elements for developing vaccine immunogens can be promoted using chemistry and topology. As a current trend, this is becoming more of a rational design exercise, which may considerably reduce both the time and costs required for vaccine development and production¹⁰. In this study, it was hypothesized that advanced computational screening could identify suitable peptides for the construction of a subunit vaccine. We used the full set of the putative structural proteins from SARS-CoV-2 as the substance for extracting antigenic elements creating both B-cell and T-cell immunity. By integrating these peptides together, we developed a multi-epitope-based vaccine polypeptide with appropriate binding to Toll-like receptor 3 (TLR-3) to elicit an effective immune response, and with least possible toxicity and hypersensitivity. We also considered incorporating an adjuvant for augmenting the host antigen-specific immune response. As this method has been validated experimentally with designs against other pathogenic species¹¹, we suggest that the proposed structural formulation would be potential to generate an effective immune response against the novel virus. The wet lab researchers are expected to validate our design, hoping to reach protection for the healthy community against the COVID-19 pandemic.

Results

The basic steps of the procedure for designing the multi-epitope vaccine are shown in Fig. 1.

Structural T-cell and B-cell epitopes of SARS-CoV-2. Final selected cytotoxic T lymphocyte (CTL) antigenic epitopes harboured in the structural proteins of SARS-CoV-2 are listed in Table 1, indicating the epitopes for spike glycoprotein (S), envelope protein (E), membrane protein (M) and nucleocapsid phosphoprotein (N). The full sets of predicted epitopes and their properties can be found in Supplementary Tables S1–S4 online, for S, E, M and N proteins, respectively. The IEDB MHC-II prediction tool was applied to predict helper T lymphocyte (HTL) (15-mer) epitopes and their MHC-II binding, results of which are shown in Table 2, and with more detail in Supplementary Table S5 online. Both Tables 1 and 2 also indicate the HLA restriction of the single predicted epitopes. The peptide DVSLVKPSFYVYSRVK contained in the E protein was identified as the only linear B lymphocyte (LBL) epitope, when applying the score value >0.75 as the criterion for selection (More detail in Supplementary Table S6 online). Using appropriate server tools, antigenic, non-toxic and non-allergenic T-cell or B-cell epitopes that were able to induce interferon-gamma (IFN- γ), interleukin-4 (IL-4) and interleukin-10 (IL-10) cytokines were selected for the design of a multi-epitope vaccine.

Multi-epitope vaccine polypeptide construction. The total of 34 CTL, and 12 HTL epitopic peptides were fused to each other by KK, and GPGPG linkers, respectively, followed by adjoining a single LBL epitope

Protein	Peptide	IC ₅₀	Antigenicity	Allele
S	³⁸⁶ KLNDLCFTNV ³⁹⁵	6.22	2.6927	HLA-A*02:03; HLA-A*02:01
S	³²⁹ FPNITNLCP ³³⁸	8.63	1.3964	HLA-B*53:01; HLA-B*35:01
S	²⁰⁰ FKIYSKHTPI ²⁰⁹	9.13	1.016	HLA-A*02:03
S	¹⁰⁶⁰ VVFLHVTVY ¹⁰⁶⁸	13.02	1.5122	HLA-A*02:03; HLA-A*02:06; HLA-A*02:01
S	⁵⁸⁷ ITPCSFGGV ⁵⁹⁵	20.32	1.3871	HLA-A*68:02; HLA-A*02:06
S	¹²⁰⁷ EQYIKWPWYI ¹²¹⁶	21.01	1.1122	HLA-A*23:01; HLA-A*24:02
S	²⁶⁵ YYVGYLQPR ²⁷³	21.9	1.4692	HLA-A*33:01
S	²²⁹ LPIGINITRF ²³⁸	25.21	1.3027	HLA-B*53:01; HLA-B*35:01
S	⁵¹² VLSFELLHA ⁵²⁰	33.91	1.0776	HLA-A*02:03
S	⁴⁰⁸ RQIAPGQTGK ⁴¹⁷	38	1.7893	HLA-A*03:01
S	¹¹⁹⁶ SLIDLQELGK ¹²⁰⁵	38.67	1.0275	HLA-A*11:01
S	⁶⁴⁴ QTRAGCLIGA ⁶⁵³	39.78	1.3933	HLA-A*68:02
E	⁶¹ RVKNLNSSR ⁶⁹	4.68	0.8998	HLA-A*31:01; HLA-A*30:01
E	¹⁸ LLFLAFVVF ²⁶	8.32	0.8144	HLA-B*15:01
E	⁵⁷ YVYSRVKNL ⁶⁵	15.74	0.702	HLA-A*02:03
E	²⁹ VTLAILTALR ³⁸	26.92	0.8404	HLA-A*68:01
E	²³ FVVFLVTL ³¹	27.7	0.7403	HLA-A*02:06
E	²⁶ FLVTLAIL ³⁴	39.95	0.9645	HLA-A*02:01
E	⁴⁵ NIVNVSLVK ⁵³	46.43	0.931	HLA-A*68:01
M	¹⁹ QWNLVIGFLF ²⁸	12.03	1.2302	HLA-A*23:01
M	¹² IAMACLVLG ²⁰	13.64	1.1306	HLA-A*68:02
M	⁶ GTITVEELK ¹⁴	18.08	1.0976	HLA-A*68:01; HLA-A*11:01
M	³⁵ RTRSMWSFNP ⁴⁴	18.67	1.591	HLA-A*30:01
M	⁵¹ SGFAAYSRYR ⁶⁰	24.22	1.0034	HLA-A*31:01
M	²² LVIGFLFLT ³⁰	25.12	1.2619	HLA-A*02:06
M	²⁶ FLFLTWICL ³⁴	32.26	1.4835	HLA-A*02:01
M	¹⁰ IAIAMACLVL ¹⁸	34.16	1.1704	HLA-A*02:06
M	⁶⁰ VTLACFVLA ⁷⁰	40.96	1.3368	HLA-A*02:03
M	²⁹ SFRLFARTR ³⁷	41.19	0.7038	HLA-A*31:01; HLA-A*33:01
N	³¹⁵ FGMSRIGMEV ³²⁴	7.16	0.88	HLA-A*02:03; HLA-A*02:01
N	³⁶¹ KTFPPTPEPK ³⁷⁰	11.43	0.7657	HLA-A*30:01; HLA-A*11:01
N	¹⁰⁰ KMKDLSR ¹⁰⁷	13.15	1.7575	HLA-A*31:01
N	¹⁹³ SSRNSTPGS ²⁰¹	32.96	1.2424	HLA-A*30:01
N	¹⁰⁴ LSPRWYFYLL ¹¹³	37.51	1.3486	HLA-B*08:01

Table 1. The final set of cytotoxic T lymphocyte (CTL) epitopes selected for multi-epitope vaccine construction.

using the KK linker, to create the multi-epitope peptide-based vaccine construct^{12–14}. Furthermore, to boost the immunogenicity of the multi-epitope vaccine, β -defensin (GIINTLQKYYCRVRRGRCVLSCLPKKEEQIGK-STRGRKCCRRK) was added as an adjuvant to the amino terminus of the polypeptide using an EAAAK linker to the first CTL epitope. The primary structure of the multi-epitope subunit vaccine construct included the total of 694 amino acids (Fig. 2 and Supplementary Fig. S1 online).

Immunogenicity, allergenicity and physicochemical properties of the vaccine candidate. Assessment of immunogenic, allergenic and solubility showed the multi-epitope vaccine candidate was of a suitable design (Table 3). The calculated molecular weight of the final vaccine polypeptide (78.35 kDa) indicated its good antigenic nature^{13,15}, and the pI value of 10.38 showed the basic nature of the final vaccine candidate. The instability index of the final polypeptide (36.73) indicated its high stability. The Grand average of hydropathicity value was -0.261 , showing that the final structure is slightly hydrophilic in nature, which can lead to better connection with other proteins. Moreover, the aliphatic index of the peptide was equal to 81.44, which implied the final peptide had high thermo-stability (Table 3). Estimated half-life was 30 h in mammalian reticulocytes in vitro, and >20 h in yeast, and >10 h in *E. coli* in vivo.

3D structure of the vaccine polypeptide. The sequence of the multi-epitope vaccine polypeptide was submitted to appropriate tools, and a refined model of its 3D structure was obtained (Supplementary Fig. S1 online). The output from ProSA-web validation tools showed a Z-score of -4.33 , indicating the good quality of the final vaccine polypeptide structure (Supplementary Fig. S1 online). In addition, RAMPAGE analysis indicated that 92.3% amino acids of the final structure were in the favoured area, 6.1% were in the allowed area, and

Protein	Peptide	Core peptide	Antigenicity	Allele
S	⁵¹¹ VVLSFELLHAPATVC ⁵²⁵	FELLHAPAT	0.8618	HLA-DRB1*01:01
S	¹⁶⁶ CTFEYVSQPFLMDLE ¹⁸⁰	EYVSQPFLM	0.57	HLA-DPA1*03:01/DPB1*04:02; HLA-DPA1*02:01/DPB1*01:01
S	⁷⁵⁰ SNLLQYGSFCTQLN ⁷⁶⁴	LLQYGSFCT	0.8305	HLA-DRB1*15:01
S	¹⁶⁸ FEYVSQPFLMDLEGK ¹⁸²	EYVSQPFLM	0.8278	HLA-DPA1*03:01/DPB1*04:02
S	⁷⁵¹ NLLQYGSFCTQLNR ⁷⁶⁵	LLQYGSFCT	0.8668	HLA-DRB1*15:01
S	¹⁴² GVYYHKNNKSWMESE ¹⁵⁶	YHKNNKSWM	0.4684	HLA-DRB3*02:02
S	¹⁴¹ LGYYHKNNKSWMES ¹⁵⁵	YHKNNKSWM	0.4937	HLA-DRB3*02:02
S	¹²¹⁰ IKWPWYIWLGFIAGL ¹²²⁴	YIWLGFIAAG	0.9153	HLA-DPA1*01:03/DPB1*02:01
S	¹⁴⁰ FLGVYYHKNNKSWME ¹⁵⁴	YHKNNKSWM	0.4793	HLA-DRB3*02:02
S	³⁴⁶ RFASVYAWNKRKRISN ³⁶⁰	FASVYAWNR	0.4243	HLA-DRB5*01:01
S	⁵⁵ FLPFFSNVTWVFHAIH ⁶⁹	FSNVTWFHA	0.4883	HLA-DPA1*02:01/DPB1*01:01
S	¹⁶⁶ CTFEYVSQPFLMDLE ¹⁸⁰	YVSQPFLMD	0.57	HLA-DPA1*01:03/DPB1*04:01; HLA-DPA1*01:03/DPB1*02:01
N	⁸³ QIGYYRRATRRIRGG ⁹⁷	YRRATRRIR	0.4614	HLA-DRB1*11:01; HLA-DRB5*01:01
N	³⁰³ QIAQFAPSASAFFGM ³¹⁷	FAPSASAFF	0.4032	HLA-DRB1*09:01

Table 2. The final set of helper T lymphocyte (HTL) epitopes selected for multi-epitope vaccine construction.

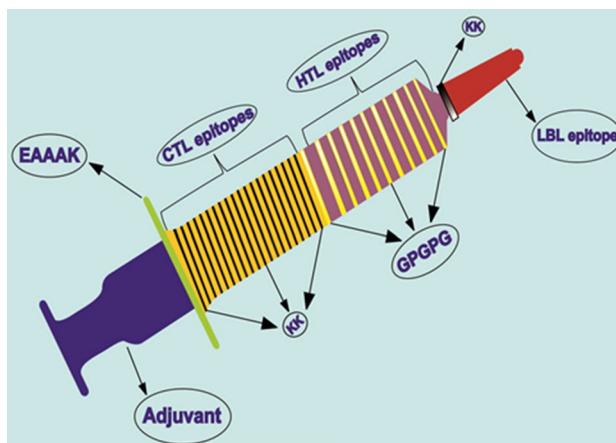


Figure 2. Schematic profile of the multi-epitope subunit vaccine construct of length 694 residues. An adjuvant was added at the N-terminal tail of the vaccine using EAAAK linker, followed by 34 CTL, 12 HTL and 1 LBL epitopes fused by KK and GPGPG linkers.

Features	Assessment
Number of amino acids	694
Molecular weight	78,351.68 Dalton
Theoretical pI	10.37
No. of negatively charged residues (Asp + Glu)	25
No. of positively charged residues (Arg + Lys)	137
Extinction coefficient (at 280 nm in H ₂ O)	120,305 M ⁻¹ cm ⁻¹
Instability index	36.73
Aliphatic index	81.44
Grand average of hydropathicity (GRAVY)	- 0.261
Antigenicity	0.7342 (VaxiJen v.2.0)
Allergenicity	Probable non-allergen (AllergenFP v.1.0) Probable non-allergen (AllerTOP v.2.0)
Solubility	0.914870 (SOLpro)

Table 3. Antigenic, allergenic and physiochemical assessments of the primary sequence of final vaccine protein.

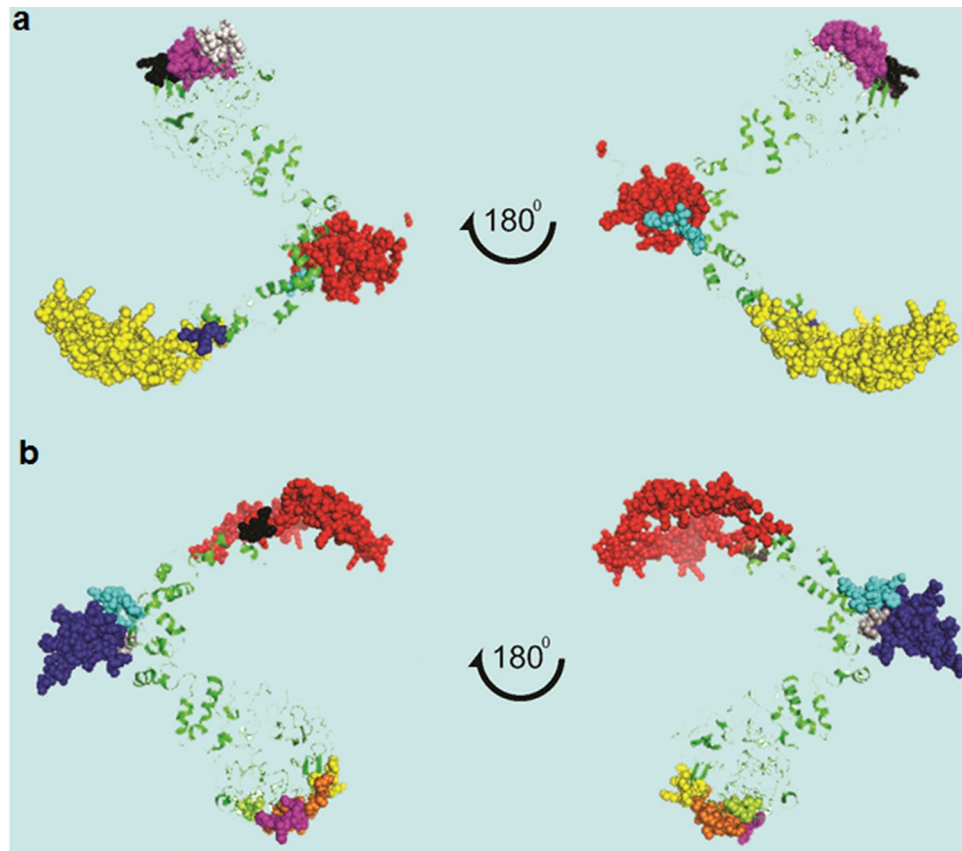


Figure 3. B lymphocyte epitopes present in the designed multi-epitope vaccine. (a) Conformational B cell epitopes shown by spheres, (b) Linear B cell epitopes shown by spheres.

1.6% were in the disallowed area of Ramachandran plot, reflecting a high structural quality of the constructed vaccine model (Supplementary Fig. S1 online).

Conformational and linear B-cell epitopes in the vaccine polypeptide structure. The total of 7 conformational epitopes with scores of 0.702 to 0.882, and 9 linear epitopes with scores of 0.7 to 0.878 were selected as the final B-cell epitopes (Fig. 3, Supplementary Table S7 online). PI value (the score given by ElliPro) of 0.882 shows that 88.2% residues are locating in the predicted ellipsoid area of the epitope and this epitope features the highest solvent accessibility.

Binding of the vaccine structure to TLR-3. The best docked complex of the immune receptor TLR-3 with the vaccine model was identified from among various server outputs by comparing the binding free energy. Analysis of the residues contributing at the protein–protein interface showed that the C-terminal domain of the polypeptide is involved in the interaction, where residues from vaccine polypeptide form polar or non-polar contacts with three domains on the TLR-3 structure (Supplementary Fig. S2 online). The docked complex was further applied for running molecular dynamic (MD) simulation investigations.

MD relaxation and analysis of the receptor-vaccine complex. The MD simulations of the docked multi-epitope-based subunit vaccine with TLR-3 as the receptor were carried out to achieve information about the conformational changes of TLR-3-vaccine polypeptide complex. Such studies were essential for several vital facets: (1) to comprehend whether the designed vaccine was stable at the bound pocket; (2) to verify that the induced conformational mobility of both TLR-3 receptor and the multi-epitope vaccine structure did not exert undesired impact on the conformation of the docked proteins; and (3) to corroborate that the epitopes of the multi-epitope vaccine were potential for efficiently being recognized by the human immune system, causing strong immune response.

Three statistical factors were assessed based on 24,000 ps of simulation trajectory (Fig. 4). The root mean squared deviation (RMSD) values of TLR-3 and multi-epitope vaccine in the complex reflected the high conformational stability of the docked molecules. An average RMSD of 0.29 nm with maximum of 0.44 nm realized at 14,000 ps was noted for the TLR-3 molecule (Fig. 4A). The RMSD value for the multi-epitope vaccine model (Fig. 4B) showed that it mostly remained stable during simulation time, with a plateau at about 10,500 ps. The

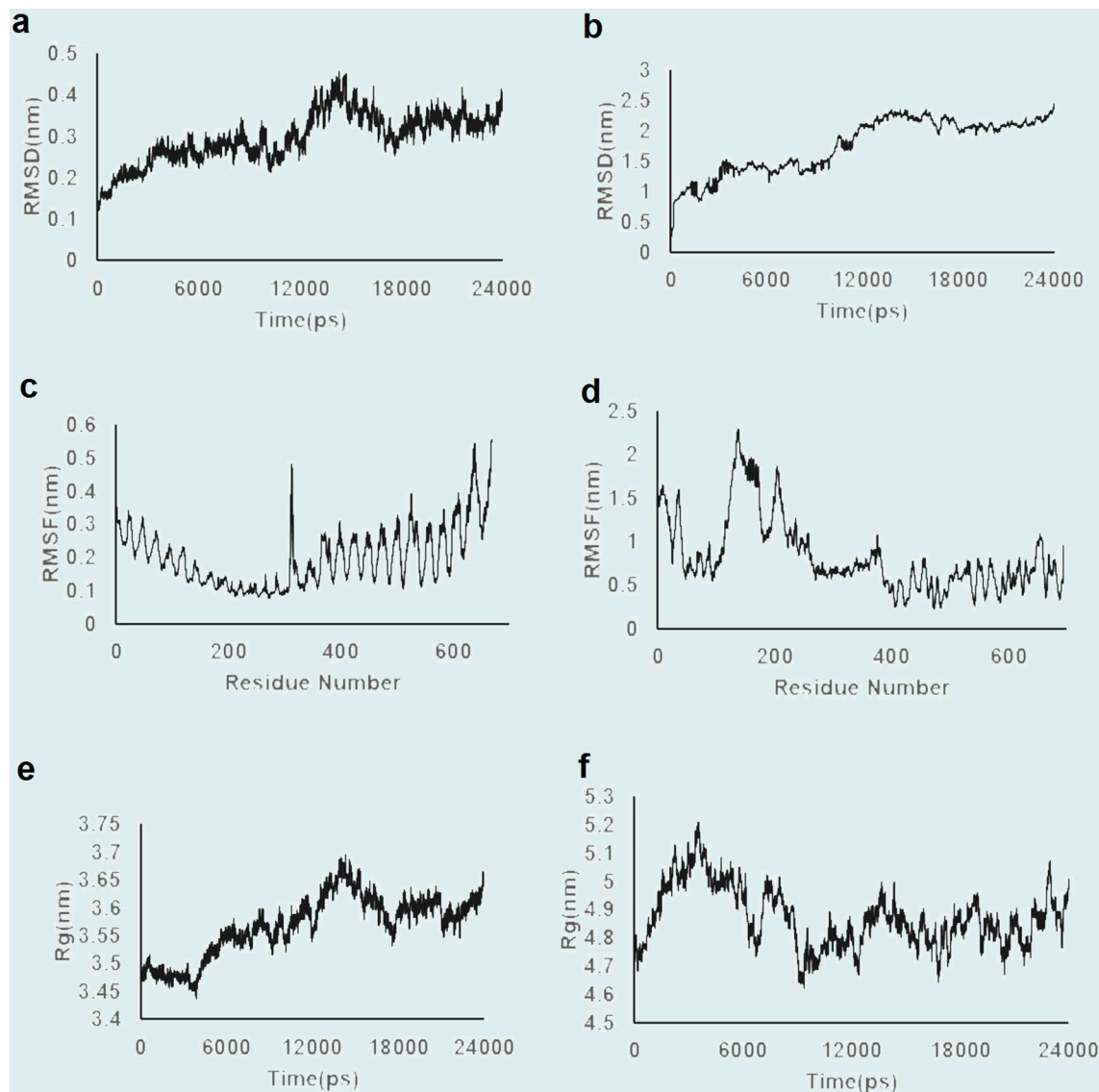


Figure 4. Illustration of the molecular dynamic equilibration for simulation outputs. **(a)** Root mean squared deviations (RMSDs) of Ca for Toll-like receptor-3 (TLR-3), and **(b)** for the multi-epitope vaccine polypeptide; **(c)** Root mean squared fluctuations (RMSFs) of Ca atoms for TLR-3, and **(d)** for the multi-epitope vaccine polypeptide; **(e)** Radius of gyration for TLR-3, and **(f)** for the multi-epitope vaccine polypeptide.

observed trends can be attributed to the moving multi-epitope vaccine at TLR-3 binding pocket in an effort to obtain a suitable and stable docked conformation.

Region-wise structural fluctuations of the TLR-3-vaccine polypeptide complex were studied by calculating the root mean squared fluctuation (RMSF) parameter (Fig. 4C,D). The mean RMSF calculated for TLR-3 and the multi-epitope vaccine were 0.2 nm and 0.83 nm, respectively, which were overall in favour of protein residues' local stability. Along the vaccine sequence, residues located at the loop regions (such as I137, T138, W204, N376 and R654) had high fluctuation (Fig. 4D). The flexibility of the loop regions is essential for proper holding of the vaccine at the binding pocket.

The compactness of the complex structure was estimated by calculating the radius of gyration (Rg) of TLR-3 and multi-epitope peptide molecules. The related graphs (Fig. 4E,F) showed that during the simulation time, TLR-3 and multi-epitope vaccine molecules had mean Rg values of about 3.56 nm and 4.87 nm, respectively. Rg fluctuations is an indication of movements and conformational changes in flexible regions of the multi-epitope vaccine peptide in the TLR-3 binding pocket. This dynamics seems essential to suitably identify the vaccine and incorporating it in the binding pocket.

Free energy of the binding between vaccine polypeptide and TLR-3. To figure out the strength of the contact between multi-epitope vaccine and TLR-3 structures, the binding free energy between the two molecules was calculated using MMPBSA approach. According to Table 4, the nonpolar element (− 136.92 kcal/

	Energy values (kcal/mol)
ΔE_{ele}^a	-241.5 ± 14.81
ΔE_{vdW}^b	-118.81 ± 9.0
ΔG_{GB}^c	240.3 ± 16.71
ΔG_{SA}^d	-18.11 ± 1.1
$\Delta E_{\text{polar}}^e$	-1.2 ± 0.4
$\Delta E_{\text{non-polar}}^f$	-136.92 ± 4.07
$\Delta G_{\text{binding}}$	-138.11 ± 9.46

Table 4. Binding free energy calculation for the multi-epitope vaccine candidate-TLR-3 complex.

^aElectrostatic contribution. ^bvan der Waals contribution. ^cPolar contribution of the solvation effect. ^dNon-polar contribution of solvation effect. ^e $\Delta E_{\text{polar}} = \Delta E_{\text{ele}} + \Delta G_{\text{GB}}$. ^f $\Delta E_{\text{non-polar}} = \Delta E_{\text{vdW}} + \Delta G_{\text{SA}}$.

mol) was an important energy term in the binding free energy of the complex. Our findings clarified that the favourable electrostatic energy ($E_{\text{ele}} = -241.5$ kcal/mol) was covered up by the huge polar energy component ($\Delta G_{\text{GB}} = 240.3$ kcal/mol) in the binding process of the multi-epitope vaccine polypeptide. Therefore, the non-polar energy was known as the main driving force in the vaccine binding to TLR-3, and this hydrophobic contribution leads to a thermodynamically favourable interaction ($\Delta G_{\text{binding}} = -138.11$ kcal/mol). To further clarify the binding mode, the calculated energy was broken down into residue-residue pairs through the binding free energy decomposition analysis. According to the data, there were several residues of vaccine polypeptide with less than -1.5 kcal/mol free energy of contribution in the binding mechanism. The binding pose of vaccine with the key residues are illustrated in Fig. S3.

Discussion

Classic live or attenuated vaccines have a long history of success, however their deployment is associated with issues of biosafety such as autoimmune or strong allergic responses, plus difficulties with their synthesis and manufacture. These drawbacks might be addressed by developing fully synthetic peptide-based vaccines¹⁶. Research on COVID-19 prevention through this approach was begun immediately after release of relevant basic data such as its full genomic sequence. Specific epitope regions in SARS-CoV-2 with high homology to SARS-CoV, the best-characterized coronavirus in terms of epitope responses, were identified^{17–19}. Antigenic properties of spike glycoprotein were more focused by theory and experimental researchers^{20–22}. Applying this basic knowledge, a number of peptide-based designs have been presented to date, which encompass those integrating epitopes from a single or few viral protein(s)^{23–27} to those considering the whole proteome of the virus²⁸. The present research contributes to current understanding in this field by offering an alternative approach for developing a potential COVID-19 vaccine, where the entire set of structural proteins were searched for most efficacious epitopic segments.

One rationale for the choice of the whole set of structural proteins of SARS-CoV-2 for epitope identification in this study, was the ample evidence on immunity-related advantages of each of the S, E, M and N proteins of SARS-CoV or SARS-CoV-2^{17,29–32}. In addition, the choice of structural proteins as the source of epitopes implies a highly-conserved epitope set, as structural proteins are subject to less sequence or structural evolution¹⁸. We found that our proposed epitopes demonstrate overlap with similar studies which have extracted the epitopes from individual structural proteins from different sequence database entries; the similarity of findings and the conservation analyses by those research groups may also indicate the high epitope conservancy among geographical strains^{26,27}. Aside from these advantages, our selection of structural epitopes aimed to cope with potential vaccine side effects, particularly the antibody-dependent enhancement of infectivity (ADEI). ADEI is a great challenge of subunit vaccines, referring to the reduced specificity in response elicitation because of the numerous conformations adopted by a peptide vaccine¹⁰. Strategies to mitigate this concern include: 1- immunofocusing by considering several most antigenic epitopes³³; 2- inclusion of structural proteins: structurally flexible coronaviral epitopes may be of limited value for in vivo immunotargeting and require to be replaced by conserved epitopes with low structural plasticity⁶. Structural proteins are characterized by highly conserved sequences, thus the linear epitopes remain highly stable in these proteins and the conformational epitopes preserve their structural patterns during various steps of the virus cycle. Our choice to include the full set of SARS-CoV-2 structural proteins has thus combined the advantages of augmented immunogenicity and geographical conservancy with precautions of both strategies proposed for reducing ADEI.

Integrating multiple immunodominant sites of viral pathogens in the vaccine structure helps augment the antigenic effect. Reports revealed that multiple antigenic peptides induce stronger B and T cells immune responses than un-conjugated peptide epitopes^{34–36}. Thus, consecutive sequences of HTL and CTL epitopic peptides were fused to each other using accepted linkers. GPGPG has been introduced as a spacer between adjacent epitopes, because of several positive contributions to the immune-modulatory function of polypeptide vaccines, including reduction of junctional epitopes, increased proteasome processing, and augmented immunogenicity³⁷. The KK linker is the target sequence of cathepsin B, a lysosomal protease of high importance for antigen processing. Connecting two peptide antigens together using the lysine linker helps avoid induction of antibodies to the amino acid sequence that is generated by joining of two peptides and most antibodies would be reactive to each peptide³⁸.

Epitope-based peptide vaccines induce relatively weak immune response, when used alone³⁹. The immunoreactivity could be improved by appending proper adjuvants. The vaccine construct in this study was prepared by fusing the epitopic peptides to β -defensin as an immunogenic adjuvant. β -Defensin has previously been reported as a potent adjuvant when conjugated with MERS-CoV antigens. Vaccines containing defensins as adjuvants have been shown, both in vivo and in vitro, to activate the primary innate antiviral immune response and mediate other immunomodulatory activities against a number of viruses, including coronaviruses^{40,41}. Use of appropriate adjuvants has also been shown to help induce durable IFN- γ responses⁶.

In this study, the binding of the built vaccine model with the immune receptor TLR-3 was assessed by performing molecular docking, MD simulation and free energy calculations. Triggering TLR-3 may help induce the TLR signalling networks which activate the immune pathways evolved specifically against viral pathogens. In addition, this TLR choice was a consensus from vaccine design research works on viral/retroviral pathogens, specifically coronavirus species^{12,14,25,26,42,43}. While we note that other Toll-like immune receptors such as TLR-2, -4, -5, -7 and -8 may be stimulated by the COVID-19 virus, they were not considered here as their exact roles are not established, with some possibly functioning even to the advantage of the coronavirus^{44,45}.

Conclusion

In the ongoing urgent situation brought about by SARS-CoV-2, it is hard to fast counteract the circulating disease through preventative or therapeutic measures. The multi-epitope-based subunit vaccine design obtained through an immunoinformatic pipeline in this study could be promising, as it incorporates a priori bioinformatics predictions and up-to-date immunological knowledge. Despite the current highly active research community as well as the efforts of the industry section in the road to find an optimal immunogenic formulation, the extreme diversity in available design choices, with no a priori image of their effectiveness, appears as a fundamental obstacle in this route. To overcome this issue, a rapid mass design and screening strategy incorporating the same immunoinformatic approach is recommended to be developed by the research community, with the hope to find one formulation with the highest immunogenicity and biosafety.

Methodology

SARS-CoV-2 structural protein sequences. The amino acid sequences of SARS-CoV-2 structural proteins, including the spike glycoprotein (S), envelope protein (E), membrane protein (M), and nucleocapsid phosphoprotein (N), were retrieved using the NCBI reference genome of the virus (accession number NC_045512.2).

Identifying cytotoxic T lymphocyte (CTL) epitopes. Predicting peptides that are capable of inducing CTL responses is a crucial step in the design of epitope-based vaccine. The MHC-I Binding tool of Immune Epitope Database and Analysis Resource (IEDB; <http://tools.iedb.org/mhci>) was used to predict the CD8⁺ T-cell epitopes borne in S, E, M and N proteins^{12,46,47}. In this step, the ANN 4.0 method was set as the prediction method. The human was selected as the source species. The maximum IC₅₀ value was set to 50 nM, and percentile rank < 1 was considered as threshold since lower score indicates high affinity^{48,49}.

Identifying helper T lymphocyte (HTL) epitopes. IEDB (<http://www.iedb.org>) was used to predict MHC-II binding of 15-mer epitopes for viral structural proteins against human HLAs such as HLA-DRB1*15:01, HLA-DRB4*01:01, HLA-DRB3*01:01, HLA-DRB5*01:01, HLA-DRB1*03:01, HLA-DRB3*02:02 and HLA-DRB1*07:01 (Supplementary Table S8 online), using NN-align 2.3 method^{13,14,50,51}. The maximum IC₅₀ value of 50 nM, 500 nM and 5000 nM indicate high, intermediate and low affinity of epitopes, respectively⁵². The 15-mer epitopes with IC₅₀ values < 50 nM were considered to be included in the vaccine polypeptide^{52–54}.

B-cell epitopes prediction. The ABCpreds server was used to predict 16-mer linear B-lymphocyte (LBL) epitopes, using a threshold of 0.51⁵⁵. In addition, the ElliPro tool of IEDB was utilized to predict conformational and linear B-cell epitopes of the vaccine polypeptide.

Assessment of identified epitopes for antigenicity, allergenicity, and toxicity. The antigenic potential of each of the T and B cells epitopes was predicted by Vaxijen v2.0 applying a threshold of 0.4⁵⁶. The predicted T and B cells epitopes were then further evaluated in terms of toxicity and allergenicity, with ToxinPred and AllergenFP v1.0 servers, respectively^{57,58}. Then, the ability of each of the HTL and B cell epitopes (CD4⁺) to induce the secretion of cytokines, such as interferon-gamma (IFN- γ), interleukin-4 (IL-4) and interleukin-10 (IL-10), to overcome the inflammatory response and prevent tissue damage was predicted by IFNepitope, IL4pred and IL10pred server tools, respectively^{59–61}. The IL4pred and IL10pred operations were carried out based on SVM method and hybrid method, respectively, with other parameters kept as default^{12,13}.

Designing the multi-epitope vaccine polypeptide construct. To develop a multi-epitope vaccine construct, we selected those predicted epitopes which demonstrated high antigenic potential, were not identified as allergic or toxic, and had good solubility when highly expressed. To construct the vaccine polypeptide, the selected CTL, HTL and LBL epitopes were fused together using appropriate linkers^{62,63}. In addition, a 45-amino acid sequence prepared from β -defensin-2 protein was added to the N terminus of the vaccine sequence using the EAAAK linker, to increase the immunogenic capacity of the multi-epitope vaccine^{12,52,63}.

Immunogenic, allergenic and physicochemical evaluation of vaccine construct. The antigenicity of the multi-epitope vaccine polypeptide was predicted utilizing the VaxiJen v2.0 tool, with the threshold value

of 0.4⁵⁶. The allergenicity of the vaccine was analysed using AllerTOP v.2.0 and AllergenFP v.1.0 servers^{57,64}. The ProtParam server was employed to evaluate the physical chemistry properties of the vaccine construct, such as amino acid composition, molecular weight, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY), aliphatic and instability index, and in vitro and in vivo half-life⁶⁵.

Vaccine polypeptide structure modelling, refinement and validation. The SOPMA server was applied for analysing the secondary structural properties of the multi-epitope vaccine polypeptide⁶⁶. The GalaxyWEB server was employed for modelling and refinement of the 3D structure model⁶⁷. The server relaxes the model structure using repacking and molecular dynamics (MD) simulation. Next, RAMPAGE server and ProSA-web tools were used to validate the refined 3D model^{68,69}. All of these tools give us the overall quality of the 3D structure of the peptide vaccine.

Molecular docking of the vaccine polypeptide to TLR-3. The binding of antigenic agents with the target immune cell protein is crucial for the creation of a suitable immune system response. For analysing the binding pattern of the multi-epitope vaccine polypeptide with TLR-3 (PDB ID: 2A0Z)⁷⁰, molecular docking analysis was performed by Hdock, Zdock, Cluspro and Hawkdock^{71–74}. Among the molecular species docked by each of these applications, the best outputs were extracted, followed by the four docked molecules uploaded into the Hawkdock program, and the free energy of binding of each complex calculated. Based on this screening, it was determined that the model selected from the Cluspro output had the best binding free energy. Therefore, this molecular species was chosen as the primary structure for initiating the MD simulation.

MD simulation of the vaccine-receptor complex. To determine the structural stability and to study the molecular details of the interactions between TLR-3 and the multi-epitope vaccine polypeptide in the docked conformation, MD simulation was performed. Briefly, the system including vaccine polypeptide-TLR-3 was simulated by the Gromacs-2020 package applying OPLS-AA force field⁷⁵. The complex system was solvated using TIP3P water model. Then, the genion module was utilized to neutralize the whole system. Next, the conjugate gradient algorithm was applied to minimize the energy of the system. In an NVT ensemble, the temperature of the system gradually increased from 0 to 310 K during 400 ps. Subsequently, in an NPT ensemble, 500 ps simulation was carried out at the pressure of 1 atm and the temperature of 310 K. Production simulation for 24,000 ps was then implemented. The particle-mesh Ewald (PME) and the LINCS algorithms were applied to assess all electrostatic connections and to restrain all bond lengths in the protein, respectively. Moreover, periodic boundary condition was utilized during the simulation. The final coordinates obtained for the complex system were analysed with classic MD analyses, plus the MMPBSA method for calculating the free energy of intermolecular interactions⁷⁶. The results were visualized by Pymol (Schrodinger L.L.C.).

Data availability

No datasets were generated or analysed during the current study. Data such as epitope sequences, which were analysed during this study, are included in this published article and its Supplementary Information file.

Received: 24 April 2020; Accepted: 12 November 2020

Published online: 30 November 2020

References

- Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192. <https://doi.org/10.1038/s41579-018-0118-9> (2019).
- World Health Organization. *Coronavirus disease 2019 (COVID-19), Situation reports, Weekly Epidemiological Update, 21 September 2020*. www.who.int/docs/default-source/coronaviruse/situation-reports/20200921-weekly-epi-update-6.pdf?sfvrsn=d9cf9496_6 (2020).
- Tseng, C. T. *et al.* Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS ONE* **7**, e35421. <https://doi.org/10.1371/journal.pone.0035421> (2012).
- Muthumani, K. *et al.* A synthetic consensus anti-spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci. Transl. Med.* **7**, 301–132. <https://doi.org/10.1126/scitranslmed.aac7462> (2015).
- Wang, L. *et al.* Evaluation of candidate vaccine approaches for MERS-CoV. *Nat. Commun.* **6**, 7712. <https://doi.org/10.1038/ncomms8712> (2015).
- Enjuanes, L. *et al.* Molecular basis of coronavirus virulence and vaccine development. *Adv. Virus Res.* **96**, 245–286. <https://doi.org/10.1016/bs.aivir.2016.08.003> (2016).
- Wang, J. *et al.* The adjuvanticity of an *O. volvulus*-derived rOv-ASP-1 protein in mice using sequential vaccinations and in non-human primates. *PLoS ONE* **7**, e37019. <https://doi.org/10.1371/journal.pone.0037019> (2012).
- Zhang, N. *et al.* Identification of an ideal adjuvant for receptor-binding domain-based subunit vaccines against Middle East respiratory syndrome coronavirus. *Cell. Mol. Immunol.* **13**, 180–190. <https://doi.org/10.1038/cmi.2015.03> (2016).
- Papaneri, A. B. *et al.* Middle East respiratory syndrome: obstacles and prospects for vaccine development. *Expert Rev. Vaccines* **14**, 949–962. <https://doi.org/10.1586/14760584.2015.1036033> (2015).
- Graham, B. S., Gilman, M. S. A. & McLellan, J. S. Structure-based vaccine antigen design. *Annu. Rev. Med.* **70**, 91–104. <https://doi.org/10.1146/annurev-med-121217-094234> (2019).
- Khan, M. K. *et al.* In silico predicted mycobacterial epitope elicits in vitro T-cell responses. *Mol. Immunol.* **61**, 16–22. <https://doi.org/10.1016/j.molimm.2014.04.009> (2014).
- Abdulla, F., Adhikari, U. K. & Uddin, M. K. Exploring T & B-cell epitopes and designing multi-epitope subunit vaccine targeting integration step of HIV-1 lifecycle using immunoinformatics approach. *Microb Pathog* **137**, 103791. <https://doi.org/10.1016/j.micpath.2019.103791> (2019).
- Khatoun, N., Pandey, R. K. & Prajapati, V. K. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci. Rep.* **7**, 8285. <https://doi.org/10.1038/s41598-017-08842-w> (2017).

14. Narula, A., Pandey, R. K., Khatoon, N., Mishra, A. & Prajapati, V. K. Excavating chikungunya genome to design B and T cell multi-epitope subunit vaccine using comprehensive immunoinformatics approach to control chikungunya infection. *Infect. Genet. Evol.* **61**, 4–15. <https://doi.org/10.1016/j.meegid.2018.03.007> (2018).
15. Berzofsky, J. A. & Berkower, I. in *Fundamental Immunology* Vol. 631–684 (ed W.E. Paul) (Raven Press, 1993).
16. Skwarczynski, M. & Toth, I. Peptide-based synthetic vaccines. *Chem. Sci.* **7**, 842–854 (2016).
17. Grifoni, A. *et al.* A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671–680. <https://doi.org/10.1016/j.chom.2020.03.002> (2020).
18. Kumar, S., Maurya, V. K., Prasad, A. K., Bhatt, M. L. B. & Saxena, S. K. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *Virusdisease* **31**, 13–21. <https://doi.org/10.1007/s13337-020-00571-5> (2020).
19. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. Preliminary identification of potential vaccine targets for the covid-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* <https://doi.org/10.3390/v12030254> (2020).
20. Baruah, V. & Bose, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J. Med. Virol.* **92**, 495–500. <https://doi.org/10.1002/jmv.25698> (2020).
21. Lucchese, G. Epitopes for a 2019-nCoV vaccine. *Cell. Mol. Immunol.* <https://doi.org/10.1038/s41423-020-0377-z> (2020).
22. Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* <https://doi.org/10.1016/j.cell.2020.02.058> (2020).
23. Bhattacharya, M. *et al.* Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J. Med. Virol.* **5**, 2. <https://doi.org/10.1002/jmv.25736> (2020).
24. Kar, T. *et al.* A candidate multi-epitope vaccine against SARS-CoV-2. *Sci. Rep.* **10**, 10895. <https://doi.org/10.1038/s41598-020-67749-1> (2020).
25. Rahman, M. S. *et al.* 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. *PeerJ* **8**, e9572. <https://doi.org/10.7717/peerj.9572> (2020).
26. Kalita, P., Padhi, A. K., Zhang, K. Y. J. & Tripathi, T. Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. *Microb. Pathog.* **145**, 104236. <https://doi.org/10.1016/j.micpath.2020.104236> (2020).
27. Lizbeth, R. G., Jazmin, G. M., Jose, C. B. & Marlet, M. A. Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine. *J. Biomol. Struct. Dyn.* <https://doi.org/10.1080/07391102.2020.1780944> (2020).
28. Crooke, S. N., Ovsyannikova, I. G., Kennedy, R. B. & Poland, G. A. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Sci. Rep.* **10**, 14179. <https://doi.org/10.1038/s41598-020-70864-8> (2020).
29. Liu, J. *et al.* The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J. Infect. Dis.* **202**, 1171–1180. <https://doi.org/10.1086/656315> (2010).
30. Ng, O. W. *et al.* Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **34**, 2008–2014. <https://doi.org/10.1016/j.vaccine.2016.02.063> (2016).
31. Peng, H. *et al.* Human memory T cell responses to SARS-CoV E protein. *Microb. Infect.* **8**, 2424–2431. <https://doi.org/10.1016/j.micinf.2006.05.008> (2006).
32. Surjit, M. & Lal, S. K. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. *Infect. Genet. Evol.* **8**, 397–405. <https://doi.org/10.1016/j.meegid.2007.07.004> (2008).
33. Ma, C. *et al.* Searching for an ideal vaccine candidate among different MERS coronavirus receptor-binding fragments: the importance of immunofocusing in subunit vaccine design. *Vaccine* **32**, 6170–6176. <https://doi.org/10.1016/j.vaccine.2014.08.086> (2014).
34. Tam, J. P. Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. *Proc. Natl. Acad. Sci.* **85**, 5409–5413 (1988).
35. Li, W., Joshi, M. D., Singhanian, S., Ramsey, K. H. & Murthy, A. K. Peptide vaccine: progress and challenges. *Vaccines* **2**, 515–536 (2014).
36. Reche, P. A., Fernandez-Caldas, E., Flower, D. R., Fridkis-Hareli, M. & Hoshino, Y. Peptide-based immunotherapeutics and vaccines. *J. Immunol. Res.* **2014**, 1–2 (2014).
37. Livingston, B. *et al.* A rational strategy to design multi-epitope immunogens based on multiple Th lymphocyte epitopes. *J. Immunol.* **168**, 5499–5506. <https://doi.org/10.4049/jimmunol.168.11.5499> (2002).
38. Yano, A. *et al.* An ingenious design for peptide vaccines. *Vaccine* **23**, 2322–2326. <https://doi.org/10.1016/j.vaccine.2005.01.031> (2005).
39. Tahir Ul Qamar, M. *et al.* Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J. Transl. Med.* **17**, 362. <https://doi.org/10.1186/s12967-019-2116-8> (2019).
40. Kim, J., Yang, Y. L., Jang, S. H. & Jang, Y. S. Human beta-defensin 2 plays a regulatory role in innate antiviral immunity and is capable of potentiating the induction of antigen-specific immunity. *Virol. J.* **15**, 124. <https://doi.org/10.1186/s12985-018-1035-2> (2018).
41. Park, M. S., Kim, J. I., Lee, I., Park, S. & Bae, J. Y. Towards the application of human defensins as antivirals. *Biomol. Ther.* **26**, 242–254. <https://doi.org/10.4062/biomolther.2017.172> (2018).
42. Srivastava, S., Kamthania, M., Singh, S., Saxena, A. K. & Sharma, N. Structural basis of development of multi-epitope vaccine against Middle East respiratory syndrome using in silico approach. *Infect. Drug Resist.* **11**, 2377–2391. <https://doi.org/10.2147/IDR.S175114> (2018).
43. Ikram, A. *et al.* Exploring NS3/4A, NS5A and NS5B proteins to design conserved subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches. *Sci. Rep.* **8**, 16107. <https://doi.org/10.1038/s41598-018-34254-5> (2018).
44. Olejnik, J., Hume, A. J. & Muhlberger, E. Toll-like receptor 4 in acute viral infection: too much of a good thing. *PLoS Pathog.* **14**, e1007390. <https://doi.org/10.1371/journal.ppat.1007390> (2018).
45. Imai, Y. *et al.* Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. *Cell* **133**, 235–249. <https://doi.org/10.1016/j.cell.2008.02.043> (2008).
46. Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* **8**, 424. <https://doi.org/10.1186/1471-2105-8-424> (2007).
47. Yadav, G., Rao, R., Raj, U. & Varadwaj, P. K. Computational modeling and analysis of prominent T-cell epitopes for assisting in designing vaccine of ZIKA virus. *J. Appl. Pharm. Sci.* **7**, 116–122. <https://doi.org/10.7324/japs.2017.70816> (2017).
48. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517. <https://doi.org/10.1093/bioinformatics/btv639> (2016).
49. Nielsen, M. & Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinform.* **10**, 296. <https://doi.org/10.1186/1471-2105-10-296> (2009).
50. Greenbaum, J. *et al.* Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different super-types and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* **63**, 325–335. <https://doi.org/10.1007/s00251-011-0513-0> (2011).
51. Weiskopf, D. *et al.* Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc. Natl. Acad. Sci. USA.* **110**, E2046–2053. <https://doi.org/10.1073/pnas.1305227110> (2013).
52. Duvvuri, V. R. *et al.* Preexisting CD4+ T-cell immunity in human population to avian influenza H7N9 virus: whole proteome-wide immunoinformatics analyses. *PLoS ONE* **9**, e91273. <https://doi.org/10.1371/journal.pone.0091273> (2014).

53. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**, 4409. <https://doi.org/10.1038/s41598-019-40833-x> (2019).
54. Zhou, W. Y. *et al.* Therapeutic efficacy of a multi-epitope vaccine against *Helicobacter pylori* infection in BALB/c mice model. *Vaccine* **27**, 5013–5019. <https://doi.org/10.1016/j.vaccine.2009.05.009> (2009).
55. Saha, S. & Raghava, G. P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**, 40–48. <https://doi.org/10.1002/prot.21078> (2006).
56. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4. <https://doi.org/10.1186/1471-2105-8-4> (2007).
57. Dimitrov, I., Naneva, L., Doytchinova, I. & Bangov, I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**, 846–851. <https://doi.org/10.1093/bioinformatics/btt619> (2014).
58. Gupta, S. *et al.* In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* **8**, e73957. <https://doi.org/10.1371/journal.pone.0073957> (2013).
59. Dhanda, S. K., Gupta, S., Vir, P. & Raghava, G. P. Prediction of IL4 inducing peptides. *Clin. Dev. Immunol.* **2013**, 263952. <https://doi.org/10.1155/2013/263952> (2013).
60. Dhanda, S. K., Vir, P. & Raghava, G. P. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* **8**, 30. <https://doi.org/10.1186/1745-6150-8-30> (2013).
61. Nagpal, G. *et al.* Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci. Rep.* **7**, 42851. <https://doi.org/10.1038/srep42851> (2017).
62. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369. <https://doi.org/10.1016/j.addr.2012.09.039> (2013).
63. Mohan, T., Sharma, C., Bhat, A. A. & Rao, D. N. Modulation of HIV peptide antigen specific cellular immune response by synthetic alpha- and beta-defensin peptides. *Vaccine* **31**, 1707–1716. <https://doi.org/10.1016/j.vaccine.2013.01.041> (2013).
64. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP vol 2—a server for in silico prediction of allergens. *J. Mol. Model.* **20**, 2278. <https://doi.org/10.1007/s00894-014-2278-5> (2014).
65. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook Springer Protocols Handbooks* (ed J.M. Walker) 571–607 (Humana Press, 2005).
66. Geourjon, C. & Deleage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *CABIOS* **11**, 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681> (1995).
67. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **40**, W294–297. <https://doi.org/10.1093/nar/gks493> (2012).
68. Wang, W. *et al.* Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in *Gossypium raimondii* and *G. arboreum*. *Data Brief* **9**, 345–348. <https://doi.org/10.1016/j.dib.2016.05.025> (2016).
69. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–410. <https://doi.org/10.1093/nar/gkm290> (2007).
70. Bell, J. K. *et al.* The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc. Natl. Acad. Sci. USA.* **102**, 10976–10980. <https://doi.org/10.1073/pnas.0505077102> (2005).
71. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278. <https://doi.org/10.1038/nprot.2016.169> (2017).
72. Pierce, B. G., Hourai, Y. & Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* **6**, e24657. <https://doi.org/10.1371/journal.pone.0024657> (2011).
73. Weng, G. *et al.* HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* **47**, W322–W330. <https://doi.org/10.1093/nar/gkz397> (2019).
74. Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S. Y. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373. <https://doi.org/10.1093/nar/gkx407> (2017).
75. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718. <https://doi.org/10.1002/jcc.20291> (2005).
76. Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**, 1951–1962. <https://doi.org/10.1021/ci500020m> (2014).

Author contributions

E.B. (First author) performed data curation, analysis, validation and visualization; and wrote the original draft. B.S. contributed in resource and funding acquisition, and writing the original draft. A.N. contributed in project administration, resources and funding acquisition, and validation. E.B. (Last author) conceptualized the work, defined methodology, and contributed in project administration, supervision, visualization, completing the draft. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-77547-4>.

Correspondence and requests for materials should be addressed to A.N. or E.B.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020