



A reverse vaccinology and immunoinformatics approach for designing a multiepitope vaccine against SARS-CoV-2

Ehsan Jahangirian¹ · Ghadir A. Jamal² · MohammadReza Nouroozi³ · Alemeh Mohammadpour¹

Received: 5 August 2021 / Accepted: 5 September 2021 / Published online: 20 September 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Since 2019, the world was involved with SARS-CoV-2 and consequently, with the announcement by the World Health Organization that COVID-19 was a pandemic, scientific were an effort to obtain the best approach to combat this global dilemma. The best way to prevent the pandemic from spreading further is to use a vaccine against COVID-19. Here, we report the design of a recombinant multi-epitope vaccine against the four proteins spike or crown (S), membrane (M), nucleocapsid (N), and envelope (E) of SARS-CoV-2 using immunoinformatics tools. We evaluated the most antigenic epitopes that bind to HLA class 1 subtypes, along with HLA class 2, as well as B cell epitopes. Beta-defensin 3 and PADRE sequence were used as adjuvants in the structure of the vaccine. KK, GPGPG, and AAY linkers were used to fuse the selected epitopes. The nucleotide sequence was cloned into pET26b(+) vector using restriction enzymes *XhoI* and *NdeI*, and HisTag sequence was considered in the C-terminal of the construct. The results showed that the proposed candidate vaccine is a 70.87 kDa protein with high antigenicity and immunogenicity as well as non-allergenic and non-toxic. A total of 95% of the selected epitopes have conservancy with similar sequences. Molecular docking showed a strong binding between the vaccine structure and toll-like receptor (TLR) 7/8. The docking, molecular dynamics, and MM/PBSA analysis showed that the vaccine established a stable interaction with both structures of TLR7 and TLR8. Simulation of immune stimulation by this vaccine showed that it evokes immune responses related to humoral and cellular immunity.

Keywords COVID-19 · Immunoinformatics · Molecular dynamics simulations · Multi-epitope vaccine

Introduction

For the first time, Wuhan, China, witnessed the outbreak of febrile respiratory disease (it was caused by the Corona 2019 virus), which resulted in patients with acute respiratory distress syndrome (ARDS), accompanied by cough, fever, and shortness of breath, being hospitalized. Besides, viral gene

sequencing related to five pneumonia patients who were hospitalized from December 18 to December 29, 2019, was reported to have an unknown β -CoV strain in 5 hospitalized patients (Choudhary et al. 2020; Hui et al. 2020). Sequencing revealed that there was high sequence similarity with the two bat-coronaviruses and a MERS-CoV (Lu et al. 2020). The virus was identified as SARS-CoV-2 that is caused by COVID-19 disease (Perlman and Netland 2009). Over the past two decades, coronaviruses have caused three epidemics, including COVID-19 in 2019, Severe Acute Respiratory Syndrome (SARS) in 2002, and Middle East Respiratory Syndrome (MERS) in 2013. Four genera of the coronavirus α , β , γ , and δ have been identified. The Human Corona Virus (H-CoV), belong to the genera α and β , is a member of the family Coronaviridae, A viral family with the largest genomic RNA (26–32 kb), among viruses known to date (Anand et al. 2003; Su et al. 2016; Ziebuhr 2005). (Perlman and Netland 2009). Coronaviruses are more commonly associated with respiratory illnesses and colds, but can also cause infections of the central nervous system (CNS) (Bergmann

Ehsan Jahangirian and GhadirA. Jamal equally contributed in this work.

✉ Ghadir A. Jamal
ghadir.jamal@ku.edu.kw

¹ Department of Animal Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

² Faculty of Allied Health Sciences, Kuwait University, Kuwait City, Kuwait

³ Department of Animal Science and Food Technology, Agriculture Science and Natural Resources University Khouzestan, Ahwaz, Iran

et al. 2006; Nicholson et al. 1993). Coronavirus infection in humans is primarily driven by the interaction between the coronavirus's coronary glycoprotein (S-protein) and the host cell receptor-converting enzyme angiotensin-2 (ACE2), which is mainly expressed in alveolar cells of type 2 in the lungs (Hoffmann et al. 2020; Wong et al. 2004). Once the virus enters the cell, the viral RNA genome is released into the cytoplasm. At the same time, as it begins to replicate, the viral genome translates into two forms of nonstructural replication proteins, Replicase 1ab, and structural proteins, spike or crown (S), envelope (E), membrane (M), and nucleocapsid (N) (Bonavia et al. 2003; Malik et al. 2020). This condition causes the coronary protein to bind to host cell receptors, releasing the viral genome into the host cell, thus facilitating virus replication (de Wit et al. 2016). Increased binding affinity between ACE2 and SARS-CoV-2 receptors has been associated with increased virus transmission (Wong et al. 2004). The S protein consists of two subunits of S1 and S2. S1 subunit is related to the receptor-binding domain (RBD). Also, the S2 subunit is known as the main factor for the fusion of the viral and host cell membrane (Tseng et al. 2012). Therefore, glycoprotein S is the most important antigenic candidate for the design of a vaccine against the coronavirus. Antiviral drugs have been used as treatment strategies to reduce the viral load or to prevent the proliferation of SARS-CoV-2 (Richardson et al. 2020; Sheahan et al. 2020; Wang et al. 2020). Besides, recently plasma exchange using COVID-19 sera has shown promising results (Derebail and Falk 2020; Keith et al. 2020). Also, the results of binding of monoclonal antibody (CR3022) to the domain binding of SARS-CoV-2 S protein receptor have shown the potential for its use as a therapeutic candidate for COVID-19 (Tian et al. 2020). However, it is important to develop an appropriate strategy to prevent further spread of the SARS-CoV-2 epidemic, for example, the COVID-19 vaccine. There are several types of vaccine candidates for protection against virulence, including complete attenuated or inactivated live virus, DNA vaccines, mRNA vaccines, and subunit vaccines, each with its own advantages and disadvantages.

Several vaccines have been proposed to prevent COVID-19 to date (Le Thanh et al. 2020). Side effects and biosafety concerns related to inactivated CoV viruses lead to a use limitation in them, despite their potential in neutralizing antibodies producing and reach of immunity protection (Tseng et al. 2012). Besides, DNA vaccines containing plasmids expressing protein S (whole or specific regions of their sequence) as well as their associated DNA primers are effective against MERS-CoV infection (Muthumani et al. 2015; Wang et al. 2015). Simultaneously with the intracellular expression of immunogenic antigens, vaccines containing plasmid DNA enable the activation of both cellular and blood (humoral) immune responses. DNA vaccines have a

proven history of immunogenicity, but plasmids are limited to delivering one or a small number of CoV antigens to the immune system (Enjuanes et al. 2016). Two studies have shown that recombinant S protein-based subunit vaccines have been very successful in immunizing against SARS and MERS (Wang et al. 2012; Zhang et al. 2016). Although DNA vaccines and mRNA vaccines can elicit a centralized antibody response or act more flexibly and effectively to manipulate antigens, their components may not be able to fully demonstrate the full antigenic capacity of the virus. As a result, they have a limited protective or immunopathological effect due to unbalanced immune responses (Zhang et al. 2019). One promising option to address this is multi-unit epitope vaccines, which contain a variety of antigenic elements and provide a wide range of virus antigens (Abdulla et al. 2021; Gupta and Kumar 2020; Krishnan et al. 2021). In addition, multi-epitope vaccines do not have problems related to mRNA stability during translation and have design potential for varied strains target (Skwarczynski and Toth 2016). The multi-epitope vaccines have high safety and cost-effectiveness which contains both B cell and T cell immune response. Because of this advantage, multi-epitope vaccines are capable of producing strong, long-lasting immunity against the target pathogen (Skwarczynski and Toth 2016). Therefore, it is necessary to predict the interaction of immunogenic epitopes with the major histocompatibility complex (MHC). Here we utilized immunoinformatics and computational tools for designing a multi-epitope vaccine candidate against SARS-CoV-2 based on the bests antigenic and immunogenic factors of S, N, M, and E proteins.

Materials and methods

Proteins sequence retrieval

The amino acid sequence of S, M, N, and E proteins were retrieved from the Uniprot server with accession numbers of P0DTC2 (<https://www.uniprot.org/uniprot/P0DTC2>), P59595 (<https://www.uniprot.org/uniprot/P0DTC5>), P0DTC9 (<https://www.uniprot.org/uniprot/P0DTC9>), and P0DTC4 (<https://www.uniprot.org/uniprot/P0DTC4>), respectively.

Prediction of linear and conformational B-cell epitopes

IEDB server (<http://tools.iedb.org/bcell/>) was utilized for predicting liner B cell epitopes. This server utilizes a method that considers general basis parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity, and the antigenic propensity of polypeptides chains to

predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.

Also, discontinuous B cell epitopes were determined using the ElliPro server in <http://tools.iedb.org/elliopro/>. ElliPro predicts discontinuous epitopes based on a protein antigen's 3D structure and allocates a PI (Protrusion Index) value averaged over epitope residues for each as a score. Finally, it defines discontinuous B cell epitopes with PI and it clusters them based on the distance R (in Å between residue's centers of mass). We used 6VSB, 6YI3, and 7K3G PDB structures related to S, N, and E proteins. For M protein PDB, we used RaptorX to model M protein 3D structure.

Cytotoxic T lymphocyte epitope prediction

NetMHC—4.0 service (<https://services.healthtech.dtu.dk/service.php?NetMHC-4.0>), was used to identify the binding of peptides to MHC class I molecules. NetMHC—4.0 offers 81 different Human MHC alleles including HLA-A, -B, -C, and -E based on trained ANNs. Here we utilized all supertypes HLA with 12mers epitopes in length and considered the strongest epitope bindings.

Helper T lymphocyte epitope prediction

Helper T lymphocytes (HTL) were predicted by NetMHCII-2.3 (<https://services.healthtech.dtu.dk/service.php?NetMHCII-2.3>). This server calculates the binding of peptides to HLA-DR, HLA-DQ, and HLA-DP MHC class II molecules through artificial neuron networks. We investigated all four S, M, N, and E proteins for 25 HLA-DR alleles, 20 HLA-DQ, and 9 HLA-DP class II alleles. Finally, we selected MHC class II epitopes (15mers) with strong epitope bindings.

Determination of epitope conservancy

Epitope conservancy analysis (<http://tools.iedb.org/conservancy/>) was used to compute the degree of the conservancy of selected epitopes within the given S, M, N, and E protein sequence set at a given identity level. Epitope conservancy analysis tool defines conservancy as the fraction of protein sequences that contain the epitope. Also, the Identity of each selected epitope is considered as the degree of correspondence or similarity between two sequences.

Population coverage

We used the population coverage tool at <http://tools.iedb.org/population/> to determine different epitope frequencies in different ethnicities. Because MHC molecules have multiple polymorphisms, predicting the binding of an epitope

to multiple MHC molecules reduces the denominated MHC restriction of T cell responses. The population coverage tool utilizes an allele frequency database to calculate multiple population coverages in both MHC I and MHC II molecules for selected epitopes.

Prediction of antigenicity, allergenicity, toxicity, and interferon-gamma inducer

T and B cell epitopes were evaluated in terms of antigenicity, allergenicity, and toxicity by VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), ToxinPred (<http://crdd.osdd.net/raghava/toxinpred/>), and AllerCatPro (<https://allercatpro.bii.a-star.edu.sg/>) servers, respectively. Here, we consider a threshold of 0.4 for antigenicity prediction. ToxinPred and AllerCatPro v1.7 server's properties were adjusted as default. Furthermore, HTL epitopes were investigated for their potential to induce interferon-gamma (IFN- γ) at <http://crdd.osdd.net/raghava/ifnepitope/index.php> server.

Construction, modeling, refinement, and validation of vaccine structure

The highest immunogenic B cell, CLT, and HLT epitopes were selected, and they fused by KK, GPGPG, and AAY linkers. Then, I-TASSER online tool (<https://zhanglab.dcm.med.umich.edu/I-TASSER/>) was used to predict the 3D structure of the vaccine. GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) provides a refined modeled 3D structure. Future validation of modeled 3D structure was evaluated by the ProSA-web server (<https://www.came.sbg.ac.at/prosa.php>). Also, the Ramachandran graph of the 3D structure was drawn by the MolProbity server (<http://molprobity.biochem.duke.edu/>).

In silico cloning, codon optimization, and physico-chemical properties of the vaccine

Nucleotide sequences of final construct optimized in the Java Codon Adaptation Tool (JCat) (<https://www.jcat.de/>) to reach maximum expression yield at *E. coli* K-12 strain. Also, GC content and codon adaptation index (CAI) score were obtained in <http://genomes.urv.es/CAIcal/E-CAI/>. *Xho*I and *Nde*I restriction enzymes were considered in N-terminal and C-terminal of construct, as well as, HisTag put in C-terminal. Finally, in silico cloning of construct was done by SnapGene tool in pET26b(+). Also, the ProtParam tool in the ExPASy server (<https://web.expasy.org/protparam/>) was used to predict the Physico-chemical properties of the vaccine.

Docking of CoVac19-TLR7 and CoVac19-TLR8

Vaccine modeled structure was minimized in terms of energy by Minimize Structure tab in Chimera 1.13.1 V. Then all water molecules were removed from the structure, and add hydrogen molecule to structure. TLR7 and TLR8 structures were retrieved from the RCSB database with 7CYN and 5WYX PDB numbers, respectively. HADDOCK server (<https://alcazar.science.uu.nl/services/HADDOCK2.2/>) was used to docking the CoVac19 structure with TLR7 and TLR8. HADDOCK results were analyzed and selected the best cluster complex with the lowest energy. Docking 2D results were evaluated by LigPlot⁺.

MD simulation and MM/PBSA analysis

We utilized GROMACS (GROningen MAchine for Chemical Simulations) to performed molecular dynamics simulation of docked complexes in Linux. Needed topology files related to structure energy minimization and equilibration were generated from OPLS-AA (Optimized Potential for Liquid Simulation-All Atom) force field constraints. The system was equilibrated with a water model, and a dodecahedron box was used, with periodic boundary conditions. The vaccine and TLRs structures were investigated in terms of net charge, and the system was neutralized via charged ion addition. The simulation run was performed for 40 ns, and MD results were visualized via the Grace tool in Linux.

We calculated different contributions for the binding free energy of complexes via the MM/PBSA approach.

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,complex}) \quad (1)$$

$$\Delta G_{solv} = G_{electrostatic\epsilon=80} - G_{electrostatic\epsilon=1} + \Delta G_{hydrophobic} \quad (2)$$

$$\Delta G_{vacuum} = \Delta E_{molecularmechanics} - T \cdot \Delta S_{normalmodeanalysis} \quad (3)$$

Here, we have the calculation of solvation free energies using solving the linearized Poisson Boltzman equation for giving the electrostatic contribution to the solvation free energy. The average interaction energy between receptor and ligand and taking the entropy change upon binding into account was used to calculates ΔG_{vacuum} .

Simulating immune responses

Profile of immune response and immunogenicity affected by Cov vaccine were evaluated via C-IMMSIM server (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>), a well-known prediction immune simulation. C-IMMSIM utilized position-specific scoring matrices (PSSM) to predict immune response based

on machine learning techniques. We assumed 4 weeks as the minimum interval recommended for the first and the second doses of injection. Time steps were assumed 1050 so that if a time step will consider about 8 h. Every three injections were given 4 weeks apart at time steps 1, 84, and 168.

Results

Protein sequence retrieval

All four S, M, N, and E proteins were retrieved from the UniProt server and were evaluated in terms of their potential to elicit B cell epitopes and T cell epitopes.

Prediction of linear and conformational B-cell epitopes

Eight liner B cell epitopes were selected based on the highest immunogenicity to induce humoral immunity by the vaccine. B cell epitopes were predicted by the IEDB server and then were evaluated by the Vaxigen server in terms of antigenicity (Table 1). Furthermore, conformational B cell epitopes were investigated by the ElliPro server, where that which conformational B cell epitopes overlapped with liner B cell epitopes (Supplementary Table S1).

Cytotoxic T lymphocyte epitope prediction

Due to the importance of inducing cellular immunity, all

four proteins S, N, M, and E were evaluated using a NetMHC I server. Fifteen CTL epitopes were used in the final structure of the vaccine, based on their binding affinity and antigenicity (Table 2). All identified CTL epitopes are included in Supplementary Table S2.

Helper T lymphocyte epitope prediction

HTL epitopes were investigated throughout the NetMHCII server. Most string binding as well as a most antigenic epitope for HLA-DR, HLA-DP, and HLA-DQ (Table 3). Also, the potential of each selected HTL epitopes to induce IFN- γ was investigated and shown in Table 3. All identified HTL epitopes are included in Supplementary Table S3.

Determination of epitope conservancy

Epitope conservancy for B cell, CTL, and HTL selected epitopes were shown a high conservancy percentage for all

Table 1 Predicted linear B cell epitopes of S, M, N, and E proteins

No	Start	End	Length	Antigenicity	Conservancy
S					
MDLEGKQGNFKNL	177	189	13	1.2592	100%
YNSASFSTFKCYGVSPKLNLCFT	369	393	25	1.4031	100%
GDEVQRQIAPGQTGKIADYNYKLP	404	426	23	1.1017	100%
SNKKFLPF	555	562	8	1.3952	99%
GQSKRVDFC	1035	1043	9	1.7790	100%
M					
PLLESE	132	137	6	0.8069	100%
N					
DPNFKD	343	348	6	2.8780	100%

four S, M, N, and E proteins so that B-cell, CTL, and HTL selected epitopes reached 99–100%, 85–100%, and 81–100% epitope conservancy, respectively. All epitope conservancy results are shown in Tables 1, 2, and 3. However, most epitope conservancies were belonged to S, while E protein has shown the lowest epitope conservancy. These results showed selected epitopes have the potential to be considered as conserved epitopes.

Population coverage

Population coverage was estimated for each selected epitope, which assessments a different epitopes frequency in different ethnicities. All three MHC I, MHC II, and combined MHC

I–MHC II class alleles were considered, and it was found that, generally, combined class alleles showed the highest population coverage for epitopes so that 94.58% (average) of selected epitopes have appeared in 16 regions (Fig. 1; Supplementary Table S4).

Construction, modeling, refinement, and validation of vaccine structure

Seven B cell, 15 CTL, and 12 HTL epitopes were evaluated and selected based on the highest antigenicity, and then fused by KK, GPGPG, and AAY linkers. Two well-known adjuvants so-called Human Beta-defensin 3 (HBD3) and PADRE were added to the N-terminal of the

Table 2 Predicted CTL epitopes related to S, N, M, and E by NetMHC I server

Epitope position	HLA	Peptide	Affinity (nM)	% Rank	Antigenicity	Allergenicity	Conservancy
Spike glycoprotein							
383	HLA-A0201	PTKLNLCFTNV	134.69	1.30	2.1308	Non-allergen	98%
897	HLA-B0801	FAMQMAYRFNGI	729.03	1.10	0.8395	Non-allergen	100%
1208	HLA-A2402	YIKWPWYIWLGF	42.03	0.08	1.4087	Non-allergen	100%
Membrane protein							
48	HLA-A0201	IKLIFLWLLWPV	13.17	0.17	0.7178	Non-allergen	94%
70	HLA-B3901	YRINWITGGIAI	3608.58	1.70	1.2741	Non-allergen	90%
81	HLA-B5801	IAMACLVLMLWL	41.73	0.25	0.9300	Non-allergen	93%
Nucleoprotein							
91	HLA-B2705	RRIRGGDGKMKD	428.31	1.30	0.7443	Non-allergen	100%
101	HLA-B0801	KDLSRWYFYLYL	290.23	0.60	1.0134	Non-allergen	100%
225	HLA-A0301	RLNQLESKMSGK	56.27	0.30	0.7580	Non-allergen	98%
312	HLA-A0201	AFFGMSRIGMEV	520.61	3.00	0.8821	Non-allergen	97%
319	HLA-B4001	IGMEVTPSGTWL	500.92	0.80	0.8546	Non-allergen	98%
319	HLA-B5801	IGMEVTPSGTWL	373.42	0.80	0.8546	Non-allergen	98%
Envelope							
19	HLA-B3901	FLAFVVFLVTL	21,440.88	12.00	0.7623	Non-allergen	86%
36	HLA-B2705	LRLCAYCCNIVN	3227.68	4.50	0.8619	Non-allergen	94%
41	HLA-A0301	YCCNIVNVSLVK	1832.88	3.00	0.9775	Non-allergen	85%
52	HLA-B0702	KPSFYVYSRVKN	15,461.92	11.00	0.9436	Non-allergen	88%

Table 3 Predicted HTL epitopes related to S, N, M, and E by NetMHCII server

Allele	Position	Peptide	Affinity (nM)	% Rank	Antigenicity	Conservancy	IFN γ	CD8
Spike glycoprotein							Default	Random Forest based
DRB4_0101	894	LQIPFAMQMAY-RFNG	22.2	0.70	0.7205	100%	Negative	IL10 inducer
DPA10301–DPB10402	508	YRVVVLSFELL-HAPA	11.3	0.12	0.7072	97%	Positive	IL10 inducer
DPA10103–DPB10201	1207	EQYIKWPWYIWL-GFI	19.8	0.60	1.0096	100%	Positive	IL10 inducer
DQA10601–DQB10402	628	QLTPTWRVYSTG-SNV	24.7	0.07	0.9276	99%	Negative	IL10 non-inducer
Membrane protein								
DRB3_0101	34	LLQFAYANRNR-FLYI	41.0	2.50	0.7387	93%	Positive	IL10 non-inducer
DPA10103–DPB10301	98	ASFRLFARTRSM-WSF	32.4	0.60	0.7304	97%	Negative	IL10 inducer
DQA10301–DQB10301	74	NWITGGIAIAMA-CLV	26.8	0.30	0.8998	92%S	Negative	IL10 non-inducer
Nucleoprotein								
DPA10201–DPB10101	386	QKKQQTVTLL-PAADL	183.2	7.00	0.7662	91%	Positive	IL10 inducer
DPA10103–DPB10201	103	DLSRWYFYYL-GTGP	50.3	3.00	1.5180	100%	Positive	IL10 non-inducer
Envelope								
DQA10101–DQB10501	37	LRLCAYCCNIVN-VSL	744.9	15.00	1.0896	87%	Positive	IL10 inducer
DQA10301–DQB10301	34	TALRLCAYCC-NIVNV	446.1	15.00	0.8876	91%	Positive	IL10 non-inducer
DQA10601–DQB10402	51	LVKPSFYVYS-RVKNL	86.3	0.70	0.7311	81%	Positive	IL10 inducer

construct (Fig. 2). Amino acid sequence was submitted to the I-TASSER server for predicting the 3D structure of the

vaccine. GalaxyRefine server improved raw modeled 3D of vaccine to a decreased RMSD structure in angstrom scale.

Fig. 1 Population Coverage of selected epitopes

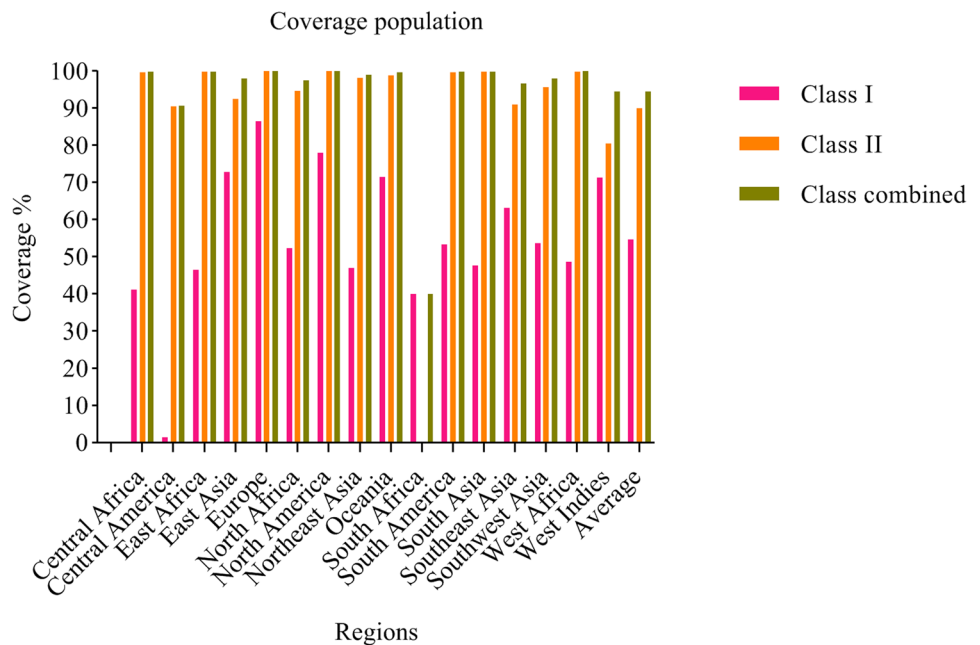
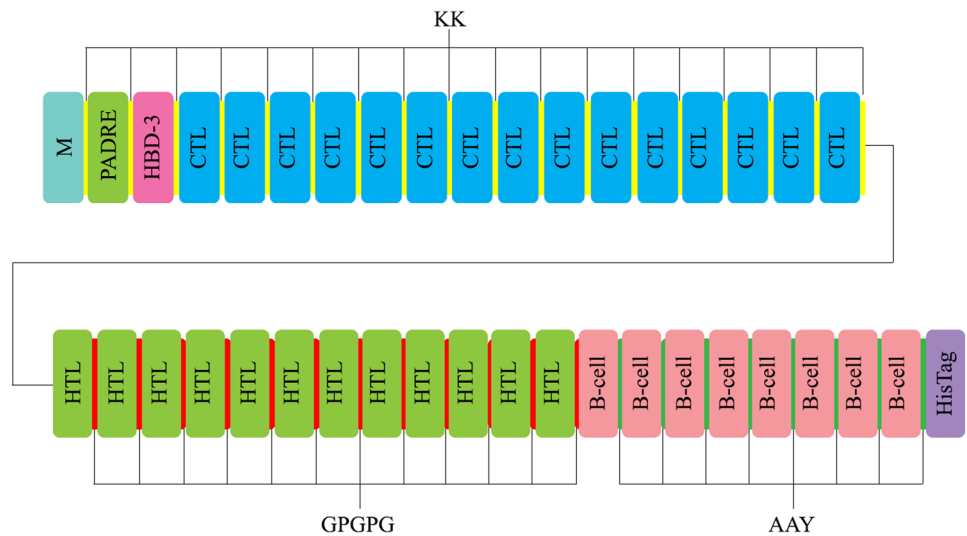


Fig. 2 Graphical representation of COVID-19 vaccine candidate constructs



Further validation parameters for the refined 3D structure shown a Z-score equal to -2.56 (Fig. 3a and b). Plots of residues scores related to local model quality by plotting energies (before and after model refining) were compared in supplementary Fig. S1. Furthermore, the residue energies of the structure, related to before and after model refining, are shown in supplementary Fig. S2. Also, the Ramachandran graph showed 90% and 98.1% amino acids are in favored and allowed regions (Fig. 3c and d). The refined modeled COVID-19 structure is shown in Fig. 3e.

Codon optimization, in silico cloning, and physico-chemical properties of the vaccine

Codon optimization was done for reaching maximum expression in *E. coli* (strain K-12). We have assumed parameters for avoiding rho-independent transcription terminators and avoiding prokaryotic ribosome binding sites. Furthermore, it avoided cleavage sites of *XhoI* and *NdeI* restriction enzymes. Nucleotide sequence of vaccine with the length of 1899 bp was submitted to Codon Adaptation Tool, and results showed CAI-Value of 0.96 and GC-Content of 50.50, after codon optimization.

The construct was posited in pET26b(+) to in silico cloning by the SnapGene tool (Fig. 4). *XhoI* and *NdeI* restriction enzymes were considered in the N-terminal and C-terminal of the construct. Also, a HisTag sequence was included in the C-terminal of the construct. Then, the final nucleotide sequence was submitted to Expsy Translate Toll and it found that the construct is in the frame.

ProtParam tool was used to calculate physico-chemical properties of vaccine, and results showed COVID-19 vaccine is consists of a protein with 633 amino acid sequence, with a molecular weight of ~ 70 kDa. Theoretical pI was equal to 9.99, and it probably has an estimated half-life of

30 h (mammalian reticulocytes, in vitro), > 20 h (yeast, in vivo), and > 10 h (*Escherichia coli*, in vivo). The instability index and Aliphatic index were calculated at about 31.98 and 77.81, respectively. Also, this protein obtained a Grand average of hydropathicity (GRAVY) of -0.141 . All calculated physico-chemical parameters of Cov protein indicated an appropriate construct for expression.

Docking of CoVac19-TLR7 and CoVac19-TLR8

HADDOCK server was used to dock the CoVac19 structure with TLR7 and TLR8. Twelve clusters related to the CoVac19-TLR7 complex and 15 clusters related to the CoVac19-TLR8 complex were obtained in each HADDOCK run. Analysis results of docking showed both CoVac19-TLR7 and CoVac19-TLR8 complexes could establish strong interaction so that CoVac19-TLR7 reached a HADDOCK score of -205.9 , as well as CoVac19-TLR8 complex HADDOCK score was -204.4 . Further analysis indicates the important role of Van der Waals and electrostatic energies in both CoVac19-TLR7 and CoVac19-TLR8 complexes (Tables 4 and 5) (Figs. 5a–c and 6a–c).

Table 4 HADDOCK results related to CoVac19-TLR7 complex

HADDOCK score	-205.9 ± 11.1
Cluster size	5
RMSD from the overall lowest-energy structure	1.9 ± 2.3
Van der Waals energy	-51.1 ± 7.8
Electrostatic energy	-273.3 ± 27.7
Desolvation energy	-123.9 ± 8.7
Restraints violation energy	237.3 ± 15.65
Buried surface area	1998.1 ± 164.1
Z-score	-1.9

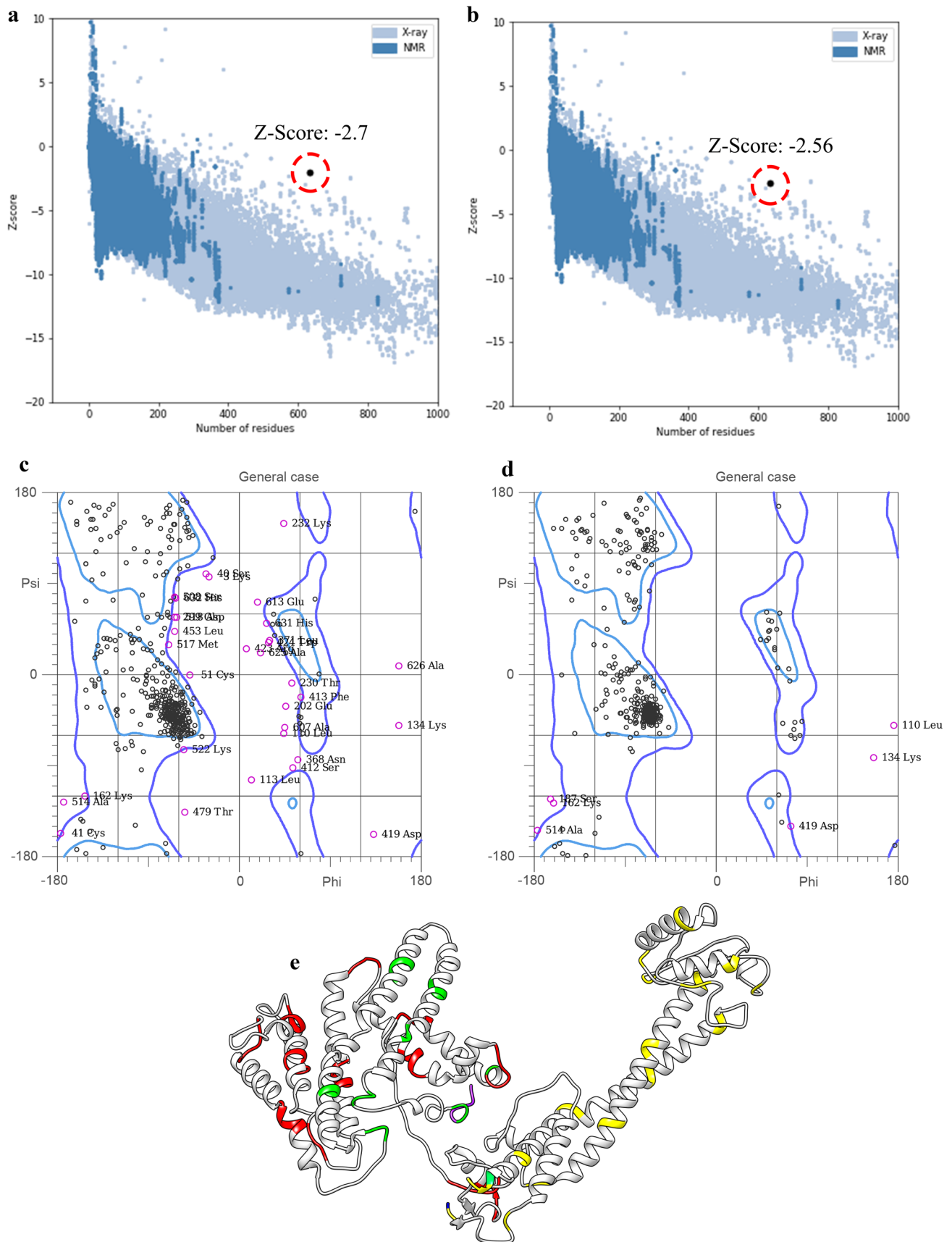


Fig. 3 Modeling, refinement, and validation of COVID-19 vaccine 3D structure. **a** and **b** Z-score of COVID-19 vaccine 3D structure, before and after structure refinement. **c** and **d** Ramachandran graph related to COVID-19 vaccine 3D structure, before and after structure refinement. **e** Refined of modeled COVID-19 vaccine 3D structure. KK, GPGPG, and AAY linkers are shown in yellow, orange, and green ribbons, respectively

As is shown in Fig. 5c, Ala16, Asn22, Thr23, Lys64, and Thr67 amino acids from the COVID-19 vaccine structure have formed hydrogen interactions with His86, Asp89, Pro120, Arg121, and Glu170 amino acids in the CoVac19-TLR7 complex. Also, a total of 14 hydrogen bonds including the interaction between Ala15, Lys18, Asn22, Ile21, Gln25, Arg32, Lys64, Thr67, Lys68, Asp71, and Leu122 of COVID-19 vaccine structure amino acids and Glu59, Asp96, Lys103, Gln117, Ile118, Pro119, Gly121, Glu124, and Glu171 TLR8 structure amino acids are responsible for forming the complex of CoVac19-TLR8.

Formed hydrogen bonds in CoVac19-TLR7 complex (9 hydrogen bonds), as well as CoVac19-TLR8 complex (14 hydrogen bonds), are shown as 2D presentation in Fig. 7a and b, respectively.

MD simulation and MM/PBSA analysis

MD simulation of docking results showed stable interactions between both CoVac19-TLR7 and CoVac19-TLR8

complexes. RMSD graph related to CoVac19-TLR7 is 0.8 nm which confirms an appropriate constancy (Fig. 8a). Also, the CoVac19-TLR7 RMSF graph is reached amino acid stability during MD simulation (Fig. 8b). The gyration radiation graph had a high rate of about 2.1 nm but it decreased, gradually, and showed an index of 1.9 nm for CoVac19-TLR7 (Fig. 8c). In addition, the H-bond graph indicates CoVac19-TLR7 complex maintains its interactions all over MD simulation because was not observed any decrease in this complex (Fig. 8d).

The CoVac19-TLR8 complex also showed an RMSD equal to 0.9 nm, in agreement with complex stability in MD (Fig. 9a). RMSF graph for CoVac19-TLR8 complex for all amino acids has without fluctuation (Fig. 9b). A rate of 0.8 nm for gyration radiation confirmed the stability of the CoVac19-TLR8 complex (Fig. 9c). However, maximum maintenance of interaction is shown in the H-bond graph (Fig. 9d). Overall MD simulation results indicate stable interaction in both CoVac19-TLR7 and CoVac19-TLR8 complexes.

Here, we did further analysis for MD simulation results and the contribution of involved amino acids in complexes, via MM/PBSA calculation. As shown in Table 6, binding energy equal to − 1842 kJ/mol is obtained for CoVac19-TLR7 complex, as well as van der Waal energy and electrostatic energy of − 363.404 kJ/mol and − 2198.164 kJ/mol, respectively. We can see that binding energy is equal

Fig. 4 In silico cloning of COVID-19 vaccine nucleotide sequence to pET26b(+) by SnapGene

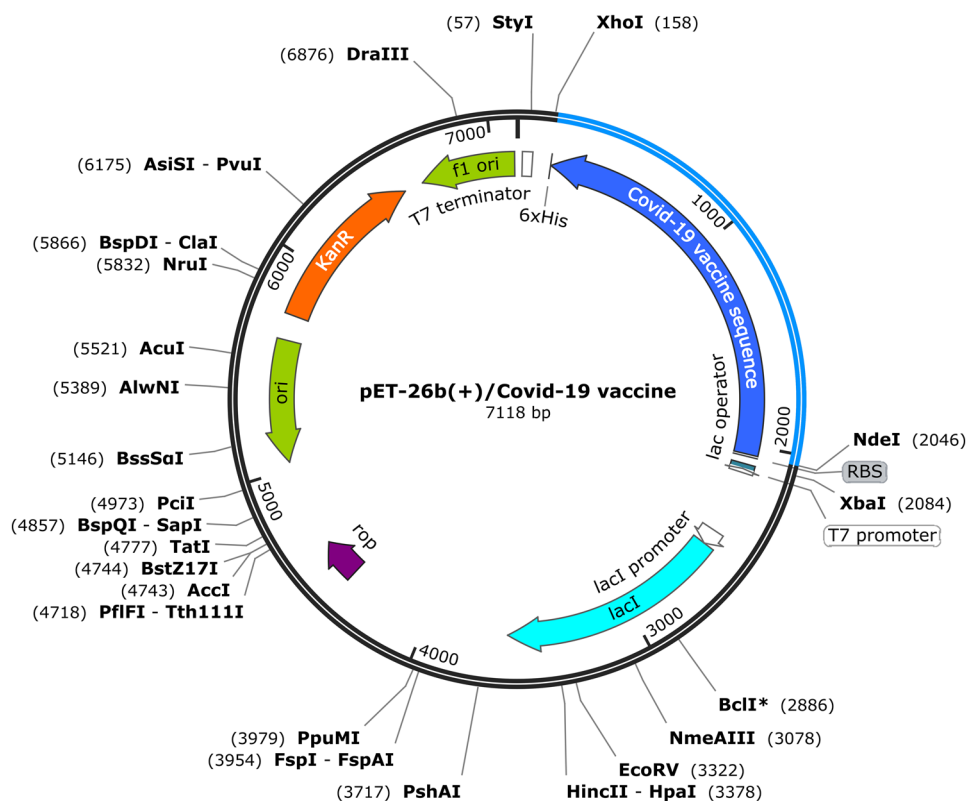


Table 5 HADDOCK results related to CoVac19-TLR8 complex

HADDOCK score	-204.4 ± 10.1
Cluster size	12
RMSD from the overall lowest-energy structure	1.7 ± 1.5
Van der Waals energy	-68.8 ± 10.3
Electrostatic energy	-438.9 ± 73.2
Desolvation energy	-93.7 ± 6.0
Restraints violation energy	459.1 ± 34.98
Buried surface area	2390.1 ± 131.4
Z-score	-1.8

to -3336.965 kJ/mol, in the CoVac19-TLR8 complex, compared to CoVac19-TLR7 complex, which is referred to the more hydrogen bonds in the CoVac19-TLR8 complex (Table 7). Furthermore, Van der Waal energy and electrostatic energy of -633.075 kJ/mol and -4043.666 kJ/mol, respectively, in CoVac19-TLR8 complex; that is negative energy similar to CoVac19-TLR7 complex. Overall MM/PBSA analysis confirmed the stability of complexes.

Simulating immune responses

C-IMMSIM immune provided us an in silico perspective from the generated immunogenic profile by the vaccine. Primary response originated from the immune stimulation was observed in much lower levels compared to the secondary and tertiary responses. We assumed 3 injections with 4-week

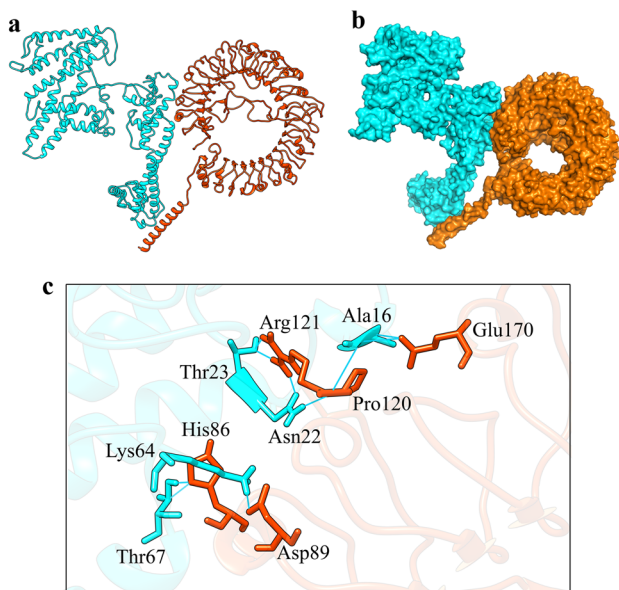


Fig. 5 HADDOCK docking results of COVID-19 vaccine-TLR7 complex. **a** CoVac19-TLR7 complex is shown as ribbons. **b** CoVac19-TLR7 complex is shown as surface. **c** CoVac19-TLR7 complex is shown as magnified sticks. CoVac19 structure is blue, and TLR7 structure is orange

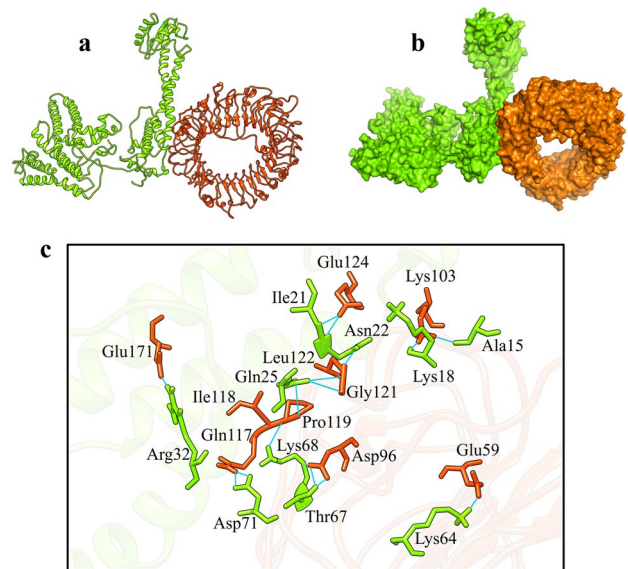


Fig. 6 HADDOCK docking results of COVID-19 vaccine-TLR8 complex. **a** CoVac19-TLR8 complex is shown as ribbons. **b** CoVac19-TLR8 complex is shown as surface. **c** CoVac19-TLR8 complex is shown as magnified sticks. CoVac19 structure is green, and the TLR8 structure is orange

intervals between each injection. Derived immune responses showed IgM + IgG level (Fig. 10a). A memory cell development is observed in tertiary response associated with robust activation level of B cells such as B isotypes of IgM and IgG (Fig. 10b and c). Active T cells in both the CD4 T-helper lymphocyte and CD4 T-helper lymphocyte responses associated with secondary and tertiary reactions increase significantly and subsequently decrease (Fig. 10d–f). The TC cell population also shows a correlation with the pre-activation of all of TCs (Fig. 10g). Because cellular immunity plays an important role in combat to COVID-19 disease, it is observed that natural killer cells and dendritic cells are consistent with the rapid reproduction of macrophages (Fig. 10h–j). Finally, high levels of generated cytokines such as IFN- γ and IL-2 affected due to injections confirm acceptable stimulation of the immune system by this vaccine (Fig. 10k).

Discussion

With the advent of SARS-CoV-2 and, consequently, with the announcement by the World Health Organization that COVID-19 was a pandemic, researchers began efforts to obtain vaccines to prevent the disease (Cucinotta and Vanelli 2020; Sohrabi et al. 2020). Classic live or attenuated vaccines have a long history of success, but cause problems such as autoimmune and allergic reactions. This has led to the use of immunoinformatics methods to eliminate such biosafety issues, along with taking advantage of time and cost savings.

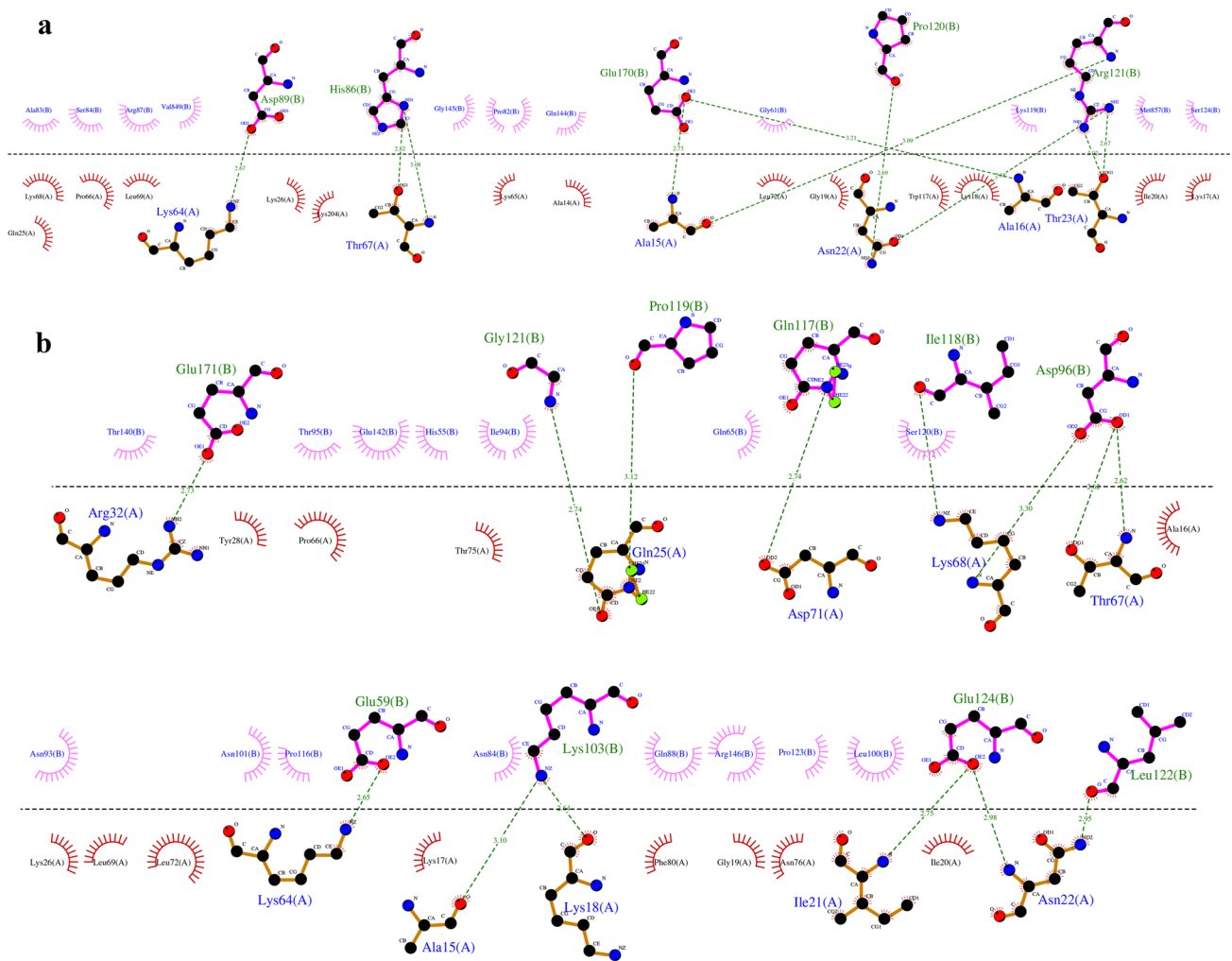


Fig. 7 2D presentation of HADDOCK results. **a** CoVac19-TLR7 complex. CoVac19 amino acids are shown as orange sticks, and TLR7 amino acids are shown as pink sticks. **b** CoVac19-TLR8 complex.

CoVac19 amino acids are shown as orange sticks, and TLR8 amino acids are shown as pink sticks

The multi-epitope vaccines are an effective and promising option to overcome the limitations of classical vaccines. One of the salient features of multi-epitope vaccines compared to conventional vaccines is that due to viral genome screening, stimulation of targeted immune responses occurs following accurate identification of immunogenic epitopes. To date, several vaccines against SARS-CoV-2 have been developed and reported (Ahmad et al. 2020; Behmard et al. 2020; Kar et al. 2020; Khairkhan et al. 2020; Mukherjee et al. 2020). Due to the existence of sufficient information about the genomics and proteomics of SARS-CoV-2, the most effective SARS-CoV-2 antigenic epitopes can be identified and used to make the COVID-19 vaccine by relying on in silico methods.

In our study, we used a variety of computational tools alongside immunoinformatics methods to design a multi-epitope vaccine capable of eliciting both humoral and

cellular immune responses (De Groot et al. 2020). In addition, multi-epitope vaccines specifically have epitopes that bind to CTL, HTL, and B cells, along with the use of protein adjuvants in this type of vaccine (Sette and Fikes 2003). Here, we used a multi-epitope vaccine against proteins S, M, N, and E. As a modulating tropism, protein S is one of the most important factors of viral entry into the host cell (Ou et al. 2020). On the other hand, since SARS-CoV-2 is an RNA virus, a mutation is expected. Most of these mutations occur in surface proteins such as protein S, which can mislead the immune system (Sanjuán and Domingo-Calap 2016). Therefore, we used protein S as the major and most important component in this multi-epitope vaccine, along with structural proteins M, N, and E, to maximize the immune response against SARS-CoV-2. Structural proteins have more protected regions because they are under less structured evolution (Satarker and Nampoothiri 2020).

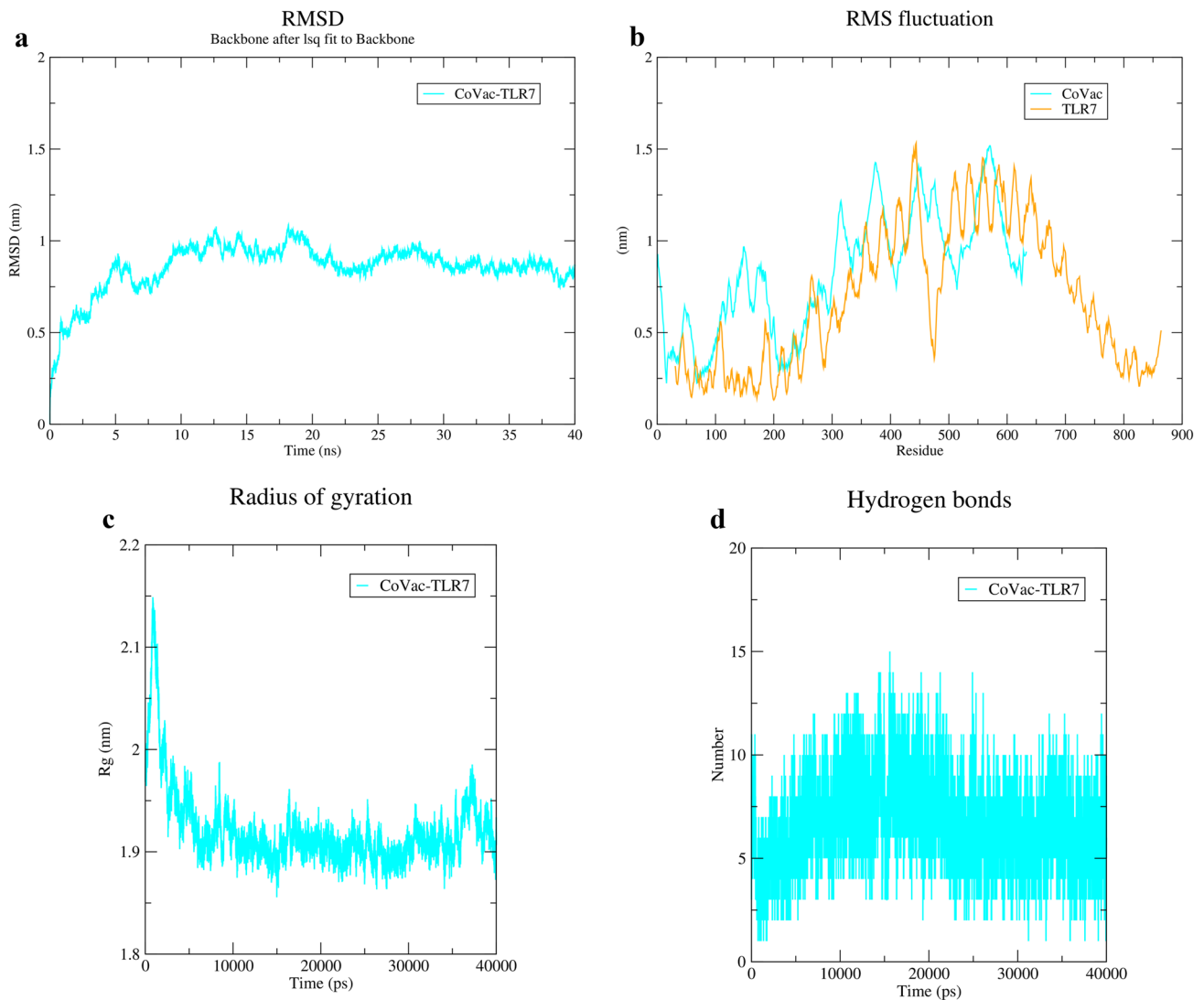


Fig. 8 MD simulation results of CoVac19-TLR7 complex. **a** RMSD, **b** RMSF, **c** Rg, and **d** hydrogen bond graphs related to CoVac19-TLR7 complex

It is expected that epitopes extracted from these protected areas will lead to increased diversity and immune response. Studies has shown M glycoprotein, as most abundant structural protein of SARS-COV-2, leads to promoting membrane fusion and viral infectivity via S protein binding and the host surface receptor (Hu et al. 2003; Thomas 2020). Furthermore, M could affect virus-induced immune responses of the host throughout involving in regulation of replication and packing the genomic RNA into viral particles (Zheng et al. 2020). N protein of SARS-COV-2 is responsible for the package the genomic viral genome into ribonucleoprotein complexes which are known as nucleocapsids (de Haan and Rottier 2005). N proteins are included in three highly conserved domains called N terminal domain 1, RNA-binding domain 2, and C-terminal domain 3 (McBride et al. 2014). Genome replication and transmission are affected by

protection on the genome via N protein (de Haan and Rottier 2005). Also, N protein plays a key role in viral transcription, translation and CoV virion via its interaction with gRNA and sgRNA molecules (Cubuk et al. 2021; Narayanan et al. 2003).

In our study, we used the beta-defensin 3 sequences as an adjuvant. Beta-defensin 3 has been shown to act as a potential adjuvant in the N-terminal region of the protein. Also, we included the PADRE sequence in the construct that acts as a peptide with very high adjuvant properties. The glycine-rich binder, such as GPGPG, was used to bind selected epitopes. GPGPG makes adjacent domains available by increasing the solubility of the protein structure and allows them to function more freely; also, it acts as the immune-modulatory (Kavoosi et al. 2007). The KK linker was also used to aid in antigen processing. Studies have

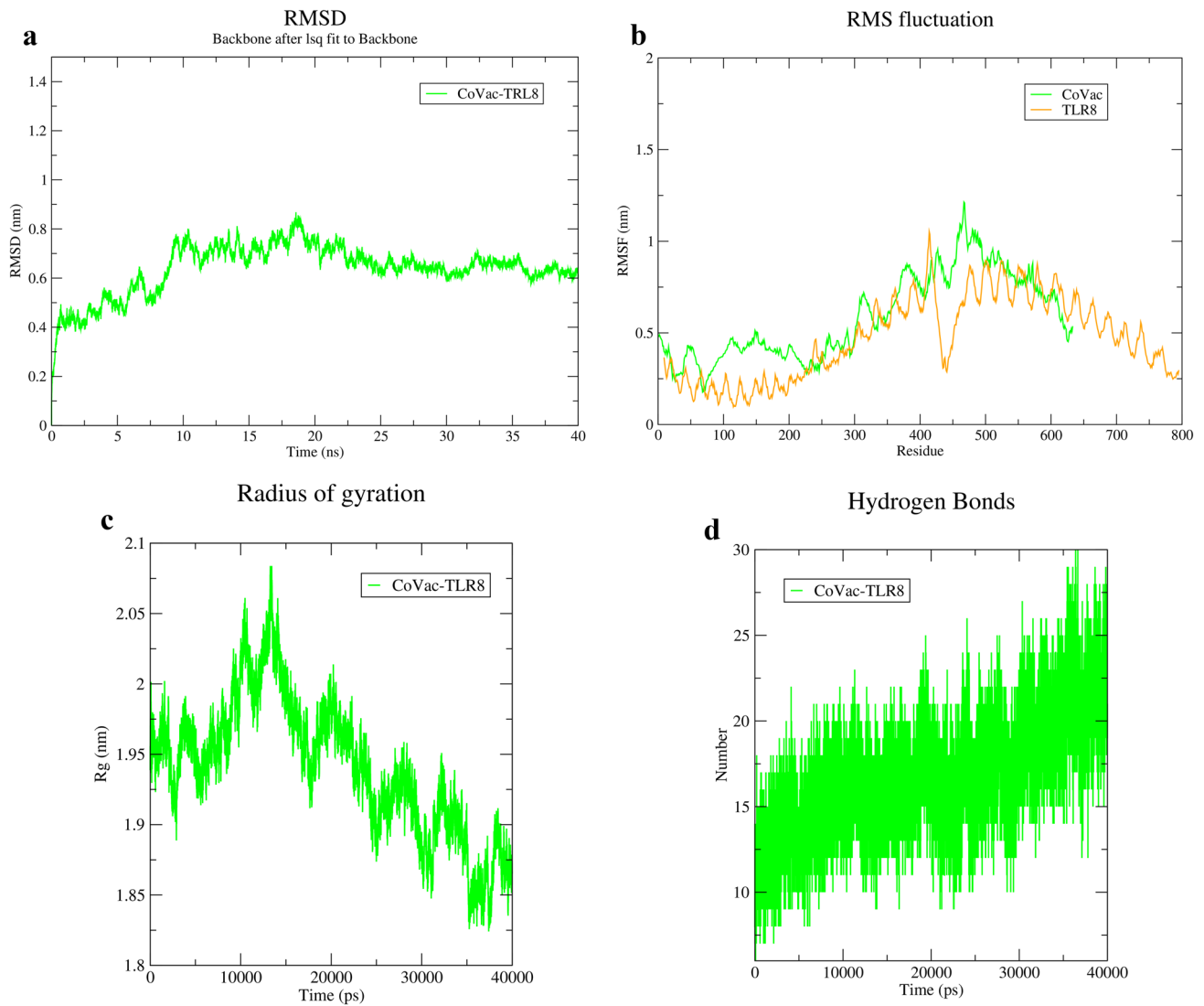


Fig. 9 MD simulation results of CoVac19-TLR8 complex. **a** RMSD, **b** RMSF, **c** Rg, and **d** hydrogen bond graphs related to CoVac19-TLR8 complex

shown that the KK linker targets the Cathepsin B sequence as a lysosomal protease for antigen processing. During the screening, the predicted CTL and HTL epitopes were examined from the total human HLA supertypes for the MHC class I and MHC class II alleles, and the most antigenic epitopes were selected.

The obtained index of more than 95% of the population coverage for CTL epitopes and helper T cells in 16 specific geographical areas showed that selected epitopes among target populations commonly bind to HLA molecules. Linear B cell epitopes were used in conjunction with conformational

Table 6 MM/PBSA analysis related to CoVac19-TLR7 complex

MM/PBSA-complexes	CoVac19-TLR7 complex
Van der Waal energy	-363.404 ± 14.338 kJ/mol
Electrostatic energy	-2198.164 ± 78.604 kJ/mol
Polar solvation energy	767.667 ± 154.711 kJ/mol
SASA energy	-48.291 ± 4.832 kJ/mol
Binding energy	-1842.191 ± 100.880 kJ/mol

Table 7 MM/PBSA analysis related to CoVac19-TLR8 complex

MM/PBSA-complexes	CoVac19-TLR8 complex
Van der Waal energy	-633.075 ± 52.833 kJ/mol
Electrostatic energy	-4043.666 ± 99.422 kJ/mol
Polar solvation energy	1418.238 ± 145.798 kJ/mol
SASA energy	-78.462 ± 6.429 kJ/mol
Binding energy	-3336.965 ± 108.220 kJ/mol

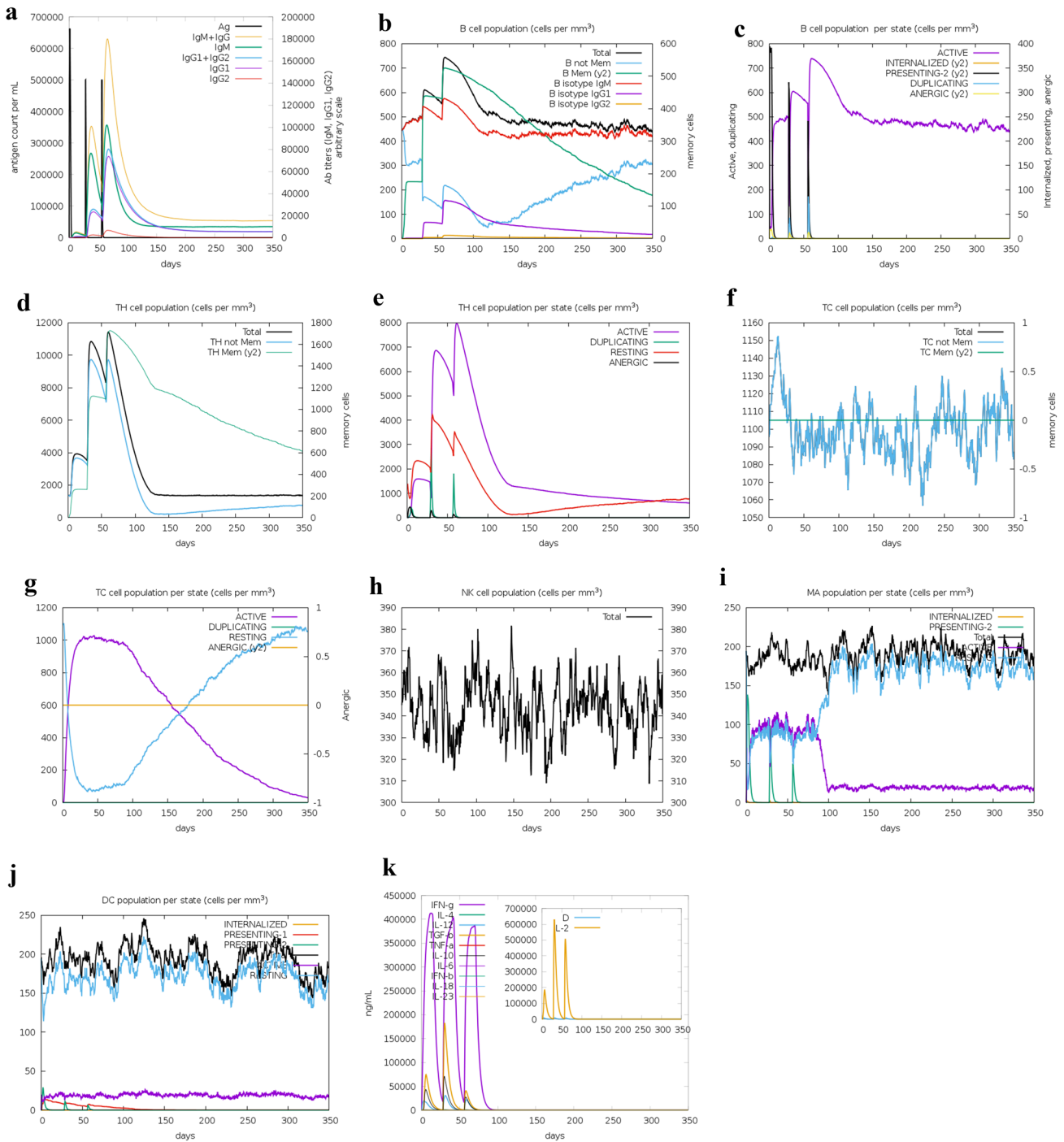


Fig. 10 Immune simulation after COVID-19 vaccine injection. **a** Antigen and immunoglobulins, **b** B lymphocytes: total count, memory cells, and sub-divided in isotypes IgM, IgG1, and IgG2; **c** B lymphocyte population per entity-state (i.e., showing counts for active, presenting on class-II, internalized the Ag, duplicating and anergic; **d** CD4 T-helper lymphocytes count. The plot shows total and memory counts; **e** CD4 T-helper lymphocyte count sub-divided per entity-state (i.e., active, resting, anergic, and duplicating), **f** CD8 T-cytotoxic

lymphocytes count. Total and memory shown: **g** CD8 T-cytotoxic lymphocytes count per entity-state, **h** natural killer cells (total count); **i** dendritic cells. DC can present antigenic peptides on both MHC class-I and class-II molecules. The curves show the total number broken down to active, resting, internalized, and presenting the ag; **j** macrophages. total count, internalized, presenting on MHC class-II, active and resting macrophages; **k** cytokines. The concentration of cytokines and interleukins. D in the inset plot is danger signal

B cell epitopes to maximize the humoral immune response by the construct.

T cells recognize able to recognize a complex between a specific MHC molecule (called HLA in humans) and a unique epitope that was derived from a pathogen. As a result, that unique epitope only evokes a response in individuals in whom the expressed MHC in that person can bind to the unique epitope. MHCs are classified into two classes, HLA I and II; HLA I molecules are important for peptides that are associated with cytotoxic T lymphocytes (CTL) responses, and HLA II molecules are important for those peptides associated with helper T lymphocyte (HTL) responses. Therefore, each of the CLT epitopes and HTL epitopes can bind to specific HLAs. On the other hand, MHCs have many polymorphisms that are different in each population. Depending on the population coverage, each epitope can bind to a specific allele of HLAs, which is determined as the percentage of the difference in population coverage for each epitope in HLA class 1 and HLA class 2. If each epitope tends to bind to more HLA molecules, it offers better population coverage in relatives. As can be seen in Fig. 1, the difference in coverage percentage between HLA class 1 and HLA class 2 epitopes is reduced in the combined HLA class 1/2 because each epitope was able to incidence the potential to bind to more HLAs (Bui et al. 2006).

Here, we used the IEDB server to predict B cell epitopes. The IEDB server estimated a total of 7 linear B cell epitopes with the highest immunogenicity. Also, out of 48 CTL epitopes, 16 epitopes, and out of 56 HTL epitopes, 12 epitopes were identified, indicating the ability of the construct designed to induce a strong cellular immunity. Given the importance of the non-allergenicity of our vaccine for the immune system, which is generally one of the major limiting factors to the vaccine designing process, we used the AllerCatPro server to predict potential allergenicity. The results confirmed that all selected epitopes are non-allergenic for the immune system. Besides, toxicity evaluation of selected epitopes confirmed each epitope was predicted as non-toxic.

To achieve stable expression in the *E. coli* expression system for *in vitro*, it is necessary to design an *in silico* cloning in which the nucleotide sequence has factors such as (1) the absence of non-specific restriction enzymes on the sequence, (2) the percentage of suitable CAI-Value/GC-Content, (3) placement of the nucleotide sequence in the Open Reading Frame (ORF), and (4) optimization of the nucleotide sequence is based on codon preference in *E. coli*. As shown for *in silico* cloning, the CAI-Value percentages before and after codon optimization were 0.59 and 0.96, respectively, confirming that the CAI-Value index of the nucleotide sequence had improved. Also, the GC-Content before and after codon optimization were 55.39 and 50.50, respectively, which is very close to the GC-Content index of 50.73 in *E.*

coli (strain K12). In addition, the placement of the amino acid sequence after nucleotide translation in the ORF indicates that the nucleotide sequences are in the appropriate frame, and the design of *in silico* cloning is similar to *in vitro* conditions that assume maximum protein expression.

The molecular weight of the construct was 70.87 kDa (633 aa) for this multi-epitope vaccine. Both the epitopes and the final structure were soluble and stable, indicating that the designed construct had high solubility and stability to initiate an immunogenic reaction. The I-TASSER server was used to predict the structure of the third protein. The modeled structure was refined using GalaxyRefine. Comparison of Z-score before refining and after refining shows that the structure improved after refining. Next, the 3D modeled construct was evaluated in terms of structural energy content. It was identified that plot of residue scores plotting energies as a function of amino acid sequence position had multiplicity negative residual energies as well as the low rate of residual positive values. This indicates that positive values corresponding to problematic or erroneous parts have a smaller share of structural energy in plotting energies, so it can be concluded that the predominance of negative energy leads to increased stability and structural stability of the protein. In addition, the residue energy of the three-dimensional structure was associated with the predominantly blue color (regions referred to lowest energy) and less red was observed in the structure (regions referred to highest energy), which indicates low energy of the structure and thus the stability of the 3D structure in terms of energetic content. Stable structures increase the ability of proteins to maintain their interactions with other proteins/ligands (Chamani and Heshmati 2008; Mokaberi et al. 2021). Also, the Ramachandran graph after refinement showed that 98.1% of amino acids are in the allowed region. This indicates that the final structure is of acceptable quality.

Recognition of pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors (PRR) is resulted in activation of response innate host antiviral immunity and plays a role in the control of viral infection (Chow et al. 2018). Generally, viral RNAs are considered as the PAMPs that are recognized by TLRs (Sallenave and Guillot 2020). Previous studies have been shown the role of the TLR7 receptor in host defense against single-stranded RNA viruses such as SARS-CoV-2 (ssRNA) (Fallerini and Daga 2021; Salvi et al. 2021; Solanich et al. 2021). TLR7 and TLR8 are the endosomal TLRs that play essential and unnecessary biological roles in host survival (Shen et al. 2010). Triggering of specific innate immune receptors such as TLR7 and TLR 8 by their agonists can activate innate immune cells which eventually leads to the production of I interferon (Iwasaki and Medzhitov 2010).

SARS-CoV-2 spike glycoprotein is identified as a structural component of the virus by TLR receptors expressed in the plasma membrane of immune cells such as monocytes, macrophages, and immature dendritic cells (Khanmohammadi and Rezaei 2021; Poulas et al. 2020). The binding of COVID-19 antigens to TLRs activates the NF- κ B, protein kinase, and finally the secretion of cytokines (Khanmohammadi and Rezaei 2021).

Although COVID-19 disease is still a relatively new disease and requires more clinical findings, studies have shown the importance of TLR7/8 stimulation in COVID-19 vaccines. In one study, Imidazoquinolines was used as an adjuvant in a SARS-CoV-2 recombinant spike protein vaccine to enhance TLR7/8 stimulation, and the results showed that the vaccine led to product the robust IgG2a and IgG1 antibody titers in a mouse model and neutralized viral infection in both *in vitro* and *in vivo* (Jangra et al. 2021). Responses of Th1 and CD8 + T cells have been shown also against HIV-1 Gag protein as RNA viral infection by TLR7/8 agonist (Wille-Reece et al. 2005a, 2005b).

As Danesh et al. have explained, molecular modeling can provide an insight to locate the exact binding site of the drug/ligand or proteins to the proteins' amino acids (Danesh et al. 2018). Here we used protein–protein molecular docking to investigate the interaction of the vaccine structure with TLR7 and TLR8. Interactions analysis of the vaccine structure with TLR7 and TLR8 showed that 10 hydrogen bonds in CoVac19-TLR7 complex and 14 hydrogen bonds in CoVac19-TLR8 complex, along with electrostatic and Van der Waals bonds, were responsible for the interactions formed in both CoVac19-TLR7 and CoVac19-TLR8 complexes. Stimulation of TLR leads to a specific immune response and activation of a cascade of immune factors. Docking results showed that the interaction between vaccine amino acids and TLR amino acids created complexes with very negative binding energy, indicating that the affinity of the vaccine structure with high TLR structures is high. In general, negative binding energies, Van der Waals, and electrostatic in the formed complexes indicate a high affinity for binding between protein–protein, ligand–protein, or DNA–protein complexes (Dehghani Sani et al. 2018). Several studies on ssRNAs such as SARS-CoV have shown the importance of TLR7 and TLR8 for eliciting an effective immune response (de Groot and Bontrop 2020; Poulas et al. 2020; Safaei and Karimi-Googheri 2020).

In this study, the docking of the vaccine model with the TLR7 was evaluated and then simulated with MD and free energy calculations. Studies have shown that triggering TLR7 is involved in inducing TLR signaling. Subsequently, it can also activate immune pathways involved in pathogenesis and viral infection (Fallerini et al. 2021; Khanmohammadi and Rezaei 2021; Poulas et al. 2020). Studies have shown that TLR8 function is also affected by TLR7

function, and in addition, ssRNAs can stimulate both TLR7 and TLR8. That is why we also docked the construct with the TLR8, and the MD complex results in CoVac19-TLR8 showed an appropriate interaction between the TLR8 and the CoVac19.

Molecular dynamic simulations of the complexes for 40 ns showed very slight oscillation and fluctuations in the RMSD and RMSF diagrams, indicating the high stability and flexibility of the vaccine structure. In addition, analyses of the gyration radius of the amino acids involved in the complexes indicated the stability of both CoVac19-TLR7 and CoVac19-TLR8 complexes during molecular dynamics simulations. In agreement with the RMSD, RMSF, and gyration radius diagrams, the amount of H-bond also indicated that the number of hydrogen bonds involved in the complexes was maintained. The results of MM/PBSA analysis showed that the binding energy of the CoVac19-TLR8 complex was equal to -3336.965 ± 108.220 kJ/mol, which was more negative compared to the binding energy of the CoVac19-TLR7 complex (-1842.191 ± 100.880 kJ/mol). In this respect, the Van der Waals energy of both complexes was approximately equal, while the electrostatic energy of equal -4043.666 ± 99.422 kJ/mol in the CoVac19-TLR8 complex was also more negative than the electrostatic energy of equal -2198.164 ± 78.604 kJ/mol in the CoVac19-TLR7 complex. These results were in agreement with Tables 4 and 5 for the docking analysis; furthermore, the simulation results of the complexes by MD showed that over time, the binding energy in both complexes became more negative, thus indicating that the complexes were stable and retained their strong interactions along the MD. In addition to binding energy, important constants such as Van der Waals and electrostatic energies in the formation of complexes have also been shown and confirmed in other studies (Danesh et al. 2018; Dehghani Sani et al. 2018; Krishnan et al. 2021). Output data from molecular dynamic simulations can be used to evaluate vaccine stability by simulation *in vivo*, wherein a study, the evaluation of TLR7/8 stimulation as *in vivo*, showed that injection of recombinant spike protein candidate vaccine in a mouse model induced the production of IgG2a and IgG1. (Jangra et al. 2021).

Because TLR7/8 is stimulated by single-stranded RNAs, so TLR7 acts as sensors that can play a role in the clearance of SARS-CoV-2 (Onofrio et al. 2020). Stimulation of TLR7/8 can be mediated by the Myeloid Differentiation Primary Response Protein 88 (MyD88) signaling pathway and messenger cascade is activated. The formation of the messenger cascade eventually leads to the nuclear transfer of transcription factors such as NF- κ B and activation of MAPK (mitogen-activated protein kinases). Activation of these transcription factors increases the level of interferon I type mRNA, activation of interferon regulatory factors including IRF7 and IRF3, induction of expression of proinflammatory

cytokines and 1-IL genes, and transcription of interferon type I (α and β) genes, resulting in immune responses and anti-viral responses are induced (Kawasaki and Kawai 2014; O'Neill et al. 2013; Sallenave and Guillot 2020). Successful interaction of the vaccine structure with TLR7/8 structures is likely to activate the messenger cascade and ultimately elicit antiviral responses to prevent SARS-CoV-2 infection, although final confirmation requires *in vivo/in vitro* experiments and clinical trial phases.

Codon optimization of vaccine nucleotide sequence was performed to ensure effective expression in the *E. coli* host, and the improvement of its GC content was acceptable to the extent of *E. coli*. Restriction enzymes were selected so that the PelB sequence was removed from the vector to provide conditions for intracellular expression during placement of the vaccine in the pET26b(+) expression vector to simulate *in silico*. In addition, for proper purification, the HisTag sequence was considered in the C-terminal of the vaccine to purify the recombinant protein with the Ni Sepharose column. Immunization simulation studies confirmed that the vaccine, after injection, was able to elicit specific immune responses needed to combat the antigen. In this study, the most antigenic epitopes binding to the MHC-I subtype class, MHC-II, and linear/conformational of four known S, M, N, and E proteins of SARS-CoV-2 were used to reach maximum immunogenicity. Although *in silico* studies of the vaccine require entry into the clinical phase for confirmation, the predictions of immunoinformatics tools confirmed the ability of the presented vaccine in this study to induce a humoral and cellular immune response.

Conclusion

Combat to pandemic diseases such as SARS-CoV-2 required early preventative strategies. Although self-quarantine methods and rapid COVID-19 diagnoses are a priority, along with the search for therapeutic solutions, the need for universal vaccination to break the pandemic chain 19 is essential. Recombinant multi-epitope vaccines are an available and viable alternative to combat the SARS-CoV-2 virus due to their presence of several immunogenic factors against target antigens as well as the simultaneous stimulation of humoral and cellular immune responses. Immunoinformatics tools, due to their ease of use and reliability, provide us with an accurate, cost-effective, and fast prediction for designing the best vaccines, without any risk factor keep in touch (virus). In this study, we presented a reverse vaccinology approach using immunoinformatics tools to design an *in silico* vaccine. Immunogenicity responses prediction results suggest that this vaccine is a candidate for the SARS-CoV-2 vaccine.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00251-021-01228-3>.

Author contribution Conceptualization, E.J and Gh.A.J; methodology, E.J, Gh.A.J, M.N and AM; formal analysis, E.J and Gh.A.J; resources, M.N and AM; data curation, E.J, Gh.A.J, M.N and AM; writing—original draft preparation, E.J, Gh.A.J, M.N and AM; writing—review and editing, E.J and Gh.A.J; project administration, E.J and Gh.A.J; funding acquisition, E.J and Gh.A.J. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing Interest The authors declare no competing interests.

References

- Abdulla F, Nain Z, Hossain MM, Syed SB, Ahmed Khan MS, Adhikari UK (2021) A comprehensive screening of the whole proteome of hantavirus and designing a multi-epitope subunit vaccine for cross-protection against hantavirus: Structural vaccinology and immunoinformatics study. *Microb Pathogen* 150:104705
- Ahmad S, Navid A, Farid R, Abbas G, Ahmad F, Zaman N, Parvaiz N, Azam SS (2020) Design of a novel multi epitope-based vaccine for pandemic coronavirus disease (COVID-19) by vaccinomics and probable prevention strategy against avenging zoonotics. *Euro J Pharmaceu Sci* 151:105387
- Anand K, Ziebuhr J, Wadhvani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300:1763–1767
- Behmard E, Soleymani B, Najafi A, Barzegari E (2020) Immunoinformatic design of a COVID-19 subunit vaccine using entire structural immunogenic epitopes of SARS-CoV-2. *Sci Rep* 10:20864
- Bergmann CC, Lane TE, Stohlman SA (2006) Coronavirus infection of the central nervous system: host-virus stand-off. *Nat Rev Microbiol* 4:121–132
- Bonavia A, Zelus BD, Wentworth DE, Talbot PJ, Holmes KV (2003) Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. *J Virol* 77:2530–2538
- 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* 7:153
- Chamani J, Heshmati M (2008) Mechanism for stabilization of the molten globule state of papain by sodium n-alkyl sulfates: Spectroscopic and calorimetric approaches. *J Colloid Interface Sci* 322:119–127
- Choudhary S, Malik YS, Tomar S (2020) Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using *in silico* structure-based virtual screening approach. *Front Immunol* 11:1664–1664
- Chow KT, Gale M, Loo Y-M (2018) RIG-I and other RNA sensors in antiviral immunity. *Annu Rev Immunol* 36:667–694
- Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, Zimmerman MI, Vithani N, Griffith D, Wagoner JA, Bowman GR, Hall KB, Soranno A, Holehouse AS (2021) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* 12:1936
- Cucinotta D, Vanelli M (2020) WHO declares COVID-19 a pandemic. *Acta Biomed* 91:157–160
- Danesh N, Navaee Sedighi Z, Beigoli S, Sharifi-Rad A, Saberi MR, Chamani J (2018) Determining the binding site and binding affinity

- of estradiol to human serum albumin and holo-transferrin: fluorescence spectroscopic, isothermal titration calorimetry and molecular modeling approaches. *J Biomol Struct Dyn* 36:1747–1763
- De Groot AS, Moise L, Terry F, Gutierrez AH, Hindocha P, Richard G, Hoft DF, Ross TM, Noe AR, Takahashi Y, Kotraiah V, Silk SE, Nielsen CM, Minassian AM, Ashfield R, Ardito M, Draper SJ, Martin WD (2020) Better epitope discovery, precision immune engineering, and accelerated vaccine design using immunoinformatics tools. *Front Immunol* 11:442
- de Groot NG, Bontrop RE (2020) COVID-19 pandemic: is a gender-defined dosage effect responsible for the high mortality rate among males? *Immunogenetics* 72:275–277
- de Haan CA, Rottier PJ (2005) Molecular interactions in the assembly of coronaviruses. *Adv Virus Res* 64:165–230
- de Wit E, van Doremalen N, Falzarano D, Munster VJ (2016) SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14:523–534
- Dehghani Sani F, Shakibapour N, Beigoli S, Sadeghian H, Hosainzadeh M, Chamani J (2018) Changes in binding affinity between ofloxacin and calf thymus DNA in the presence of histone H1: Spectroscopic and molecular modeling investigations. *J Lumin* 203:599–608
- Derebail VK, Falk RJ (2020) ANCA-associated vasculitis — refining therapy with plasma exchange and glucocorticoids. *N Engl J Med* 382:671–673
- Enjuanes L, Zuñiga S, Castaño-Rodríguez C, Gutierrez-Alvarez J, Canton J, Sola I (2016) Molecular basis of coronavirus virulence and vaccine development. *Adv Virus Res* 96:245–286
- Fallerini C, Daga S, Mantovani S, Benetti E, Picchiotti N, Francisci D, Paciosi F, Schiaroli E, Baldassarri M, Fava F, Palmieri M, Ludovisi S, Castelli F, Quiros-Roldan E, Vaghi M, Rusconi S, Siano M, Bandini M, Spiga O, Capitani K, Furini S, Mari F, Study G-CM, Renieri A, Mondelli MU, Frullanti E (2021) Association of Toll-like receptor 7 variants with life-threatening COVID-19 disease in males: findings from a nested case-control study. *eLife* 10:e67569
- Gupta N, Kumar A (2020) Designing an efficient multi-epitope vaccine against *Campylobacter jejuni* using immunoinformatics and reverse vaccinology approach. *Microb Pathogen* 147:104398
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181:271–280.e8
- Hu Y, Wen J, Tang L, Zhang H, Zhang X, Li Y, Wang J, Han Y, Li G, Shi J, Tian X, Jiang F, Zhao X, Wang J, Liu S, Zeng C, Wang J, Yang H (2003) The M protein of SARS-CoV: basic structural and immunological properties. *Genomics Proteomics Bioinformatics* 1:118–130
- HuiE JA, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, McHugh TD, Memish ZA, Drosten C, Zumla A, Petersen E, DS (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int J Infect Dis* 91:264–266
- Iwasaki A, Medzhitov R (2010) Regulation of adaptive immunity by the innate immune system. *Science* 327:291–295
- Jangra S, De Vrieze J, Choi A, Rathnasinghe R, Laghlali G, Uvyn A, Van Herck S, Nuhn L, Deswarte K, Zhong Z, Sanders NN, Lienenklaus S, David SA, Strohmeier S, Amanat F, Krammer F, Hammad H, Lambrecht BN, Coughlan L, García-Sastre A, De Geest BG, Schotsaert M (2021) Sterilizing immunity against SARS-CoV-2 infection in mice by a single-shot and lipid amphiphile imidazoquinoline TLR7/8 agonist-adjuvanted recombinant spike protein vaccine **. *Angew Chem Int Ed* 60:9467–9473
- Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, Srivastava AP (2020) A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep* 10:10895
- Kavoosi M, Creagh AL, Kilburn DG, Haynes CA (2007) Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. *Biotechnol Bioeng* 98:599–610
- Kawasaki T, Kawai T (2014) Toll-like receptor signaling pathways. *Front Immunol* 5
- Keith P, Day M, Perkins L, Moyer L, Hewitt K, Wells A (2020) A novel treatment approach to the novel coronavirus: an argument for the use of therapeutic plasma exchange for fulminant COVID-19. *Crit Care* 24:128
- Khairkhan N, Aghasadeghi MR, Namvar A, Bolhassani A (2020) Design of novel multiepitope constructs-based peptide vaccine against the structural S, N and M proteins of human COVID-19 using immunoinformatics analysis. *PLoS One* 15:e0240577
- Khanmohammadi S, Rezaei N (2021) Role of Toll-like receptors in the pathogenesis of COVID-19. *J Med Virol* 93:2735–2739
- Krishnan G S, Joshi A, Akhtar N, Kaushik V (2021) Immunoinformatics designed T cell multi epitope dengue peptide vaccine derived from non structural proteome. *Microb Pathogen* 150:104728
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395:565–574
- Malik YS, Sircar S, Bhat S, Sharun K, Dhama K (2020) Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments. 40:68–76
- McBride R, van Zyl M, Fielding BC (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* 6:2991–3018
- Mokaberi P, Babayan-Mashhadi F, Amir Tehrani Zadeh Z, Saberi MR, Chamani J (2021) Analysis of the interaction behavior between nano-Curcumin and two human serum proteins: combining spectroscopy and molecular stimulation to understand protein-protein interaction. *J Biomol Struct Dyn* 39:3358–3377
- Mukherjee S, Tworowski D, Detroja R, Mukherjee SB, Frenkel-Morgenstern M (2020) Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets. *Vaccines (Basel)* 8
- Muthamani K, Falzarano D, Reuschel EL, Tingey C, Flingai S, Villarreal DO, Wise M, Patel A, Izmirly A, Aljuaid A, Seliga AM, Soule G, Morrow M, Kraynyak KA, Khan AS, Scott DP, Feldmann F, LaCasse R, Meade-White K, Okumura A, Ugen KE, Sardesai NY, Kim JJ, Kobinger G, Feldmann H, Weiner DB (2015) A synthetic consensus anti-spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates. *Sci Translation Med* 7:301ra132
- Narayanan K, Kim KH, Makino S (2003) Characterization of N protein self-association in coronavirus ribonucleoprotein complexes. *Virus Res* 98:131–140
- Nicholson KG, Kent J, Ireland DC (1993) Respiratory viruses and exacerbations of asthma in adults. *BMJ* 307:982–986
- O'Neill LAJ, Golenbock D, Bowie AG (2013) The history of Toll-like receptors — redefining innate immunity. *Nat Rev Immunol* 13:453–460
- Onofrio L, Caraglia M, Facchini G, Margherita V, Placido SD, Buonerba C (2020) Toll-like receptors and COVID-19: a two-faced story with an exciting ending. *Future Sci OA* 6:FSO605
- Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z (2020) Characterization of spike glycoprotein of SARS-CoV-2

- on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 11:1620
- Perlman S, Netland J (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7:439–450
- Poulas K, Farsalinos K, Zanidis C (2020) Activation of TLR7 and innate immunity as an efficient method against COVID-19 pandemic: imiquimod as a potential therapy. *Front Immunol* 11
- Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, Rawling M, Savory E, Stebbing J (2020) Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* 395:e30–e31
- Safaei S, Karimi-Googheri M (2020) Letter to the Editor: Toll-like receptor antagonists as a potential therapeutic strategy against cytokine storm in COVID-19-infected patients. *Viral Immunol* 34:361–362
- Sallenave J-M, Guillot L (2020) Innate immune signaling and proteolytic pathways in the resolution or exacerbation of SARS-CoV-2 in Covid-19: key therapeutic targets? *Front Immunol* 11
- Salvi V, Nguyen HO, Sozio F, Schioppa T, Gaudenzi C, Laffranchi M, Scapini P, Passari M, Barbazza I, Tiberio L, Tamassia N, Garlanda C, Del Prete A, Cassatella MA, Mantovani A, Sozzani S, Bosisio D (2021) SARS-CoV-2-associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI Insight*
- Sanjuán R, Domingo-Calap P (2016) Mechanisms of viral mutation. *Cell Mol Life Sci* 73:4433–4448
- Satarker S, Nampoothiri M (2020) Structural proteins in severe acute respiratory syndrome coronavirus-2. *Arch Med Res* 51:482–491
- Sette A, Fikes J (2003) Epitope-based vaccines: an update on epitope identification, vaccine design and delivery. *Curr Opin Immunol* 15:461–470
- Sheahan TP, Sims AC, Leist SR, Schäfer A, Won J, Brown AJ, Montgomery SA (2020) Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV. 11:222
- Shen N, Fu Q, Deng Y, Qian X, Zhao J, Kaufman KM, Wu YL, Yu CY, Tang Y, Chen JY, Yang W, Wong M, Kawasaki A, Tsuchiya N, Sumida T, Kawaguchi Y, Howe HS, Mok MY, Bang SY, Liu FL, Chang DM, Takasaki Y, Hashimoto H, Harley JB, Guthridge JM, Grossman JM, Cantor RM, Song YW, Bae SC, Chen S, Hahn BH, Lau YL, Tsao BP (2010) Sex-specific association of X-linked Toll-like receptor 7 (TLR7) with male systemic lupus erythematosus. *Proc Natl Acad Sci U S A* 107:15838–15843
- Skwarczynski M, Toth I (2016) Peptide-Based Synthetic Vaccines *Chem Sci* 7:842–854
- Sohrabi C, Alsafi Z, O’Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 76:71–76
- Solanich X, Vargas-Parra G, van der Made CI, Simons A, Schuurs-Hoeijmakers J, Antolí A, del Valle J, Rocamora-Blanch G, Setién F, Esteller M, van Reijmersdal SV, Riera-Mestre A, Sabater-Riera J, Capellá G, van de Veerdonk FL, van der Hoven B, Corbella X, Hoischen A, Lázaro C (2021) Genetic screening for TLR7 variants in young and previously healthy men with severe COVID-19. *Front Immunol* 12
- Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF (2016) Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* 24:490–502
- Le Thanh T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, Mayhew S (2020) The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 19:305–306
- Thomas S (2020) The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter SemiSWEET. *Pathogens & Immunity* 5:342–363
- Tian X, Li C, Huang A, Xia S, Lu S, Shi Z (2020) Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. 9:382–385
- Tseng C-T, Sbrana E, Iwata-Yoshikawa N, Newman PC, Garron T, Atmar RL, Peters CJ, Couch RB (2012) Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS Virus. *PLoS One* 7:e35421
- Wang J, Tricoche N, Du L, Hunter M, Zhan B, Goud G, Didier ES, Liu J, Lu L, Marx PA, Jiang S, Lustigman S (2012) 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
- Wang L, Shi W, Joyce MG, Modjarrad K, Zhang Y, Leung K, Lees CR, Zhou T, Yassine HM, Kanekiyo M, Yang Z-y, Chen X, Becker MM, Freeman M, Vogel L, Johnson JC, Olinger G, Todd JP, Bagci U, Solomon J, Mollura DJ, Hensley L, Jahrling P, Denison MR, Rao SS, Subbarao K, Kwong PD, Mascola JR, Kong W-P, Graham BS (2015) Evaluation of candidate vaccine approaches for MERS-CoV. *Nat Commun* 6:7712
- Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. 30:269–271
- Wille-Reece U, Flynn BJ, Loré K, Koup RA, Kedl RM, Mattapallil JJ, Weiss WR, Roederer M, Seder RA (2005a) HIV Gag protein conjugated to a Toll-like receptor 7/8 agonist improves the magnitude and quality of Th1 and CD8+ T cell responses in nonhuman primates. *Proc Natl Acad Sci U S A* 102:15190–15194
- Wille-Reece U, Wu CY, Flynn BJ, Kedl RM, Seder RA (2005b) Immunization with HIV-1 Gag protein conjugated to a TLR7/8 agonist results in the generation of HIV-1 Gag-specific Th1 and CD8+ T cell responses. *J Immunol* 174:7676–7683
- Wong SK, Li W, Moore MJ, Choe H, Farzan M (2004) A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2 *. *J Biol Chem* 279:3197–3201
- Zhang C, Maruggi G, Shan H, Li J (2019) Advances in mRNA vaccines for infectious diseases. *Front Immunol* 10:594
- Zhang N, Channappanavar R, Ma C, Wang L, Tang J, Garron T, Tao X, Tasneem S, Lu L, Tseng C-TK, Zhou Y, Perlman S, Jiang S, Du L (2016) Identification of an ideal adjuvant for receptor-binding domain-based subunit vaccines against Middle East respiratory syndrome coronavirus. *Cell Mol Immunol* 13:180–190
- Zheng Y, Zhuang M-W, Han L, Zhang J, Nan M-L, Zhan P, Kang D, Liu X, Gao C, Wang P-H (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) membrane (M) protein inhibits type I and III interferon production by targeting RIG-I/MDA-5 signaling. *Signal Transduct Target Ther* 5:299
- Ziebuhr J (2005) The coronavirus replicase. *Curr Top Microbiol Immunol* 287:57–94

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