



OPEN

## Immunoinformatics design of a structural proteins driven multi-epitope candidate vaccine against different SARS-CoV-2 variants based on fynomer

Javad Sarvmeili<sup>1</sup>, Bahram Baghban Kohnehrouz<sup>1✉</sup>, Ashraf Gholizadeh<sup>2</sup>, Dariush Shanehbandi<sup>3</sup> & Hamideh Ofoghi<sup>4</sup>

The ideal vaccines for combating diseases that may emerge in the future require more than simply inactivating a few pathogenic strains. This study aims to provide a peptide-based multi-epitope vaccine effective against various severe acute respiratory syndrome coronavirus 2 strains. To design the vaccine, a library of peptides from the spike, nucleocapsid, membrane, and envelope structural proteins of various strains was prepared. Then, the final vaccine structure was optimized using the fully protected epitopes and the fynomer scaffold. Using bioinformatics tools, the antigenicity, allergenicity, toxicity, physicochemical properties, population coverage, and secondary and three-dimensional structures of the vaccine candidate were evaluated. The bioinformatic analyses confirmed the high quality of the vaccine. According to further investigations, this structure is similar to native protein and there is a stable and strong interaction between vaccine and receptors. Based on molecular dynamics simulation, structural compactness and stability in binding were also observed. In addition, the immune simulation showed that the vaccine can stimulate immune responses similar to real conditions. Finally, codon optimization and in silico cloning confirmed efficient expression in *Escherichia coli*. In conclusion, the fynomer-based vaccine can be considered as a new style in designing and updating vaccines to protect against coronavirus disease.

### Abbreviations

CPORT	Consensus prediction of Interface residues in transient complexes
HADDOCK	High ambiguity driven protein–protein docking
MDS	Molecular dynamics simulation
OPLS	Optimised potential for liquid simulation
PRODIGY	Protein binding energy prediction
Rg	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
rpL	50S ribosomal protein L7/L12
SH3	Src homology 3
SMM	Stabilized matrix method
VOC	Variant of concern

Since the end of 2019, humanity has seen the spread of the contagious disease COVID-19 (2019 CORonaVirus Disease) caused by SARS-CoV-2 and its different variants<sup>1</sup>. Most countries have been involved in this disease to some extent regarding health, social, and economic issues. Also, some pharmaceutical and non-pharmaceutical

<sup>1</sup>Department of Plant Breeding and Biotechnology, University of Tabriz, Tabriz 51666, Iran. <sup>2</sup>Department of Animal Biology, University of Tabriz, Tabriz 51666, Iran. <sup>3</sup>Department of Immunology, Tabriz University of Medical Sciences, Tabriz 51666, Iran. <sup>4</sup>Department of Biotechnology, Iranian Research Organization for Science and Technology, Tehran 33131, Iran. ✉email: bahramrouz@yahoo.com

companies have spent significant time and money to combat this disease<sup>2</sup>. The World Health Organization (WHO) has classified this pandemic as an emergency health problem and an ongoing global health crisis due to its capacity to spread rapidly and the multiple alterations of the virus that have resulted in variants of concern (VOCs). Following the issuance of emergency licenses for the production and distribution of vaccines, the organization declared that over 150 countries had joined the Global COVID-19 Immunization Access Initiative (COVAX) to address the issue of rapid and equitable access to vaccines around the world<sup>3-5</sup>. Producing and developing high-efficacy vaccines is the most effective approach to reducing disease burden, controlling its spread, and achieving eradication<sup>6</sup>. However, ongoing research is also being conducted on various aspects of the virus and its associated disease, encompassing the understanding of disease characteristics, mutations, variants, prevention strategies, treatment methods, vaccine efficacy, immune evasion, and the roles of neutralizing and binding antibodies.

The large enveloped virus SARS-CoV-2 with a single-stranded RNA genome prone to rapid mutation with a positive sense of approximately 29.9 kb in length, is a member of the beta-coronavirus belonging to the *Coronaviridae* that can cause infection and pneumonia in birds, mammals, and humans<sup>7,8</sup>. This virus is composed of four structural proteins: spike (S), nucleocapsid (N), membrane (M), and envelope (E), as well as several non-structural proteins (nsp)<sup>9</sup>. The availability of genomic and proteomic sequences has made it possible to obtain more genetic information, identify genes, and develop vaccines for this disease. Researchers propose using vaccines as a consistent and effective approach that could provide long-term protection against all virulent strains<sup>10</sup>. This study aims to develop an epitope-based vaccine that activates innate and acquired immunity. Antigenicity is attributed to the viral structural proteins, especially the S protein<sup>11</sup>; therefore, many epitopes have been identified for this protein's T and B-cell epitopes. The importance of the S protein in infection and pathogenesis makes it an ideal target and focus for SARS-CoV-2 vaccine design and development, as it is the leading viral antigen neutralized by antibodies during infection<sup>7</sup>. Epitopes located on structural proteins are part of the antigenic determinants that can initiate cellular immune responses and are recognized by specific receptors on the surface of T or B-cells<sup>12,13</sup>. Accordingly, identifying and studying epitopes is essential for developing diagnostic methods and epitope-based vaccines<sup>14</sup>.

The presence of advanced *in silico* immunoinformatics tools by predicting and screening candidate epitopes and designing effective vaccines against diseases reduces the burden of experimental immunity on the model organism<sup>15</sup>. These tools offer a precise, rapid, cost-effective approach to developing potential multi-epitope vaccines against infectious diseases<sup>16</sup>. On the other hand, instead of using complete proteins and attenuated pathogens, which can cause sensitization and other immune reactions such as local redness, pain swelling, or increased body temperature after vaccination, designing a multi-epitope vaccine will be more reliable<sup>16</sup>. Furthermore, this approach has several advantages, including engineering epitopes for greater potency, allowing immune responses to conserved epitopes, and improving immune profiles<sup>17</sup>, which apply to different SARS-CoV-2 strains.

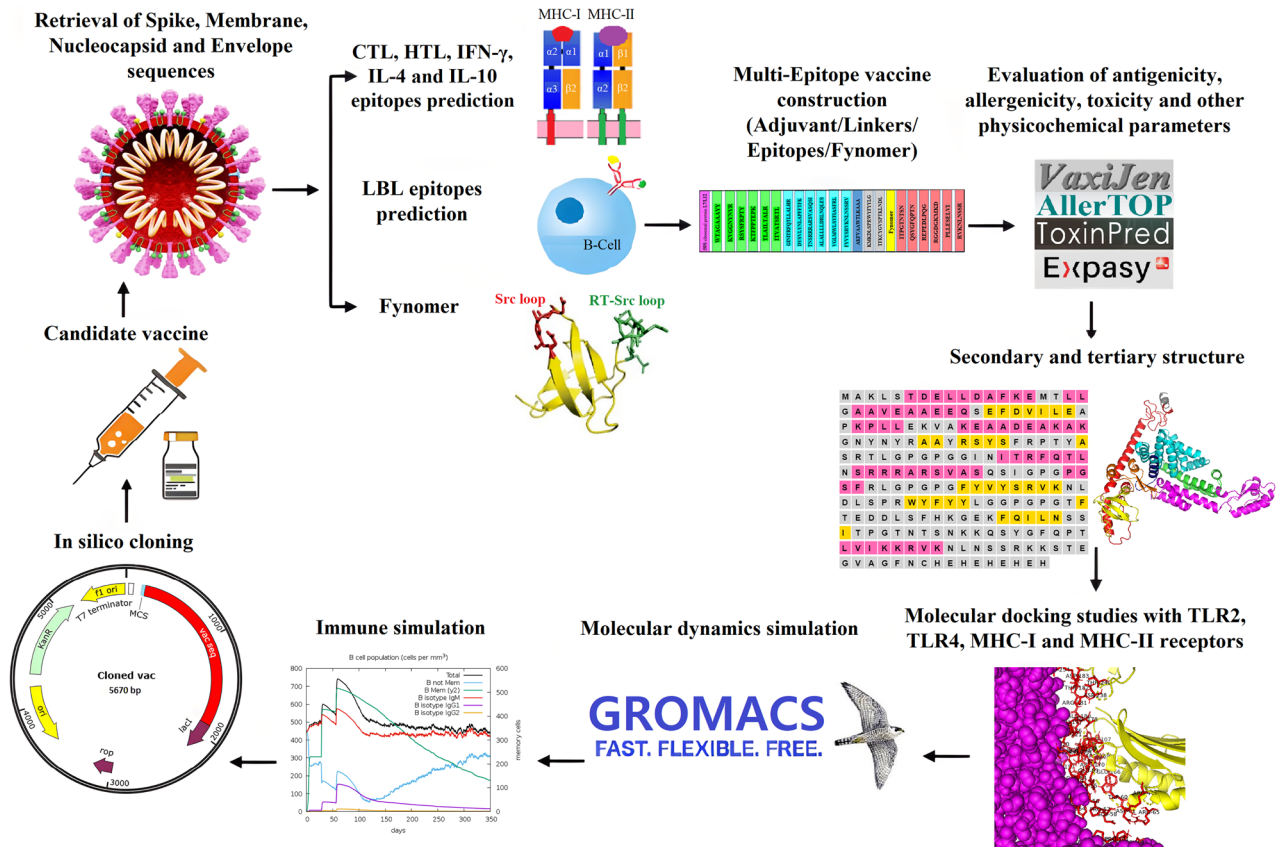
In this study, epitopes of the spike, nucleocapsid, envelope, and membrane structural proteins of 32 different variants of the SARS-CoV-2 virus, which are antigenic, have been used as targets for T and B-cells in the design of potential vaccine structure. Furthermore, to enhance therapeutic efficacy and stabilize the multi-epitope state of the structure, the example of different types of antibody mimetics known as fynomer has also been proposed for the first time. This led to the combination of the chosen epitopes around the fynomer scaffold. Fynomers are small globular binding proteins with a molecular weight of 7 kDa, composed of amino acids 83 to 145 of the Src homology 3 (SH3) domain of the human tyrosine-protein kinase Fyn<sup>18</sup>. Their structure contains two antiparallel  $\beta$ -sheets and two flexible loops called RT and Src loops, essential in interactions with other proteins<sup>19</sup>. Advantages associated with fynomer scaffolds include fast and easy purification, no tendency to form aggregates and being monomeric, high stability with a melting temperature ( $T_m$ )  $\sim 70$  °C, lac of cysteine residue that allows combining with other proteins without misfolding problems, expressing in *Escherichia coli* at a high level in soluble form, having a human origin and the existence of a completely conserved amino acid sequence between humans and mice, which indicates that it is non-immunogenic in mammals<sup>20</sup>. Based on these features, it can be claimed that fynomer not only has the potential to be engineered to produce non-hemoglobin proteins with specific binding domains and high affinity to target pathogens specifically but can also be incorporated into the structure of vaccines to increase their stability and effectiveness.

In order to understand safety and effectiveness, the physicochemical, immunogenic, and dynamic characteristics of the designed vaccine were examined after obtaining the final structure. Finally, *in silico* cloning and immune simulations were also carried out (Fig. 1). Vaccination is the most effective way to prevent infectious diseases, which generally calls for the use of thorough yet precise methods. As it is crucial for the design of multi-epitope vaccines to monitor genetic and antigenic changes in the viral population globally to deal with emerging variants, this study, with consideration for changes in protein structures, suggests a candidate vaccine with the capacity to elicit humoral and cellular immune responses to deal with broad variants.

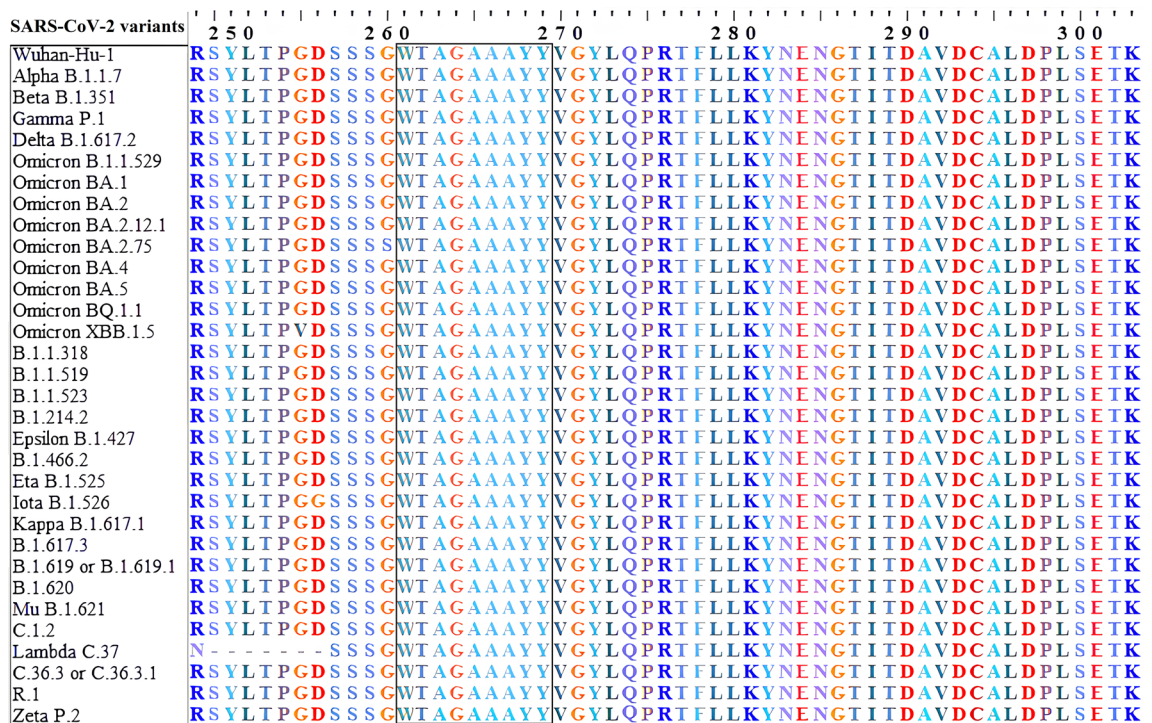
## Results

### Prediction of antigenicity and alignment of protein sequences

According to the antigenicity screening of the amino acid sequences of four structural proteins (Spike, Nucleocapsid, Membrane, and Envelope) from 32 SARS-CoV-2 strains (Supplementary Tables S1–S3), the S protein had the most excellent antigenic value, followed by proteins E, M, and N, in that order. It was also determined that these proteins could serve as desirable antigens. Hence, these proteins were selected to predict T and B-cell epitopes and vaccine development. After examining the results of sequence alignment, protective and non-protective regions of these structural proteins in the desired strains were identified by the identity of the amino acid sequences (Fig. 2). In contrast to the other three proteins, spike protein has effective mutations, particularly in the RBD (Receptor binding domain) and NTD (N-terminal domain) regions, which are effective in the



**Figure 1.** Schematic workflow of in silico multi-epitope vaccine design process.



**Figure 2.** Multiple sequence alignment (MSA) for structural proteins of 32 SARS-CoV-2 variants using ClustalW. One predicted epitope was highlighted with a black box line in the conserved region.

severity of pathogenicity, antibody escape, and disease transmission<sup>21</sup>. To enhance the efficacy of the designed vaccine against wild-type and mutant SARS-CoV-2, it is suggested that the vaccine design strategy be based on the identification and selection of epitopes that are recognized by both predictive tools as stimulating T and B lymphocytes in the immune system and are among the non-mutated and fully conserved parts of the virus. This approach ensures the vaccine targets parts of the virus that are less likely to undergo mutations and remain effective against different strains. However, it is essential to note that the immune response can be complex and may involve factors beyond the predicted conservancy of epitopes. Therefore, it is necessary to confirm the effectiveness of antigenically protected epitopes (ability to stimulate an immune response) and provide complete protection through experimental evidence<sup>22</sup>.

### Prediction of T-cell epitopes: CTL and HTL

Pathogen or vaccine peptides in the cytosol undergo fragmentation by the proteasome, which unfolds the protein and cleaves it into peptide fragments. They are transported to the endoplasmic reticulum by the protein transporter TAP. Only peptides of the right size stably bind to the MHC I groove<sup>23</sup>. The NetCTL 1.2 server assesses these features, evaluating the proteasomal cleavage, TAP transport efficiency, and MHC class I binding. Higher predicted values for these indicators increase the likelihood of peptide positioning in the MHC-I groove, facilitating CTL recognition<sup>24</sup>. NetCTL1.2 predicted 440 epitopes for S protein, 104 for N, 126 for M, and 51 for E protein. These epitopes showed binding solid affinity to multiple alleles within the 12 MHC class I supertypes. All these epitopes were included among the results obtained from IEDB as selected items by considering all MHC alleles available in this server for these structural proteins. On the other hand, the NetMHC II pan 3.2 server, which was used to predict HTL epitopes, predicted 1605 epitopes for S, 359 for N, 169 for M, and 54 for E. After checking, it was found that IEDB also suggested these same epitopes as top epitopes for the alleles that are in this server, which is a kind of confirmation of the predicted epitopes through both servers. Among all these predicted cytotoxic T lymphocytes (CTLs) and helper T lymphocytes (HTLs) epitopes, the best ones were selected based on antigenicity, allergenicity, toxicity, and most importantly, 100% conservation in all SARS-CoV-2 strains. According to these criteria, 39 CTL epitopes of S protein, 14 of N, 20 of M, and 13 of E were selected. Additionally, 52 HTL epitopes of S, 28 of N, 19 of M, and 11 of E were screened (see Supplementary Tables S4–S11). Through the use of IFNepitopes, IL4pred, and IL10pred servers, an investigation was conducted to determine the inducibility of interferon- $\gamma$  (IFN- $\gamma$ ), interleukin-4 (IL-4), and Interleukin-10 (IL-10). The results revealed 4 epitopes in S, 1 in N, and 1 in M, all meeting the necessary criteria. It should be noted that no epitopes meeting these criteria were identified in E. In this study, to achieve the highest immunogenicity and reduce epitope overload in the final vaccine construct, only one epitope from the S and N proteins with the highest scores for the investigated criteria was selected (Supplementary Table S12). The T-cell epitopes used in the final vaccine formulation are listed in Tables 1, 2 and 3. Supplementary Fig. S1 shows the conformational positions of these epitopes on each structural protein derived from it.

### Prediction of linear and conformational B-cell epitopes

After carefully examining the results of 5 linear B-cell (LBL) epitope prediction servers and considering the similarities in their results, 27 epitopes were predicted for S protein, 14 for N, 12 for M, and 6 for E. 12 linear B-cell epitopes of S, 3 of N, 3 of M and 2 epitopes of E were obtained. They were chosen based on the binding score, antigenicity, allergenicity, toxicity, flexibility, hydrophilicity, accessibility on the surface, and complete conservation criteria (Table 4 and Supplementary Table S13). The positions of these LBL epitopes in the vaccine structure are shown in Supplementary Fig. S2. 9 epitopes were selected from these epitopes to be part of the final vaccine structure. Based on the results obtained from the ElliPro server, 3 B-cell conformational epitopes were also identified from the refined Three-dimensional (3D) model, which amino acid residues, sequence location, number of residues, and their scores are shown in Table 5. In Supplementary Fig. S3, a graphic representation of the 3D model of these epitopes is presented.

### Analysis of human population coverage around the world, protection assessment, and auto-immunity identification

Population coverage analysis revealed that 12 T-cell epitopes selected in this study represented 96.77% of the worldwide human population. The percentage coverage of epitopes in other regions is also presented separately in Table 6 and Supplementary Fig. S4. Therefore, the designed multi-epitope vaccine can be recommended to combat SARS-CoV-2 in most regions (Supplementary Table S14). In this study, North America had the highest coverage (97.97%), while Central America had the least (36.29%). The low population coverage of these epitopes in Central America can be related to the fact that the populations of these regions are a complex array of different admixture processes with varying degrees of ancestral population proportions that came in different migration waves. So, results from population genetics comparisons show a wide variation in the HLA profiles from the populations that correlate with different admixture proportions<sup>25</sup>. Since Central America exhibits significant genetic diversity among its populations if a peptide vaccine is designed based on specific epitopes and these epitopes are not familiar or diverse in the genetic composition of Central American populations, it could lead to epitope coverage decrease. According to the IEDB conservation analysis tool results, as expected, selected T and B-cell epitopes in regions without alteration in the studied structural protein sequence showed high conservation among different variants (Supplementary Fig. S5). In addition, all selected epitopes were non-homologous with the human proteome, indicating the multi-epitope vaccine can induce a safe antigenic response but no cross-reaction with proteins in humans.

Proteins/epitopes (position)	Supertypes/HLA class I alleles	Immunogenicity score	Antigenicity/allergenicity	Toxicity	Epitope conservation (%)	Homology to human proteins
Spike/WTAGAAAY (258–266)	A1, A26, B58, B62/ HLA-A*68:23, HLA-B*15:17, HLA-C*03:03, HLA-A*26:02, HLA-A*29:02, HLA-A*32:15, HLA-B*40:13, HLA-C*12:03, HLA-A*30:02, HLA-B*15:02, HLA-A*26:01, HLA-A*32:07, HLA-B*35:01, HLA-B*27:20, HLA-A*68:01, HLA-C*14:02	0.1525	0.6306 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Spike/KVGGNYNYR (442–450)	A3/ HLA-A*68:23, HLA-A*31:01, HLA-A*32:07, HLA-A*02:50, HLA-C*12:03, HLA-B*27:20, HLA-B*40:13, HLA-A*32:15, HLA-C*03:03	0.0277	1.5212 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike/RSYSFRPTY (490–498)	A1, A3, B27, B58, B62/ HLA-B*15:17, HLA-B*15:03, HLA-A*32:07, HLA-C*12:03, HLA-C*03:03, HLA-A*68:23, HLA-B*27:20, HLA-B*40:13, HLA-A*30:02, HLA-A*32:01, HLA-B*58:01, HLA-A*30:01, HLA-B*15:01, HLA-A*02:17, HLA-C*07:02	0.0083	0.9553 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Nucleocapsid/ KTFPPTEPK (361–369)	A3/ HLA-A*11:01, HLA-A*68:23, HLA-A*30:01, HLA-B*27:20, HLA-C*12:03, HLA-B*40:13, HLA-A*32:07, HLA-C*14:02, HLA-A*03:01, HLA-A*31:01, HLA-A*68:01, HLA-C*03:03, HLA-A*32:01	0.1306	0.9553 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Membrane/ITVATSRTL (168–176)	B62/ HLA-C*03:03, HLA-A*68:23, HLA-B*15:17, HLA-A*32:07, HLA-A*02:50, HLA-C*15:02, HLA-B*15:02, HLA-C*12:03, HLA-B*27:20, HLA-A*02:17, HLA-C*14:02, HLA-B*40:13, HLA-A*02:11	0.1306	0.7571 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Envelope/TLAILTALR (30–38)	A3/ HLA-B*27:20, HLA-A*68:01, HLA-A*32:07, HLA-A*02:50, HLA-A*68:23, HLA-A*31:01, HLA-C*12:03, HLA-A*32:15, HLA-C*03:03, HLA-A*33:01, HLA-B*40:13	0.1989	0.7223 (antigen)/non-allergen	Non-toxin	100%	Non-homolog

**Table 1.** The selected CTL epitopes for the final vaccine construction are provided by the NetCTL and IEDB servers.

### Making the final multi-epitope vaccine

The structure of the linear vaccine includes epitopes CTL, HTL, IFN- $\gamma$ , IL-4, IL-10, LBL, Pan HLA-DR reactive epitope (PADRE), 50S ribosomal protein L7/L12 (rpIL) adjuvant, fynomer sequence, linkers (EAAAK, AAY, GPGPG, and KK) and finally the H5E tag. Figure 3A illustrates the general outline of this structure.

### Evaluation of antigenicity, allergenicity, toxicity, solubility, and physicochemical properties of the vaccine

Antigenicity, allergenicity, toxicity, and solubility were also assessed for the final sequence of the designed vaccine, and it met all the criteria. The vaccine's antigenicity was confirmed by Vaxijen v2.0 servers<sup>26</sup> with a score of 0.6387 and ANTIGENpro<sup>27</sup> with a score of 0.9238, indicating the final vaccine had a high antigenicity. AllergenFP 1.0 and AllerTOP 2.0 servers<sup>28,29</sup> showed that the final vaccine is non-allergenic and does not induce allergic reactions. Based on the results of the SVM prediction mode in the ToxinPred server<sup>30</sup>, the entire final vaccine sequence, which includes the adjuvant sequence, all epitopes, linkers, fynomer sequence, and H5E tag, does not contain any toxin part. The values predicted by SolPro<sup>31</sup> (0.7368) and Protein-sol server<sup>32</sup> (0.496) also show that the vaccine has good solubility (Table 7). Solubility is a critical factor in vaccine development because it influences

Proteins/epitopes (position)	HLA class II alleles	Antigenicity/allergenicity	Toxicity	Epitope conservation (%)	Homology to human proteins
Spike/GINITRFQTLALHR (232–246)	HLA-DRB5*01:01, HLA-DRB1*04:04, HLA-DRB1*01:01, HLA-DRB1*11:01, HLA-DRB1*04:05, HLA-DRB1*04:01, HLA-DRB1*15:01, HLA-DRB4*01:01, HLA-DRB1*07:01, HLA-DRB1*09:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01/DPB1*04:01	0.5582 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Spike/TNSRRRARSVASQSI (676–690)	HLA-DRB1*07:01, HLA-DRB1*01:01	0.7267 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Spike/DYSVLYNLAPFFTFK (361–375)	HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DRB1*08:03, HLA-DRB1*12:01, HLA-DRB1*13:02, HLA-DRB3*02:02, HLA-DRB3*01:01, HLA-DPA1*01:03/DPB1*04:01, HLA-DPA1*01:03/DPB1*04:02, HLA-DPA1*01:03/DPB1*23:01, HLA-DQA1*01:02/DQB1*06:04, HLA-DPA1*01:03/DPB1*02:01	0.7838 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Nucleocapsid/ALALLLLDRLNQLES (218–232)	HLA-DRB4*01:01, HLA-DRB1*03:01, HLA-DRB1*11:01, HLA-DRB1*04:04, HLA-DRB1*01:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01	0.5057 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Membrane/VGLMWLSYFIAS-FRL (88–102)	HLA-DRB5*01:01, HLA-DRB1*04:04, HLA-DRB1*01:01, HLA-DRB1*15:01, HLA-DRB1*04:05, HLA-DRB1*07:01, HLA-DRB1*04:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01/DPB1*04:01	0.6658 (antigen)/non-allergen	Non-toxin	100%	Non-homolog
Envelope/FYVYSRVKLNLSRV (56–70)	HLA-DRB1*01:01, HLA-DRB1*04:04, HLA-DRB1*07:01, HLA-DRB1*11:01, HLA-DRB1*04:01, HLA-DRB1*04:05, HLA-DRB1*09:01, HLA-DRB1*13:02, HLA-DRB1*15:01	0.6103 (antigen)/non-allergen	Non-toxin	100%	Non-homolog

**Table 2.** The selected HTL epitopes for the final vaccine construction are provided by the NetMHC II pan 3.2 and IEDB server.

Proteins/epitopes (position)	IFN- $\gamma$	IL-4	IL-10	Antigenicity/allergenicity	Toxicity	Epitope conservation (%)	Homology to human proteins
Spike/TFKCYGVSPTKLNLDL (25–39)	Positive	Inducer	Inducer	1.4626 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Nucleocapsid/KMKDLSRPWFYFYYLG (100–114)	Positive	Inducer	Inducer	1.4297 (antigen)/non-allergen	Non-toxin	100	Non-homolog

**Table 3.** The selected IFN- $\gamma$ , IL-4, and IL-10 epitopes in the final vaccine construct, predicted by IFNepitopes, IL4pred and IL10pred servers.

Proteins	Epitopes (position)	Antigenicity/allergenicity	Toxicity	Epitope conservation (%)	Homology to human proteins
Spike	IPTGTNTSN (598–606)	1.0158 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike	QSYGFQPTN (491–499)	1.4068 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike	REPEDLPQG (208–216)	0.7997 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike	STEIQAGNCYFP (469–480)	0.6932 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike	IGSKPCNGVEGFN (487–499)	0.4402 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Spike	NKPCNGVAGFNC (474–485)	0.5067 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Nucleocapsid	RGGDGKMKD (95–103)	0.8805 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Membrane	PLLESELVI (132–140)	0.5354 (antigen)/non-allergen	Non-toxin	100	Non-homolog
Envelope	RVKLNLSRV (61–69)	0.8998 (antigen)/non-allergen	Non-toxin	100	Non-homolog

**Table 4.** Details of linear/continuous B-cell (LBL) epitopes in the vaccine construct.

No	Residue	Number of residues	Score
1	_:A522, _:G523, _:N524, _:C525, _:Y526, _:F527, _:P528, _:K529, _:K530, _:I531, _:G532, _:S533, _:K534, _:P535, _:C536, _:E540, _:G541, _:F542, _:N543, _:K544, _:K545, _:N546, _:K547, _:P548, _:C549, _:N550, _:A553, _:G554, _:F555, _:N556, _:C557, _:H558, _:E559, _:H560, _:E561, _:H562, _:E563, _:H564, _:E565, _:H566	40	0.9
2	_:L67, _:A69, _:A70, _:G71, _:D72, _:K73, _:K74, _:I75, _:G76, _:V77, _:I78, _:K79, _:V80, _:V81, _:R82, _:E83, _:I84, _:V85, _:S86, _:G87, _:L88, _:G89, _:L90, _:K91, _:E92, _:A93, _:K94, _:D95, _:L96, _:V97, _:D98, _:G99, _:A100, _:P101, _:K102, _:L104, _:D114, _:E115, _:A116, _:K117, _:A118, _:K119, _:L120, _:E121, _:A122, _:A123, _:G124, _:A125, _:T126, _:V127	50	0.888
3	_:Y395, _:D396, _:Y397, _:E398, _:A399, _:R400, _:T401, _:E402, _:D403, _:D404, _:L405, _:S406, _:F407, _:H408, _:K409, _:E411, _:T430, _:T431, _:G432, _:E433, _:T434, _:G435, _:Y436, _:N470, _:R473, _:E474, _:P475, _:E476, _:D477, _:L478, _:P479, _:Q480, _:G481, _:K482, _:R484	35	0.847

**Table 5.** List of conformational/discontinuous B-cell epitopes predicted over final vaccine construct.

Population/area	Class combined		
	Coverage <sup>a</sup>	Average hit <sup>b</sup>	pc90 <sup>c</sup>
World	96.77%	6.18	2.05
East Asia	96.58%	6.13	2.12
Northeast Asia	94.78%	4.66	1.41
South Asia	96.9%	6.04	2.01
Southeast Asia	91.84%	3.89	1.12
Southwest Asia	86.48%	3.71	0.74
Europe	97.87%	6.88	2.5
East Africa	90.14%	3.74	1.01
West Africa	91.73%	4.03	1.15
Central Africa	88.91%	3.59	0.9
North Africa	94.15%	5.04	1.49
South Africa	92.14%	3.28	1.12
West Indies	91.4%	4.11	1.12
North America	97.97%	6.48	2.61
Central America	36.29%	1.3	0.16
South America	86.76%	4.08	0.76
Oceania	89.12%	3.92	0.92
Average	89.4	4.53	1.36
Standard deviation	13.76	1.39	0.65

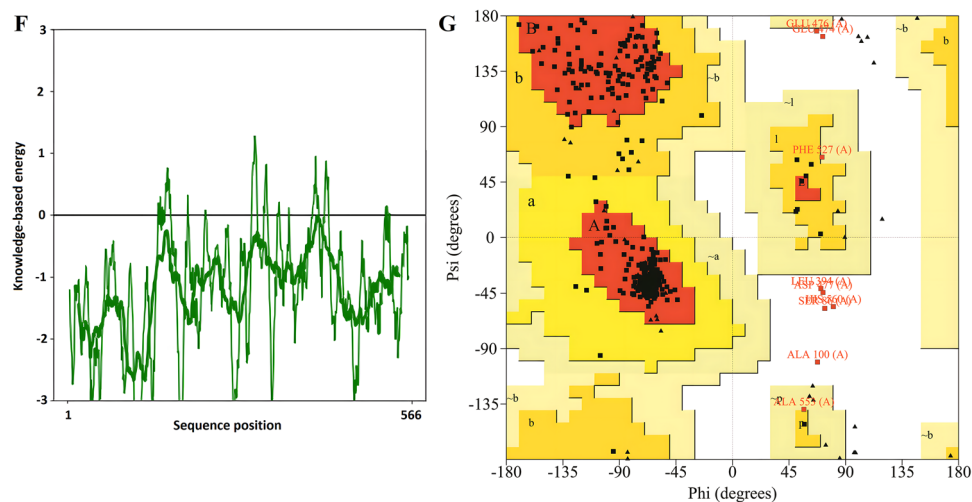
**Table 6.** Population coverage of the selected epitope included in the vaccine construct. <sup>a</sup>Coverage of population on projected. <sup>b</sup>Population recognized by HLA combinations/epitope hits on the average number. <sup>c</sup>90% of the population recognized by HLA combinations/epitope hits on the minimum number.

the distribution, stability, efficacy, vaccine formulation, administration, adsorption, manufacturing process, storage conditions, bioavailability of the vaccine components, and overall success of vaccination programs<sup>33</sup>.

According to the results of the evaluation of the physicochemical properties of the vaccine, the final structure has 566 residues and a molecular weight of 60.77 kDa. This structure has a theoretical isoelectric point (pI) of 9.36, indicating that it is essential. There were 56 negatively charged residues (Asp + Glu) and 75 positively charged residues (Arg + Lys). The charge is 19.00, which decreases in an alkaline environment, and it is generally preferable to have positive charge values. The coefficient of extinction is 71,740 M<sup>-1</sup> cm<sup>-1</sup>, with absorption values (0.1%) (g/l) 1185, including all cysteine pairs under aqueous conditions at 280 nm. In addition, the estimated half-life of the vaccine in mammalian reticulocytes (in vitro) is 30 h, in yeast (in vivo) approximately 20 h, and in *E. coli* (in vivo) 10 h, which indicates the stability of the vaccine in different phases. The half-life of a vaccine varies across different organisms, indicating its stability and effectiveness in different biological environments. A shorter half-life may require more frequent administration, affecting the dosing schedule and practicality. However, a longer half-life can offer prolonged protection but may pose safety concerns or require rapid adaptation to emerging viral variants. The optimal half-life depends on the vaccine's nature, target pathogen, and practical administration considerations<sup>34</sup>. The prediction showed that the structure of the vaccine, with a 34.09 (< 40) instability index, has a high degree of stability to elicit an immune response. The aliphatic index of the vaccine is 73.02, which indicates it has a thermostable nature, and when the aliphatic index of a protein increases, it becomes resistant to heat. The overall average hydrophilicity (GRAVY) was -0.368, and lower GRAVY scores indicate better solubility, which suggests that the vaccine candidate is hydrophilic, improving interaction with blood and water and making it more straightforward to identify the target<sup>35</sup>. All the results of the physicochemical properties analysis supported that the vaccine structure meets the criteria for vaccine formulation (Table 7 and







**Figure 3.** (continued)

Final vaccine	Vaxijen score	Antigen pro score	Allergen FP/AllerTOP result	Solubility by SolPro	Solubility by protein-sol	Toxicity	Mol. weight (kDa)
	0.6387	0.9238	Non-allergen	0.7368	0.496	Non-Toxin	60.773
	Theoretical pI	Charge	Half-life (in vitro)	Half-life (in vivo)	Instability index	Aliphatic index	Hydropathicity (GRAVY)
	9.36	19.00	30 h	> 20 h > 10 h	34.09 (stable)	73.02	-0.368

**Table 7.** Antigenicity, allergenicity, solubility, toxicity, and physicochemical properties prediction results for final vaccine construction.

with default parameters including 39.39% alpha-helix, 18.20% extended strand, 5.83% beta-turn, and 42.58% random coil. A graphical representation of the secondary structure features is shown in Fig. 3B. The high percentage of random coil, as can be understood from the figure, indicates the presence of epitopes in different regions of the construct<sup>39</sup>.

### Modeling, refinement, and validation of the 3D structure of multi-epitope vaccine

After creating the 3D vaccine model using the Robetta server<sup>40</sup>, one model was selected as the best prediction model, and based on this model, the 3D structure of the final vaccine was made. Later than refining the 3D structural model obtained from the modeling stage by GalaxyRefine server<sup>41</sup>, model 1 was selected as the best final vaccine model based on various parameters, including GDT-HA (0.9832), RMSD (0.287), MolProbity (1.954), Clash score (15.9) and weak rotamers (0.5) among other refined models (Fig. 3C and Supplementary Table S15). The Z-score for the vaccine structure in the PROSA diagram was -9.78. This score is close to the range of native proteins of similar size, indicating the lowest error rate and accuracy of the simulation and the overall reliability of the predicted model (Fig. 3E, F). In addition, the quality factor of 95.5036 obtained using the ERRAT server indicates the optimal quality of the protein model<sup>42</sup> (Fig. 3D). Grouping of amino acids based on phi and psi angles by Ramachandran plot<sup>43,44</sup> using PROCHECK analysis in the PDBsum server for the refined structure, revealed that 94.6% of the residues were classified in the most favorable, 3.5% additional allowed, 0.4% generously allowed, and 1.5% disallowed region (Fig. 3G). According to the results, the total residuals in the desired area were in the ideal value range, i.e., more than 90%, which confirms the reliability of this model.

### Molecular docking of the vaccine construct with TLR3, TLR4, MHC-I, and II receptors and binding affinity evaluation

The process of protein-protein docking between refined 3D models of the final vaccine with immune receptors TLR2 and TLR4 and MHC-I and II receptors, utilizing the HADDOCK web server 2.4, classifies the structures as refined models with water into multiple clusters by percentage values. The most reliable clusters are those whose complexes have the lowest HADDOCK score<sup>45</sup>. Table 8 displays the statistical parameters and their respective values for each docked vaccine-receptor complex. Supplementary Figs. S9-S12 shows the HADDOCK Refinement interface server's graphical results for each complex. The HADDOCK score, RMSD, and other presented criteria were evaluated for complexes (Table 9). Better connection is indicated by a low HADDOCK score (negative score)<sup>46</sup>. Also, the buried surface area (BSA) indicates the proximity and less surface of the protein exposed to water. On the other hand, the RMSD scores allow identifying the sets with the lowest energy

Complexes	Vac-TLR2	Vac-TLR4	Vac-MHC-I	Vac-MHC-II
HADDOCK score (kJ mol <sup>-1</sup> )	-143.8 ± 3.7	-141.9 ± 3.4	-93.1 ± 11.8	-74.0 ± 6.1
Cluster size (nm)	116	23	8	18
RMSD from the overall lowest energy structure (nm)	2.8 ± 1.6	0.7 ± 0.4	0.6 ± 0.3	25.5 ± 1.3
Van der Waals energy (kJ mol <sup>-1</sup> )	-112.3 ± 3.9	-100.2 ± 5.8	-93.4 ± 10.7	-84.7 ± 7.8
Electrostatic energy (kJ mol <sup>-1</sup> )	-337.2 ± 29.0	-428.1 ± 32.4	-439.5 ± 39.3	-226.9 ± 18.6
Desolvation energy (kJ mol <sup>-1</sup> )	-13.8 ± 2.7	-19.7 ± 6.0	-1.4 ± 3.0	-28.0 ± 5.7
Restraints violation energy (kJ mol <sup>-1</sup> )	498.3 ± 31.7	635.9 ± 30.0	895.1 ± 67.0	840.4 ± 61.2
Buried surface area	3737.9 ± 70.6	3128.7 ± 84.0	3019.9 ± 156.7	2661.8 ± 143.3
Z-score	-2.5	-2.7	-1.9	-1.8

**Table 8.** Statistics of the highest-scored vaccine-TLR2, TLR4, MHC-I, and MHC-II docked clusters. The most reliable cluster was selected based HADDOCK score for each complex. A negative value of the HADDOCK score projects good protein–protein interaction potential.

Complexes	Vac-TLR2	Vac-TLR4	Vac-MHC-I	Vac-MHC-II
HADDOCK score (kJ mol <sup>-1</sup> )	-211.6 ± 2.4	-234.4 ± 2.2	-227.0 ± 3.4	-181.9 ± 2.3
Cluster size (nm)	20	20	20	20
RMSD from the overall lowest energy structure (nm)	0.6 ± 0.4	0.6 ± 0.3	0.6 ± 0.4	0.6 ± 0.4
Van der Waals energy (kJ mol <sup>-1</sup> )	-129.0 ± 2.7	-111.3 ± 7.8	-107.2 ± 8.3	-92.1 ± 5.9
Electrostatic energy (kJ mol <sup>-1</sup> )	-349.7 ± 31.1	-504.1 ± 21.7	-613.6 ± 30.6	-261.0 ± 20.0
Desolvation energy (kJ mol <sup>-1</sup> )	-12.7 ± 3.1	-22.2 ± 2.8	2.9 ± 1.3	-37.6 ± 4.0
Restraints violation energy (kJ mol <sup>-1</sup> )	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Buried surface area	3919.4 ± 85.7	3088.6 ± 24.0	3353.8 ± 30.8	2989.4 ± 87.4
Z-score	0.0	0.0	0.0	0.0

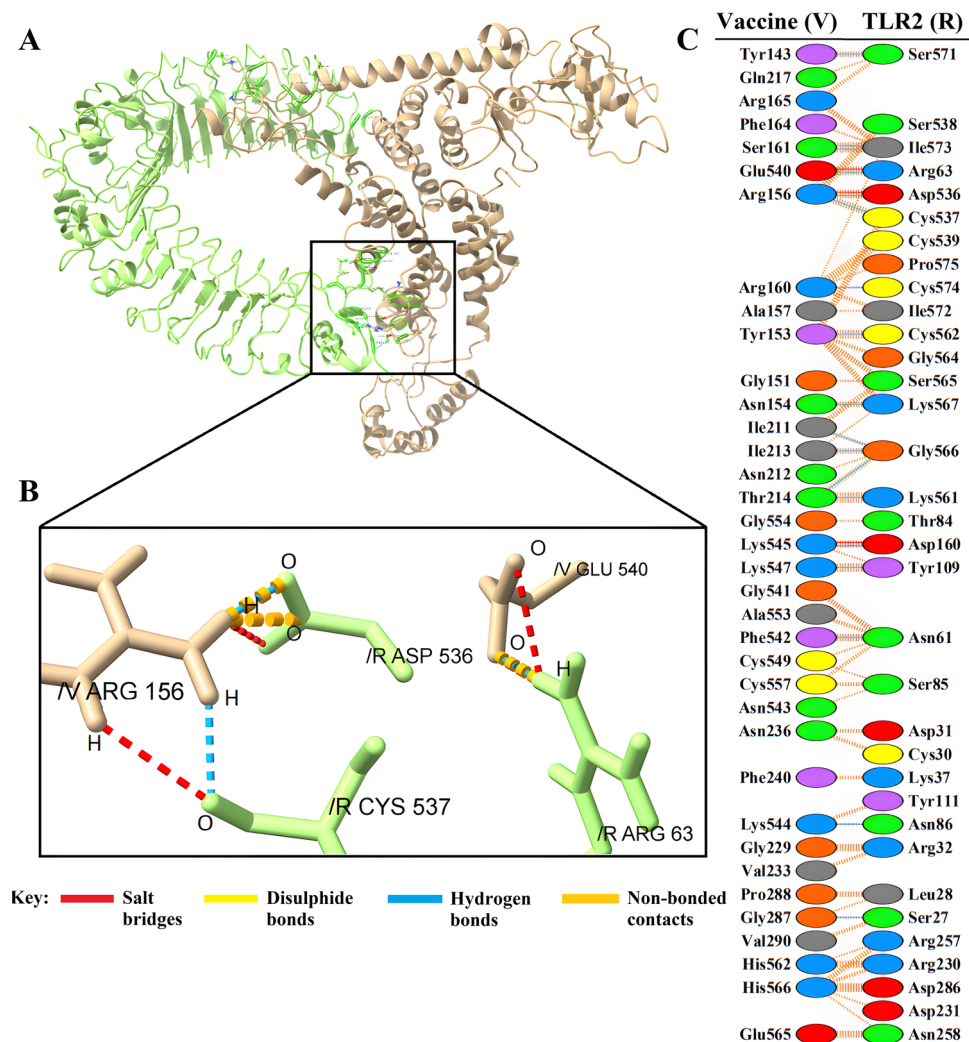
**Table 9.** Statistics of a single cluster were obtained after applying refinements on top-ranked docked vaccine-TLR2, TLR4, MHC-I, and MHC-II structures.

and minimum structural deviation. Furthermore, the low RMSD score for the docked complexes represents a good-quality model.

An overview of molecular docking and the amino acids that directly interact at the binding sites of the vaccine structure and receptors are shown in Figs. 4A, B, 5A, B, 6A, B, 7A, B. Amino acids and, in more detail, their atoms predicted to be involved in vaccine-receptor interactions are presented in Supplementary Materials S2–S5.

According to PDBsum results, there were 36 residues of the vaccine and 34 residues of TLR2 in the complex between the vaccine and TLR2, and the interface area (Å<sup>2</sup>) for the vaccine and TLR2 was 1965 and 1953, respectively. For the vaccine-TLR4 complex, 28 vaccine residues interacted with 31 residues of TLR3, and the interface area was 1552 Å<sup>2</sup> for the vaccine and 1457 Å<sup>2</sup> for TLR4. The interacting residues of the vaccine-MHC-I complex were 33 residues for the vaccine and 34 residues for MHC-I, and the interface region was 1677 Å<sup>2</sup> for the vaccine and 1606 Å<sup>2</sup> for MHC-I. Also, for the vaccine-MHC-II complex, 22 vaccine residues interacted with 26 MHC-II residues, and the interface region for the vaccine was 1426 Å<sup>2</sup>, while this region was 1358 Å<sup>2</sup> for MHC-II. On the other hand, the molecular interaction established between the vaccine with TLR2; 3 salt bridges, 18 hydrogen bonds, and 155 non-bonding contacts; with TLR4, 5 salt bridges, 23 hydrogen bonds, and 187 non-bonding contacts; with MHC-I, 8 salt bridges, 20 hydrogen bonds and 196 non-bonded contacts and finally with MHC-II; there were 2 salt bridges, 9 hydrogen bonds and 113 non-bonded contacts (Figs. 4C, 5C, 6C, and 7C). Hydrogen bonds are essential in molecular recognition because they are important in achieving a stable compound<sup>47</sup>. Electrostatic bonds also play an essential role in the interactions produced in complex<sup>47</sup>. Since, in this study, most of the hydrogen bond distances between vaccine-receptor interacting residues are around 2–3 Å, there is a high interaction between them<sup>48</sup>. In protein–protein complexes, salt bridges can be crucial for maintaining the stability of the interaction interface. Also, they can contribute to the specificity and affinity of the interaction between two proteins and be involved in the recognition and binding of the interacting partners<sup>49</sup>. In addition, the Ramachandran plots presented in the PDBsum results were also investigated for structural validation of the docked sets, which were also confirmatory (Supplementary Figs. S13–S16).

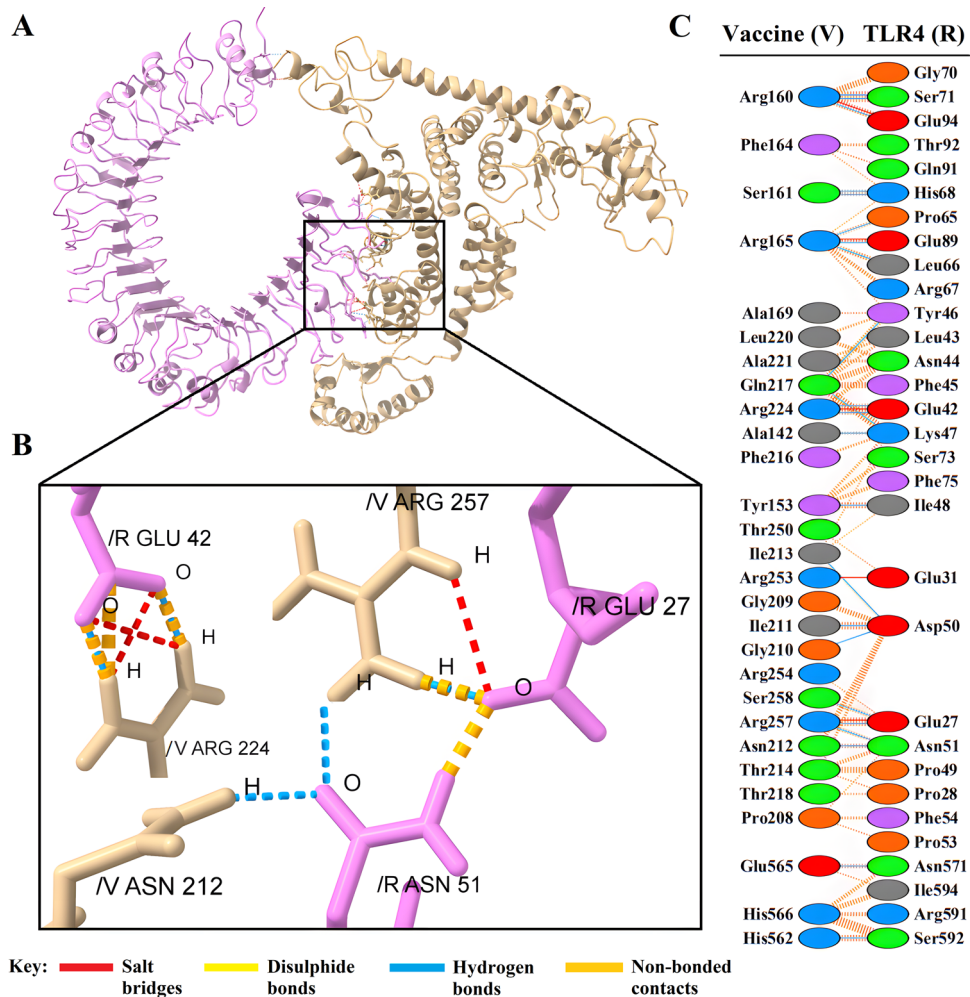
After binding affinity analysis using the PRODIGY (PROtein binDing enerGY prediction) web server, ΔG values (Gibbs free energy) for TLR2-vaccine, TLR4-vaccine, MHC-I-vaccine, and MHC-II-vaccine complex were obtained -18.4, -11.2, -13.7, -9.7 kcal mol<sup>-1</sup> (kilocalories per mole), respectively. The negative values of ΔG indicate that all four docked complexes are energetically feasible. The dissociation constant (Kd) presents binding affinity and bond strength. As binding affinity and Kd share an inverse correlation, a lower Kd signifies a heightened binding affinity. This implies that vaccine-receptor bonds are securely and tightly bound when the dissociation constant is lower (Table 10).



**Figure 4.** (A) Visualization of docking results for the vaccine-TLR2 complex. The vaccine construct is shown in gold, while TLR2 is depicted in green. (B) Magnified residues and their atoms are shown as sticks and labeled with chain, code, and number. Hydrogen bonds, salt bridges, and other interactions were represented by colored dashed lines. (C) Map of total interacting residues and bonds between the vaccine and TLR2 protein chains.

### Energy minimization and molecular dynamics simulation of vaccine structure

Molecular dynamics simulation (MDS), a technique for studying the atomic behavior of molecular systems, was employed in this study to estimate and evaluate the stability and physical movements of the vaccine structure under various conditions<sup>50</sup>. The purpose of molecular dynamics simulation for the designed vaccine is to observe how it functions in a real-life biological system<sup>51</sup>. With the aid of GROMACS software, some analyses, including energy minimization, pressure assessment, temperature, and potential energy calculations, were carried out<sup>52</sup>. Through the steepest descent algorithm, energy minimization was performed for the vaccine, and the energy of the protein was considered minimized when it was less than 1000 kJ mol<sup>-1</sup>. Energy minimization was applied for 2,482 steps, where the force reached less than 1000 kJ mol<sup>-1</sup>. The system's potential energy was computed as -3.06 e5228478 kJ mol<sup>-1</sup> with a total drift of -300,757 kJ mol<sup>-1</sup> and an average potential energy of -3.06 e43701 kJ mol<sup>-1</sup>. The average temperature at the end of 5000 NVT steps was 299.85 K, with a temperature drift of 1.14 K (Fig. 8A). The calculated average density of the system was 1011.18 kg m<sup>-3</sup> with a total drift of 0.154 kg m<sup>-3</sup> (Fig. 8B). The pressure of the system was 0.327 bar with a total drift of 4.57 bar (Fig. 8C). After a simulation time of 10 ns, a trajectory analysis was performed to confirm the flexibility and stability of the candidate vaccine. The RMSD plot illustrates the fluctuations in the overall of the vaccine and conveys its stability over time. During the 10 ns simulation process, the RMSD of the vaccine has relatively mild fluctuations, which indicates the construct is stable (Fig. 8D). The RMSF and particularly the high peaks in the RMSF demonstrate the flexibility in the vaccine structure. The RMSF plot, particularly the high peaks in it, demonstrates the flexibility of the vaccine structure (Fig. 8E). Also, the stability and compactness of the structure during the simulation have been confirmed by the radius of the gyration plot (Rg) (Fig. 8F).



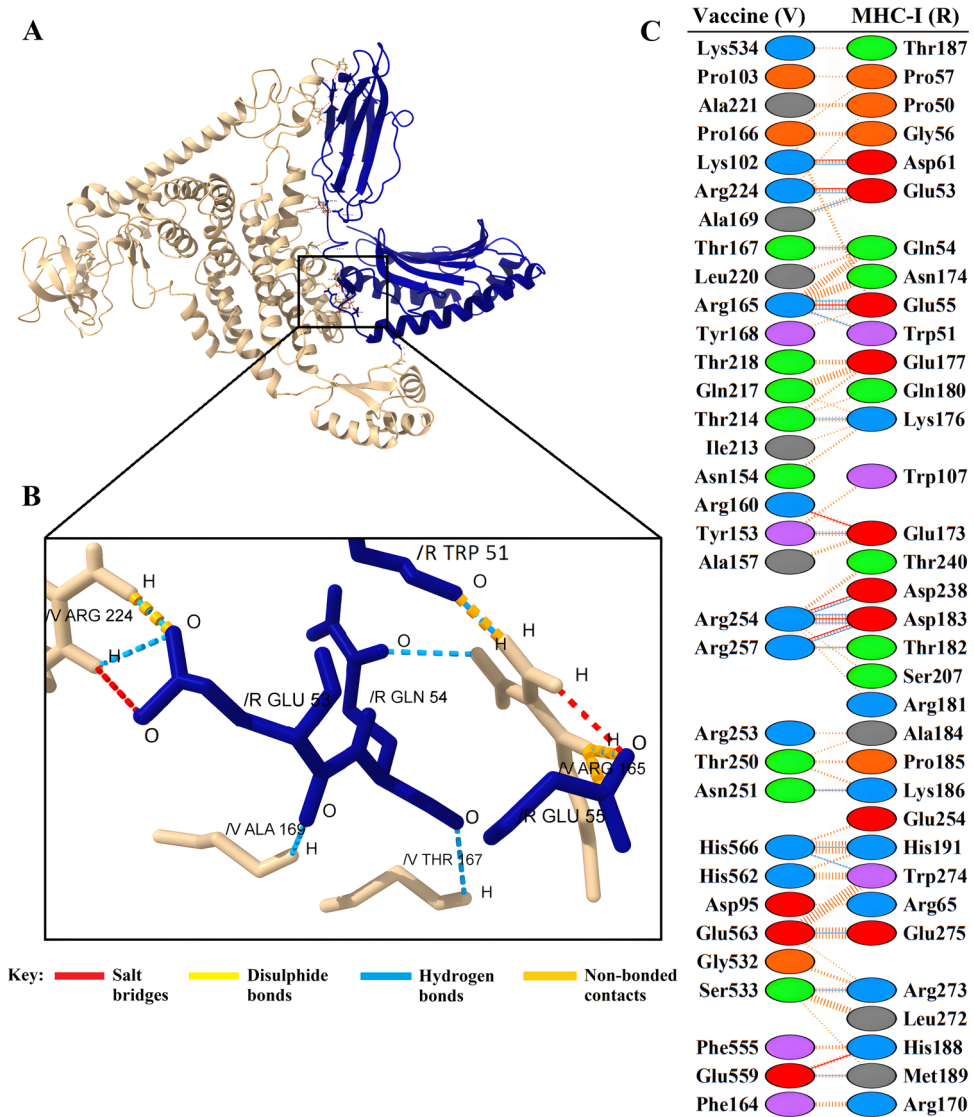
**Figure 5.** (A) Visualization of docking results for the vaccine-TLR4 complex. The vaccine construct is shown in gold, while TLR4 is depicted in violet. (B) Magnified residues and their atoms are shown as sticks and labeled with chain, code, and number. Hydrogen bonds, salt bridges, and other interactions were represented by colored dashed lines. (C) Map of total interacting residues and bonds between the vaccine and TLR4 protein chains.

### Codon optimization and in-silico cloning of vaccine

The candidate vaccine nucleotide sequence was obtained after codon optimization to maximize protein expression in the *E. coli* strain K12 as the expression host organism<sup>53</sup>. For the 1698 nucleotide optimized codon sequence, the values for the guanine-cytosine content (GC-Content) and codon adaptation index (CAI Value) were 0.9335 and 50.73, respectively (Supplementary Material SM6). Generally, a CAI above 0.8 or almost 154 and GC between 30 and 70%<sup>35</sup> indicates good protein expression in the host system. This also supports the practical expression of the vaccine formulated in this strain. Figure 9, the SnapGene software output, shows the cloning of the vaccine sequence inside the pET-28a (+) vector to generate a recombinant plasmid for developing an efficient in silico cloning strategy.

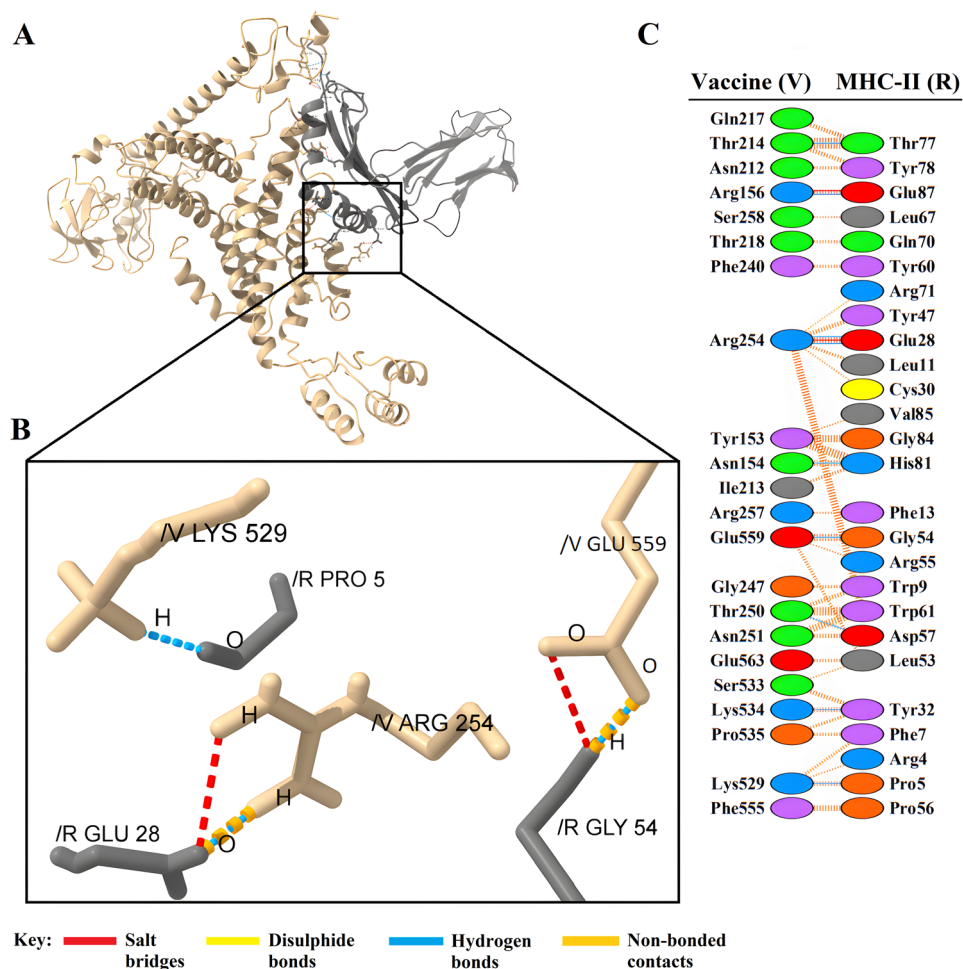
### Immune simulation

Based on immunological simulation data, administration of this vaccine with three injections potentially induces a range of immunoglobulins<sup>54</sup> (Fig. 10). In the initial response, the level of IgM increases, while in the second and third responses produced by the simulation, the levels of IgM + IgG, IgG1 + IgG2, IgG1, and IgG2 were significantly higher than in the first response. Subsequent administration of three vaccine injections resulted in a decrease in the concentration of specific antigens with typically high immunoglobulin concentrations, i.e., IgG1 + IgG2, IgM, and IgG + IgM (Fig. 10A). In addition, some high-stability B-cell isotypes are identified that provide the potential for isotype switching and memory formation (Fig. 10B, C). There was also an increase in the production of CTL/HTL cells and memory cells (TCs) during vaccination, indicating immunogenicity in the presence of T-cell epitopes in the vaccine framework (Fig. 10D–F). In general, according to these results, after the initial activation and proliferation of immune cells, the immune system undergoes a contraction phase, which leads to a decrease in some immunoglobulins, including IgM and IgG, between time intervals after vaccination. This reduction in response does not necessarily indicate vaccine failure, and it is a normal part of the



**Figure 6.** (A) Visualization of docking results for the vaccine-MHC-I complex. The vaccine construct is shown in gold, while MHC-I is depicted in navy. (B) Magnified residues and their atoms are shown as sticks and labeled with chain, code, and number. Hydrogen bonds, salt bridges, and other interactions were represented by colored dashed lines. (C) Map of total interacting residues and bonds between the vaccine and MHC-I protein chains.

immune response. In these conditions, other immune system components, such as memory T and B-cells, play an essential role in maintaining immunity. By creating immunological memory, they enable faster and more efficient responses in re-exposure to future pathogens, which depends on the incubation time of the infection, memory response quality, and memory B-cell antibody levels. Therefore, even if the level of some antibodies decreases after the initial response, the immune system is not defenseless<sup>55</sup>. Also, each time of exposure, there was an increase in natural killer cells (NK cells), dendritic cells (DCs), and macrophage activities (Fig. 10G, H). High levels of IFN- $\gamma$ , IL-23, IL-10, and IL-12 are markers of a favorable immune response, which increased significantly after exposure (Fig. 10I). Finally, immune simulations for 3 and 12 doses confirm that the vaccine induces a robust immune response and that the level of immunity increases even with repeated exposures in pandemic situations. It is important to note that interpreting vaccine efficacy and protection involves a comprehensive assessment of various immune components, including antibodies, memory cells, and other immune responses. Additionally, the specific dynamics of the immune response may vary depending on the vaccine and the individual characteristics of the immune system<sup>56</sup>.



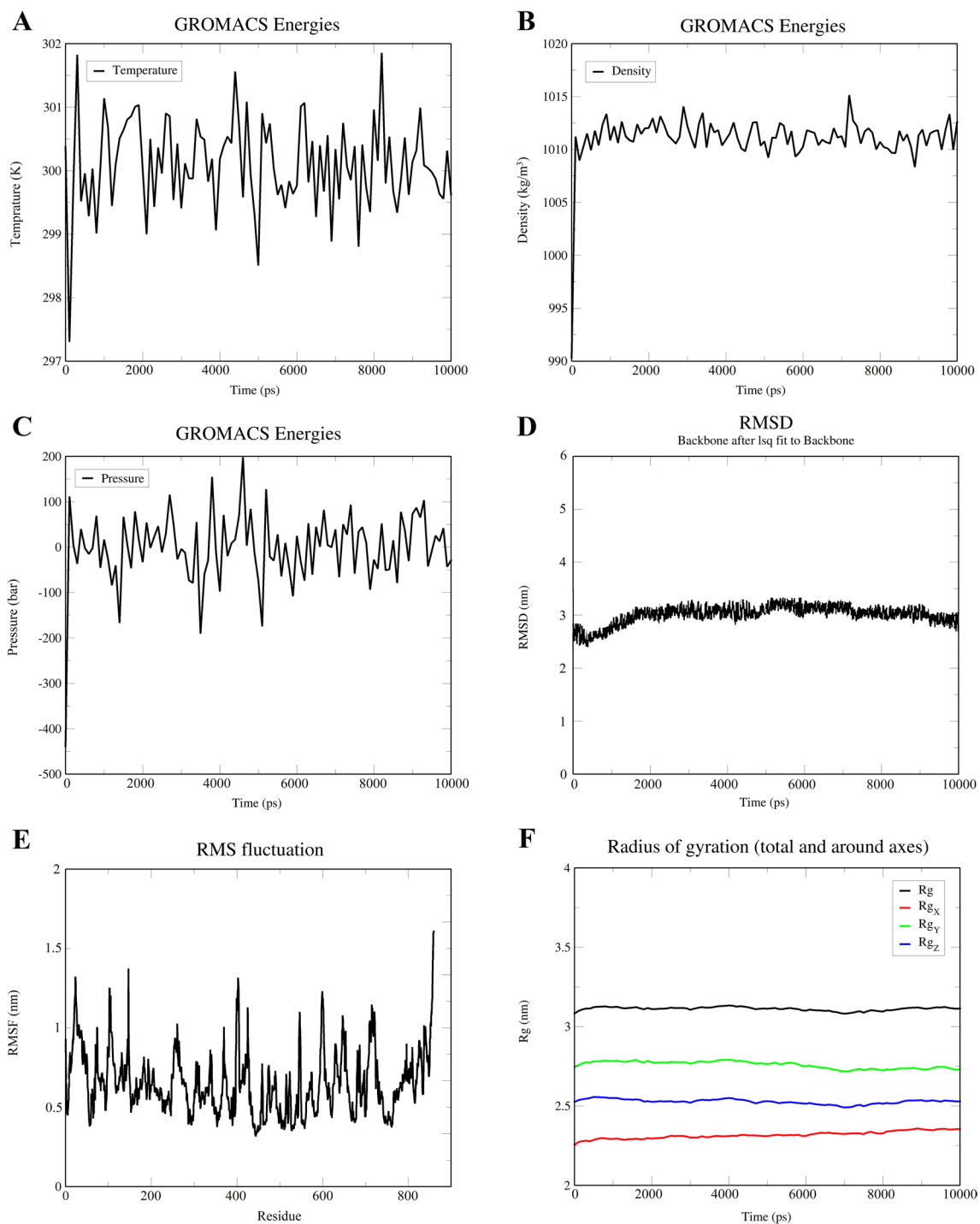
**Figure 7.** (A) Visualization of docking results for the vaccine-MHC-II complex. The vaccine construct is shown in gold, while MHC-II is depicted in gray. (B) Magnified residues and their atoms are shown as sticks and labeled with chain, code, and number. Hydrogen bonds, salt bridges, and other interactions were represented by colored dashed lines. (C) Map of total interacting residues and bonds between the vaccine and MHC-II protein chains.

Complexes	Gibbs free energy (kcal mol <sup>-1</sup> )	Kd (M)
Vaccine-TLR2	-18.4	1.1E-13
Vaccine-TLR4	-13.7	2.2E-10
Vaccine-MHC class I receptor	-11.8	4.5E-09
Vaccine-MHC class II receptor	-9.7	1.3E-07

**Table 10.** Binding affinities of the docked complexes of the vaccine with TLR2, TLR4, MHC-I and MHC-II, as predicted by the PRODIGY server.

## Discussion

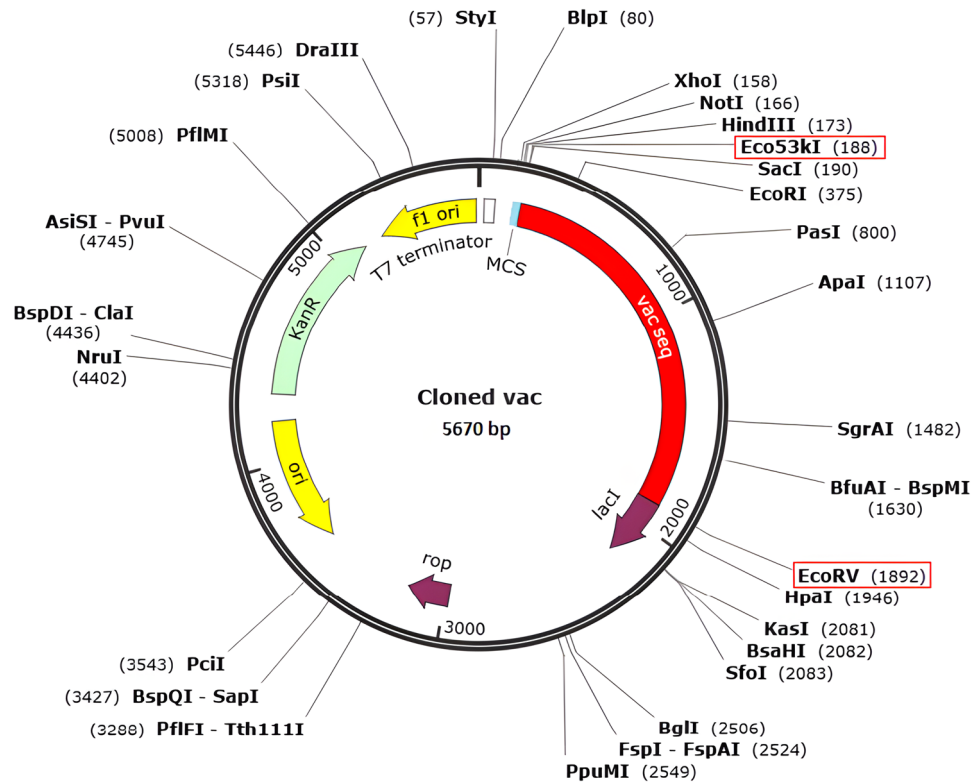
The SARS-CoV-2 pandemic has raised many questions about human weakness and susceptibility to this pathogen and viral diseases that may emerge. COVID-19 has inspired researchers to develop effective drugs and up-to-date vaccines to help control the epidemic because it threatens public health<sup>57</sup>. Despite the chance that new variants could emerge as a result of genetic mutations in the SARS-CoV-2 genome, it is evident that there is a lot more research that needs to be done on potential candidates. However, new vaccine candidates may have fewer or milder side effects, which would satisfy those concerned about the severity of side effects from other vaccines<sup>58</sup>. Using live or attenuated viral vaccines has several drawbacks, like the risk of reversion to a more acute strain, being expensive, being time-consuming, causing allergic reactions, and having other destructive immunological responses<sup>59</sup>. So, ensuring you are getting a safe and effective vaccine is essential. Designing epitope-based vaccines



**Figure 8.** The results of molecular dynamics simulation of vaccine for analysis of structural stability. **(A)** Graph showing the equilibrated temperature during energy minimization. **(B)** Graphical presentation of density during simulation. **(C)** Graph showing the pressure of the system during simulation. **(D)** RMSD plot of the vaccine construct indicating stability. **(E)** RMSF plot illustrates high fluctuations, the peak-like regions with a higher degree of flexibility. **(F)** The Rg plot showing the vaccine construct stays compact around its axes, supporting its stability during simulation.

in silico alongside immunoinformatics approaches is highly efficient because it enables researchers to quickly predict and accurately evaluate potential antigenic epitopes while saving time and money<sup>47</sup>.

We designed a multi-epitope vaccine based on a fynomer antibody mimetic that can induce a robust immune response against SARS-CoV-2 and is a candidate to prevent the COVID-19 pandemic. With the help of computational tools and molecular modeling, it is also possible to generate humoral and cellular immunity as immune responses for a set of short immunogenic sequences, i.e., epitopes, instead of large proteins or whole genomes

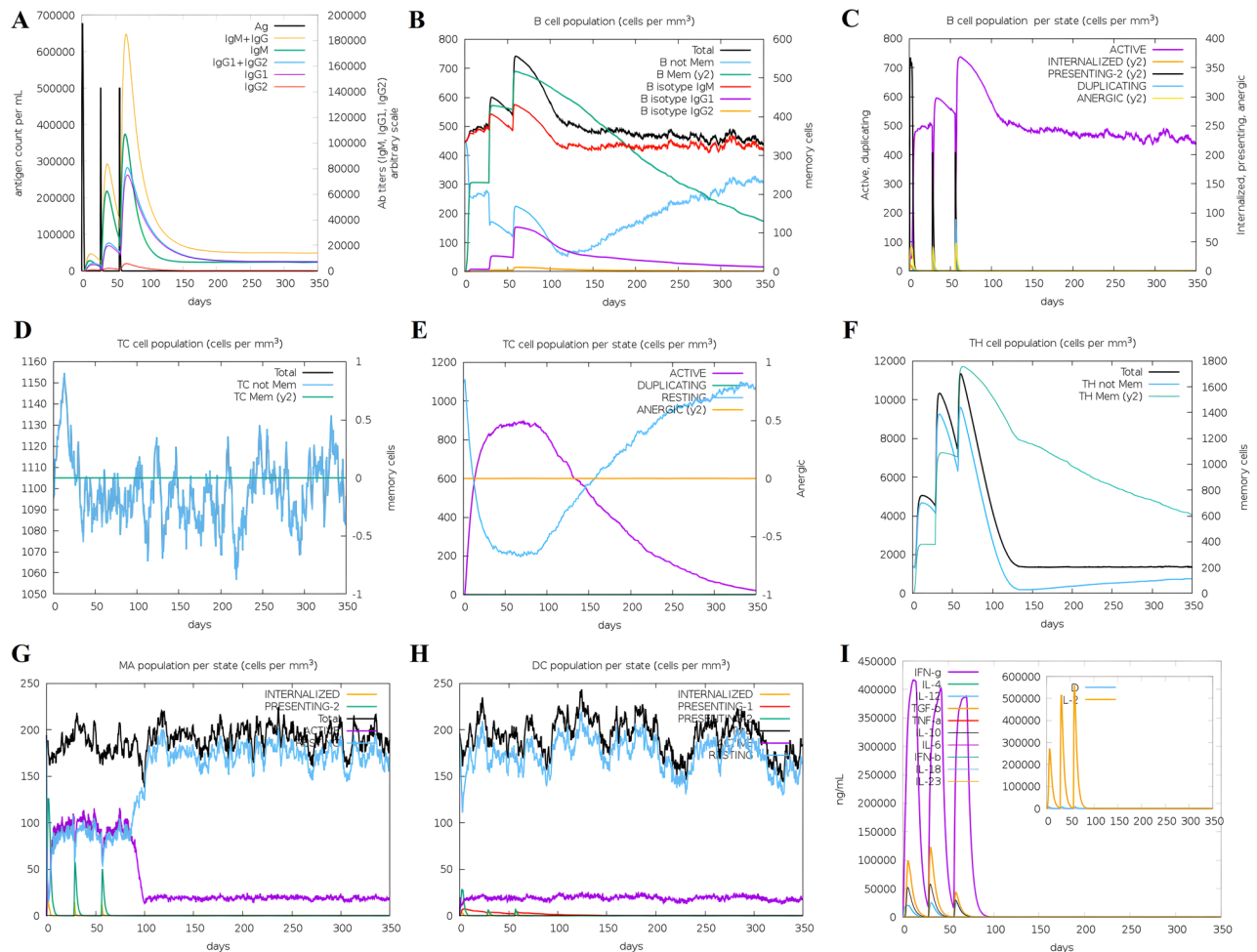


**Figure 9.** In silico restriction cloning of the designed vaccine into the pET-28a (+) expression vector. The red bar represents the codon-optimized gene of the vaccine, and the black circle represents the vector backbone.

that are typically employed in recombinant vaccine technology<sup>60,61</sup>. It is important to note that the specific design of a vaccine depends on the characteristics of the target virus, the goals of the vaccination strategy, and ongoing research findings. Researchers and vaccine developers carefully consider these factors when creating vaccines that offer the best protection against the targeted pathogen<sup>62</sup>. This study has provided a comprehensive and improved understanding of the design of a vaccine against the coronavirus due to the utilization of four structural proteins from various coronavirus strains to screen for T and B-cell epitopes. The decision to include additional proteins in a vaccine design alongside the spike protein is often strategic based on various considerations. While the spike protein is a significant target for immune response due to its high antigenicity, other reasons, including broad protection, stability, efficacy, enhanced immune response, reducing escape mutants, and pre-existing immunity, are considered for the inclusion of other components in the vaccine<sup>63</sup>. Including multiple proteins in a vaccine offers several advantages. It broadens the immune response, providing potential protection against various virus variants. This strategy safeguards against mutations in the spike protein by ensuring other components can still trigger an immune response. Different proteins may offer enhanced stability, production feasibility, or immunogenicity, optimizing vaccine efficacy. Diversifying antigens stimulates various aspects of the immune system, potentially yielding a more robust and comprehensive response for improved protection. Including other proteins reduces the risk of the virus developing escape mutants, enhancing immune detection. Additionally, accounting for pre-existing immunity to the spike protein by incorporating other proteins can boost overall vaccine effectiveness<sup>64</sup>. The epitopes derived from the structural proteins used in the analysis of T lymphocytes tended to interact with a variety of MHC-I and MHC-II alleles. Epitopes were prioritized in conservation, antigenicity, allergenicity, non-toxicity, comprehensive coverage in human populations, lack of homology with human proteins, and effective molecular interaction with the relevant HLA alleles. They were selected as candidate CTL and HTL epitopes for the vaccine (Tables 1, 2 and 3). Additionally, predictive tools were employed to evaluate B-cell epitopes for linearity, surface accessibility, antigenicity, flexibility, allergenicity, and toxicity (Table 4).

The final product of this study was a vaccine with a 566 amino acid length that included 6 CTL epitopes, 6 HTL epitopes, 2 stimulating epitopes IFN- $\gamma$ , IL-4 and IL-10, 9 LBL epitopes, fynomer sequence with 61 amino acids, 50S ribosomal protein L7/L12 adjuvant, PADRE sequence and an H5E tag (Fig. 3A). With the aid of the appropriate tools, the vaccine structure was also assessed for antigenicity, sensitivity, toxicity, physicochemical properties, solubility, and the presence or absence of transmembrane helices (TMHs) and signal peptides<sup>36</sup>. All of the scores in different parameters for this proposed construct increase the likelihood that this vaccine can be shown to be safe and effective against SARS-CoV-2 and confirm its applicability (Table 7). The second and third structures of the multi-epitope construct play a critical role in vaccine formulation, and one of the fundamental challenges of structural biology is to identify discrepancies between experimental and theoretical predictions of protein structures<sup>65</sup>. Secondary structure analysis showed that the vaccine structure includes alpha helices,





**Figure 10.** In silico simulation of immune response triggered by the designed vaccine as an antigen after three subsequent injections. (A) Antigen and subtypes of immunoglobulin levels are represented as different colored peaks. The Immunoglobulin (IgG) production represents the proliferation of primary, secondary, and tertiary immune responses after the vaccine administration. (B,C) B lymphocytes by total count and population per entity state (active, presenting, internalized, duplicating, or anergic). (D) Cytotoxic T-cell population. (E) Cytotoxic T-cell population per state. (F) Helper T-cell population. (G) Macrophages population per state. (H) Dendritic cell population per state. (I) The concentration of cytokines and interleukins is at three different stages.

extended strands, beta turns, and random coiled structures, and identified which of these structures were present in the sequence amino acid regions provided to the software (Fig. 3B). The 3D structure of the vaccine was predicted by reliable software based on deep learning and then reconstructed with the help of other software in terms of conformational shape, energy minimization, and atom distance (Fig. 3C). Refinement and validation of the vaccine model showed that the quality of the predicted model was good. The result was also confirmed by Ramachandran plot predictions<sup>66</sup> (Figs. 3D–G).

Protein–protein docking aimed to investigate the interaction and binding affinity of the vaccine structure with TLR2, TLR4, and MHC class I and II receptors. The expression of TLR2 and TLR4 receptors in the plasma membrane of different cell types, including immune cells, is linked to identifying pathogen molecular patterns and triggers of the immune response. By presenting microbe peptide fragments or vaccine epitopes on their surface, MHC molecules allow T lymphocytes to identify cells such as macrophages and stimulate and activate the immune response<sup>67</sup>. The docking results in the absorption of binding energy between the complexes, indicating a high degree of binding affinity, efficient docking, and stability of the complexes, which are expressed as negative binding energy values<sup>68</sup> (Tables 8, 9 and 10). In the docking analysis of the vaccine with receptors, the presence of salt bridges and hydrogen bonds is proof of the existence of interaction (Figs. 4, 5, 6 and 7). Studies on the structural integrity of the vaccine in a simulation environment closely related to natural systems were performed by MDS. This structure is characterized by its system balance, stability, and high flexibility, as demonstrated by the results of plots RMSD, RMSF, and Rg, which align with the findings of other studies<sup>69</sup>. In order to imitate the usual immune responses, an immune simulation was performed, and simultaneously with the repeated injection of the vaccine, a significant increase in immunoglobulins was observed. This means that memory B-cells have been created. In addition, elevated levels of CTL, HTL, and IFN- $\gamma$  are critical in ensuring a humoral and

adaptable immune response<sup>70</sup>. Due to an overall improvement in the immune response to the second and third doses of the vaccine, the immune simulation has produced results consistent with the anticipated immune response. In other words, three injections generated a sufficient immunogenic response (Fig. 10). However, if the vaccine is administered through 12 consecutive injections, the immune system will respond more strongly (Supplementary Fig. S17). The expression of the vaccine construct in a suitable expression vector is necessary for the production of recombinant proteins<sup>71</sup>. In this regard, the sequence was reverse transcribed and adapted for the *E. coli* strain K12 before cloning into the pET-28a (+) vector to ensure the translation efficiency of the designed vaccine construct in a specific expression system. An appropriate codon adaptation index and high GC content suggest a high protein expression level in the host. Finally, the gene of the vaccine construct, which has cleavage sites for specific restriction enzymes, was successfully cloned in the vector's multiple cloning site (MCS) (Fig. 9).

From the immunoinformatic approaches used in this study for SARS-CoV-2 in the design of multi-epitope vaccines against other pathogens such as MERS-CoV<sup>47</sup>, Nipah virus<sup>72</sup>, Zika<sup>73</sup>, dengue<sup>74</sup>, Hendra virus<sup>75</sup>, Malaria<sup>76</sup>, *Acinetobacter baumannii*<sup>77</sup>, *Leishmania infantum*<sup>78</sup>, *Pseudomonas aeruginosa*<sup>79</sup>, *Klebsiella pneumoniae*<sup>80</sup>, Kaposi sarcoma<sup>81</sup> and several other examples have also been used. A similar strategy has also been used to make a vaccine against cancer antigens. The results obtained from computational and immunoinformatics tools and molecular dynamics simulation analyses confirm the reduction of antigen growth, the induction of appropriate immune response, and the safety and stability of the candidate vaccine. Therefore, the multi-epitope vaccine designed based on fynomer has promising results against SARS-CoV-2, which requires further studies in vitro conditions to understand the efficacy and validation of the vaccine.

## Conclusions

The present study employed computational approaches to design a multi-epitope vaccine based on four viral structural proteins, conserved epitopes, and a fynomer antibody mimetic scaffold. These vaccines with such structures can be used to describe the immunogenic target in a novel way. This vaccine had favorable immunological properties with the ability to stimulate the innate and acquired immune systems and high population coverage. It also displayed acceptable physicochemical properties, stable structures, and proper binding with immune receptors. Thus, the product can be presented as a potential vaccine against the COVID-19 outbreak. However, these computational works require experimental validation to prove their effectiveness, which is possible by conducting additional studies and clinical trials at the in vitro and in vivo levels to determine the immunogenicity and safety of the vaccine.

## Methods

### Retrieval of SARS-CoV-2 protein sequences and antigenicity prediction

First, the complete sequences of structural proteins S, N, M, and E from the SARS-CoV-2 proteome for 32 different SARS-CoV-2 strains, including the variants Wuhan-Hu-1<sup>8</sup>, B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta), B.1.1.529 (Omicron), BA.1 (Omicron), BA.2 (Omicron), BA.2.12.1 (Omicron), BA.2.75 (Omicron), BA.4 (Omicron), BA.5 (Omicron), BQ.1.1 (Omicron), XBB.1.5 (Omicron), B.1.1.318, B.1.1.519, B.1.1.523, B.1.214.2, B.1.427 or B.1.429 (Epsilon), B.1.466.2, B.1.525 (Eta), B.1.526 (Iota), B.1.617.1 (Kappa), B.1.617.3, B.1.619 or B.1.619.1, B.1.620, B.1.621 (Mu), C.1.2, C.37 (Lambda) C.36.3 or C.36.3.1, P.2 (Zeta) and, R.1 from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/genbank/>), Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (<https://www.bv-brc.org>) and ViralZone from ExPasy (<https://viralzone.expasy.org>) were retrieved in FASTA format to use these sequences in predicting suitable epitopes for the making a multi-epitope vaccine against COVID-19. Additionally, the accuracy of some of these stored sequences was verified by similar sequences in the UniProtKB/Swiss-Prot (<https://www.expasy.org/resources/uniprotkb-swiss-prot>).

The antigenicity of these structural proteins was predicted using ANTIGENpro on scratch protein (<http://scratch.proteomics.ics.uci.edu/>) and VaxiJen (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). In ANTIGENpro, a threshold above 0.8<sup>82</sup> and VaxiJen above 0.4<sup>83</sup> for the viral category is considered to test the antigen level. In the following stage of the study, sequences with scores above the threshold were further analyzed. A set of predicted CTL, HTL, and LBL epitopes were included in an epitope bank created using these amino acid sequences.

### Prediction of T-cell epitopes

CTL epitopes using the NetCTL 1.2 server<sup>84</sup> (<https://www.cbs.dtu.dk/services/NetCTL/>) and the Immune Epitope Database (IEDB) servers<sup>85</sup> (<https://tools.iedb.org/mhci/>) were predicted. Epitope prediction by NetCTL is performed with 94–99% specificity and 54–89% sensitivity<sup>86</sup>. In this study, the most prevalent HLA class I alleles (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62) were tested on all nine amino acid peptide sequences to predict and evaluate potential CTL epitopes<sup>87</sup>. For the TAP transporter, C-terminal proteasomal cleavage and epitope recognition parameters, the default thresholds of 0.05, 0.15, and 0.75 were used, respectively<sup>36</sup>. Additionally, another set of epitopes for all HLA class I alleles in the IEDB was identified by the stabilized matrix method (SMM). Epitopes with percentile rank  $\leq 2$ , IC<sub>50</sub> < 200 nM, and high rank were investigated and considered strong binders<sup>88</sup>. Epitopes predicted for different alleles by NetCTL 1.2 and IEDB servers were selected for further analysis. Based on both servers' outcomes, common epitopes predicted by multiple alleles and favorable for the desired indices were picked for additional analysis.

On the other hand, HTL epitopes with a length of 15 residues were predicted by HLA class II alleles, including human HLA-DR, HLA-DP, and HLA-DQ alleles, using NetMHC II pan 3.2 server<sup>89</sup>. Based on the determined percentiles of 2, 10, and more than 10%, the presented peptides were categorized as strong, moderate, and non-adhesive, respectively<sup>36</sup>. Using the SMM-align method (NetMHCII 1.1) in the server IEDB<sup>90</sup>, another group of

HTL epitopes of the same length by 54 HLA class II alleles from the set of HLA-DR, HLA-DQ, and HLA-DP alleles were identified<sup>85</sup>. Epitopes with  $IC_{50} < 200$  nM, percentile rank  $\leq 2$ , and predicted by multiple alleles by both methods as strong binders were considered for further analyses. IEDB predicts the peptide binding to each MHC-II molecule using artificial neural networks (ANN) on a dataset trained with more than 500,000 binding affinity (BA) and eluted ligand mass spectrometry (EL) measurements and provides reliable results<sup>88</sup>. Also, using the IFNepitope<sup>91</sup> (<https://crdd.osdd.net/raghava/ifnepitope/>), IL4pred<sup>92</sup> (<https://webs.iitd.edu.in/raghava/il4pred/design.php>) and IL10pred<sup>93</sup> (<https://webs.iitd.edu.in/raghava/il10pred/predict3.php>), HTL epitopes that specifically induce IFN- $\gamma$ , IL-4 and IL-10 have been evaluated. The IFNepitope server detects with a maximum accuracy of 81.39% of all MHC class II binding overlapping peptides in a protein or antigen capable of inducing IFN- $\gamma$  from CD4<sup>+</sup> T-cells through methods such as machine learning strategy, motif-based analysis, and hybrid integrity<sup>94</sup>.

For further analysis, each selected epitope (CTL and HTL) was subjected to a series of selectivity filters (i.e., immunogenicity, antigenicity, allergenicity, and toxicity), which were predicted with the aid of an IEDB class I immunogenicity server, VaxiJen v2.0, AllerTOP 2.0 and ToxinPred, respectively.

### Prediction of LBL epitopes

This study predicted the linear (continuous) B-cell epitopes with default parameters by five servers: ABCPred<sup>95</sup>, BCPreds<sup>96</sup>, BepiPred<sup>97</sup>, ElliPro<sup>98</sup>, and SVMtrip<sup>99</sup>. To improve accuracy, ABCpred (<http://www.imtech.res.in/raghava/abcpred/>) uses a recurrent neural network to distinguish between epitopes and non-epitopes<sup>95</sup>. BCPreds prediction (<http://ailab.ist.psu.edu/bcpred/>) is based on kernel methods and the SVM model with an AUC value of 0.758<sup>96</sup>. IEDB Emini is the tool that allows access to the B-cell level in this server<sup>100</sup>. The training dataset for the random forest algorithm, which is used to train the BepiPred-2.0 web server (<http://www.cbs.dtu.dk/services/BepiPred/>) based on machine learning, includes numerous linear B-cell epitopes from the IEDB database<sup>97</sup>. The ElliPro tool (<https://tools.iedb.org/ellipro/>) from the IEDB server was also employed to predict linear B-cell epitopes using default parameters. SVMTrip (<http://sysbio.unl.edu/SVMTriP/>) uses a support vector machine to integrate similarity, tri-peptide affinity, and antigen epitope prediction, and its AUC reaches 0.702<sup>99</sup>.

Here, BepiPred was first used for the primary prediction, and the results from this server were checked and completed using four other servers. A linear B-cell epitope predicted by BepiPred was only if it was predicted by at least one of the other four servers. Predicting the surface accessibility of epitope sequences from structural proteins is also crucial for placing these sequences in the solvent-exposed region of antigens for B-cell fusion<sup>101</sup>, which Emini tool IEDB<sup>100,102</sup> has investigated. As a result, the non-surface-placed epitopes are also eliminated. Linear B-cell epitopes identified on the surface of structural proteins were selected based on their antigenic, non-allergic, and non-toxic nature.

### Multiple alignment and selection of fully conserved epitopes

Following multiple alignments of the desired structural protein sequences for various strains using ClustalW<sup>103</sup> supplied by MEGA<sup>104</sup> and Bioedit software, the modifications made, such as deletions, additions, and substitutions, were identified in each strain<sup>68</sup>. Then, the predicted CTL, HTL, and LBL epitopes were compared with the results of this alignment, and epitopes that remained constant in all strains were selected as fully conserved epitopes for inclusion in the final vaccine construct<sup>68</sup>.

### Population coverage, epitope protection analysis, and autoimmunity identification

The IEDB population coverage analysis tool (<https://tools.iedb.org/population/>) with default parameters was employed to test the immune response of CTL and HTL epitopes for a vaccine designed against their respective HLA genotype frequencies and adequate coverage of the human population worldwide<sup>105</sup>. On the other hand, the IEDB conservancy analysis tool (<http://tools.iedb.org/conservancy>) that is part of the IEDB (<http://www.immuneepitope.org>) was used to help select T and B-cell epitopes with a favorable level of conservation or variability and to assess the degree of conservation of the selected epitopes in comparison to other similar sequences<sup>106</sup>. This server output page displays the degree of epitope conservation from 0 (minimum) to 100% (maximum)<sup>107</sup>. In addition, to avoid inducing host autoimmune diseases and cross-reactions, each selected epitope was further analyzed with BLASTp against the UniProt database to check their homology to the human proteome, and any vaccine epitope with an identity equal to or higher than 35% was considered a homologous protein with the human proteome and removed. Therefore, only the non-homologous epitopes of the human proteome were selected for vaccine construction<sup>108,109</sup>.

### Multi-epitope vaccine design

In this study, an overlapping method was considered to prevent the selection of repetitive epitopes when choosing the appropriate epitopes from the various epitopes obtained for making a vaccine. Epitope types were chosen based on immunogenicity, non-allergenicity, high population coverage, lack of overlap with human proteins, and accessibility. They were then incorporated into the final structure of the multi-epitope vaccine<sup>81</sup>. The final vaccine consists of 566 amino acid residues, which include 6 CTL epitopes, 6 HTL epitopes, 2 cytokine-inducing epitopes (IFN- $\gamma$ , IL-4, and IL-10), 9 linear B-cell epitopes, PADRE universal epitope, and fynomer sequence (VTLFVALYDYEARTEDDLFSFHKGKGFQILNSSEGDWWEARSLTTGETGYIPSNYVAPVDSI). In the structure of this vaccine candidate, CTL epitopes were linked by the AAY (Ala-Ala-Tyr) linkers, the GPGPG (Gly-Pro-Gly-Pro-Gly) linkers linked HTL epitopes, and the LBL epitopes were linked by KK (Lys-Lys) linkers<sup>110,111</sup>. Fynomer was added to the vaccine structure by the GPGPG linker between the T and B-cell epitope regions (Supplementary Fig. S6). PADRE, which has the amino acid sequence AKFVAAWTLKAAA, is a simple carrier epitope. Combined with an adjuvant, it provides an effective response and can be used in making multi-epitope

and recombinant vaccines<sup>112</sup>. To activate helper T-cells (CD4<sup>+</sup> T-cells) and reduce the polymorphism of HLA-DR molecules in the population, this epitope has been added to the vaccine structure with the GPGPG linker. In addition, to enhance and heighten the long-term immune response, the adjuvant 50S ribosomal protein L7/L12 (rpIL) (Locus RL7\_MYCTU) (accession number P9WHE3 in NCBI) was added to the N-terminal of the vaccine via the EAAAK linker<sup>113</sup>. Since the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) discourage the 6His tag due to its potential adverse immune response, alternative tags, including humanized versions derived from human BLAST proteins such as H3A (HAAHAH), H5T (HTHTHTHTH), and H5E (HEHEHEHEH), were designed and tested, showing favorable potential in purification and functional properties in human biopharmaceutical production<sup>114</sup>. Finally, for this structure, a tag H5E was put on the C-terminal of the construct to aid in protein purification and identification.

### Evaluation of antigenicity, allergenicity, toxicity, solubility and physicochemical properties

The VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) with a threshold value of 0.4 with viral databases was used to extract prediction models of total protein antigenicity to verify the antigenicity of the vaccine. VaxiJen relies on the automatic cross-covariance (ACC) transformation of protein sequences into uniform vectors with original amino acid features<sup>83</sup>. AntigenPro is another server for this, a protein antigenicity predictor that uses a sequence-based, unaligned, pathogen-independent approach<sup>82</sup>. Scores of more than 0.4 for the VaxiJen viral model and more than 0.8 for ANTIGENpro are considered in this study as indicators of antigenicity. The AllergenFP 1.0 (<http://ddg-pharmfac.net/AllergenFP/>) and the AllerTOP 2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) were utilized to evaluate allergenicity. AllergenFP uses a binary classification to categorize allergens and non-allergens. Both servers use automatic cross-covariance (ACC) transformation to convert strings into uniform vectors after E-descriptors<sup>28,29</sup> describe the dataset. ToxinPred forecasts the toxicity of the vaccine and each of its epitopes. This server employs the SVM model and dataset with 1805 toxic peptides ( $\leq 35$  residues)<sup>30</sup> to distinguish between toxic and non-toxic peptides. The Protein-sol (<https://protein-sol.manchester.ac.uk>) and the SolPro (<http://scratch.proteomics.ics.uci.edu>) perform solubility checks. In contrast to SolPro, which is an SVM-based tool for predicting the solubility of a protein sequence with an estimated overall accuracy of more than 74% on tenfold cross-validation<sup>31</sup>, Protein-sol takes advantage of existing data on the solubility of *E. coli* proteins in a cell-free expression system<sup>32</sup>. The ExpASY ProtParam server predicted several physicochemical characteristics of the vaccine structure, including molecular weight, theoretical isoelectric point value (pI), charge, extinction coefficient, half-life, instability index, aliphatic index, and GRAVY score<sup>115</sup> (<https://web.expasy.org/protparam/>). Additionally, SignalP4.1<sup>116</sup> (<https://www.cbs.dtu.dk/services/SignalP/>) and the TMHMM v2.0 server<sup>117</sup> (<https://www.cbs.dtu.dk/services/TMHMM/>) examined the vaccine structure for the presence of signal peptides and transmembrane helices, respectively.

### Secondary structure prediction

The secondary structure of the vaccine construct has various characteristics, such as extended strands, alpha helix regions, random coils, and beta strands. These were analyzed using the SOPMA servers<sup>38</sup> ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)) and the PSIPRED v4.0<sup>118</sup> (<http://bioinf.cs.ucl.ac.uk/psipred/>). Both servers carry out the prediction process with an accuracy of over 80%. PSIPRED employs two feed-forward neural networks (FFN) to analyze PSI-BLAST (Position-Specific Iterated-BLAST) outputs<sup>118</sup>. After entering the vaccine sequence into SOPMA, the output width was set to 70, and the parameters for the number of conformational states, similarity threshold, and window width were set to 4 (Helix, sheet, turn, and coil), 8, and 17, respectively.

### Modeling, refinement, and validation of 3D structure

Robetta (<https://rosetta.bakerlab.org/>), an online protein structure prediction service that continuously evaluates different aspects of a prediction (coverage, local accuracy, completeness, etc.), modeled the final 3D vaccine structure via a cameo social project<sup>40,119</sup>. Protein folding in this server is based on deep learning, RoseTTAFold, TrRosetta, and an interactive submission interface that allows for custom sequence alignments for homology modeling, constraints, local fragments, and more in a fast and accurate manner<sup>120</sup>.

In the next step, the generated 3D model was refined by GalaxyRefine<sup>121</sup> (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). Using molecular dynamics simulations, this server first rebuilds and repacks the side chains, followed by general relaxation and structural disruption. According to CASP10 assessments, this method has performed best in improving the local structure quality<sup>122</sup>. GDT-HA, RMSD, MolProbability, clash score, and Ramachandran plot evaluated the quality of the refined model.

Model validation was performed using the ProSA web server<sup>65</sup> (<https://prosa.services.came.sbg.ac.at/prosa.php>) and the SAVES v6.0 server (<https://saves.mbi.ucla.edu>). Based on the Z-score, ProSA evaluates and confirms the overall and local quality of the predicted model and calculates an overall quality score for the desired input structure. If the Z-score is outside the specified range for the native protein, the probability of errors in the predicted structure increases. A local quality score plot is shown in the 3D molecular viewer<sup>65</sup> to make it easier to spot problematic areas of the model. The subsets of ERRAT<sup>123</sup>, VERIFY3D<sup>124</sup>, PROVE<sup>125</sup>, PROCHECK<sup>126</sup>, and WHATCHECK<sup>127</sup>, which are made available to users through SAVES v6.0, can be used to predict the structural quality of the chosen model and the evaluation of various stereochemical parameters of the protein structure. The PROCHECK tool evaluates the anticipated model's stereochemical quality by analyzing the model's overall geometry with residues through the residue geometry<sup>126,128</sup>. To generate the Ramachandran plot, there are several tools and servers available, including the PDBsum server<sup>129</sup>, MolProbability<sup>130</sup>, STAN Server<sup>131</sup>, and RAMPAGE<sup>43</sup>. However, in this study, we used the PDBsum server. Using UCSF ChimeraX software, the final model of the vaccine structure was drawn<sup>132</sup>.

### Prediction of conformational B-cell epitopes

After predicting, refining, and verifying the 3D model of the vaccine, conformational (discontinuous) B-cell epitopes were predicted using the ElliPro server<sup>98</sup> (<http://tools.iedb.org/elliopro>). Based on the geometrical features of the protein structure, solvent accessibility, and flexibility, this method predicts B-cell conformational epitopes, which are longer than other epitopes. The minimum score and maximum distance (Angstrom), two prediction parameters, were regarded as 0.8 and 6, respectively. ElliPro is a reliable tool for identifying antibody epitopes and predicting the B-cell conformational epitopes with an AUC score of 0.732<sup>98</sup>.

### Molecular docking and binding affinity analysis

The 3D structures of the TLR2, TLR4, MHC-I, and MHC-II receptors were retrieved from the Protein Data Bank (PDB) (<https://www.rcsb.org>) with the identifiers 2Z7X, 2Z63, 1I1Y, and 1KG0, respectively. Then, all ligands and water molecules attached to these structures were eliminated. The CPORIT<sup>133</sup> (Consensus Prediction Of Interface Residues in Transient complexes) predicted the active and inactive amino acid residues involved in the protein–protein interactions. These results were then submitted to the HADDOCK 2.4 server (<https://wenmr.science.uu.nl/haddock2.4/>) for docking<sup>134</sup> (Supplementary Table S16). HADDOCK (High Ambiguity Driven protein–protein DOCKing) uses Python scripts to compute the structure using crystalline and NMR structures<sup>135</sup>. Based on the lowest HADDOCK score, we selected the best cluster and corrected it with the HADDOCK correction interface. Using the PRODIGY web server, the binding affinity of each docked complex was determined by selecting the best-refined structure from the best cluster. The server offers web services for identifying biological interfaces and predicting affinity in complexes using crystallography<sup>136,137</sup>. The binding affinity of the four docked complexes was analyzed using this server at 37 °C, and its values were represented as Kd. Additionally,  $\Delta G$  values in kcal mol<sup>-1</sup> for the complexes were also obtained. Finally, Residues involved in interactions between vaccine and target receptors were visualized with ChimeraX and mapped by PDBsum<sup>138</sup>.

### Energy minimization and molecular dynamics simulation

The GROMACS version 5.1, a Linux-based and open-source program, was employed to conduct molecular dynamic simulation and energy minimization<sup>139</sup>. This program provides various calculation types, preparation, and analysis tools and is supported by several advanced techniques for free energy computing<sup>51</sup>. The vaccine pdb file was employed to initiate molecular dynamics simulations. The topology file generation required for energy minimization and equilibrium was performed under the OPLS (Optimised Potential for Liquid Simulation-All Atom) force field<sup>140</sup>. This structure fits inside an octagonal box of water molecules, represented by the three-point charge model spc216, with the boundary at most minuscule 10 Å away from any protein atom. After determining the net charge of the dissolved protein (vaccine), charged ions were added to the system for neutralization. The LINCS algorithm helped to limit covalent bonds with hydrogen atoms<sup>141</sup>, and long-range electrostatic interactions were treated with Particle Mesh Ewald (PME) using a 10 Å real-space line cutoff<sup>99</sup>. With the main chain atoms constrained to initial coordinates, this system was first minimized under 50,000 steps of steepest descent energy minimization to remove close contacts and spatial overlap. Then, it was put into a two-step NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) equilibrium phase<sup>142</sup>. The restrained system was gradually heated up to 300 °K (kelvins) under constant volume conditions at 100 ps. The system was then equilibrated for 1 ns using the constant isothermal-isobaric set at 1 atmosphere and 300 °K without any restrictions, which was accomplished by the V-Rescale thermostat and Parrinello–Rahman barostat with an integration time step of 2 fs<sup>143,144</sup>. Using the PME method, electrostatic forces for both NVT and NPT were calculated<sup>145</sup>. In the last step of MD generation, a complete simulation was conducted for 10 ns, and the structure coordinates were recorded every 10 ps. The GROMACS tool was then utilized to analyze the trajectory files. So, The Root Mean Square Deviation (RMSD) for the backbone, The Root Mean Square Fluctuation (RMSF) for the side chain, and the radius of gyration (Rg) were determined to check convergence. The graphs obtained from various analyses were drawn using the Xmgrace design tool<sup>146</sup>.

### Codon optimization and in-silico simulation of vaccine

The Java Codon Adaptation Tool (JCat) (<http://www.prodoric.de/JCat>) was used for codon optimization and reverse translation. JCat performs codon adaptation for most selected prokaryotic and eukaryotic organisms sequenced<sup>53</sup>. After codon optimization in the *E. coli* strain K12, the vaccine cDNA sequence was created, which can be used for efficient expression in this strain. Further sequence evaluation included selecting parameters to avoid rho-independent transcription terminators, prokaryotic ribosome binding sites, and restriction enzyme cleavage sites. The result includes two measures, guanine-cytosine (GC) content, and codon adaptation index (CAI) score, that assess protein expression levels<sup>147</sup>. Additionally, restriction enzyme sites *Eco53KI* and *EcoRV* were added to the N and C-terminal of the optimized codon sequence for in-silico cloning in the pET-28a (+) vector using SnapGene 6.2 software<sup>148</sup> (From Insightful Science, available at <https://www.snapgene.com>), respectively.

### Simulation of the immune system

Multi-epitope vaccine immunological simulations were carried out on the C-IMMSIM server (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>) to characterize the immunogenicity of chimeric peptides and real immunogenicity profiles<sup>54</sup>. This server is an immune response agent-based simulator and a flexible tool that uses position-specific scoring matrices (PSSM) and machine learning to predict immune epitopes and interactions. At the same time, it simulates three distinct anatomical regions in mammals, including the thymus, bone marrow, and tertiary lymph organs<sup>149</sup>. To simulate immunity, the vaccine was administered in three injections, each containing 1000 molecules, with a four-week interval between them. The random seed, simulation volume, and simulation step

parameters were set to 12,345, 10  $\mu$ l, and 1050, respectively. According to studies, the minimum interval for two injections should be four weeks. The three injections were followed by time steps of 1, 84, and 168, each representing approximately 8 h of real-life<sup>113</sup>. It is suggested that consecutive vaccine injections be administered at specific time intervals to evaluate the effects of vaccine exposure on repeated SARS-CoV-2 infections.

## Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Received: 13 November 2023; Accepted: 30 April 2024

Published online: 04 May 2024

## References

- Peng, X. L. *et al.* Advances in the design and development of SARS-CoV-2 vaccines. *Mil. Med. Res.* **8**(1), 67 (2021).
- Choudhary, S., Malik, Y. S. & Tomar, S. Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using in silico structure-based virtual screening approach. *Front. Immunol.* **11**, 1664 (2020).
- World-Health-Organization. *Novel Coronavirus (2019-nCoV)*. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200201-sitrep-12-ncov.pdf>. Accessed on 1 Feb 2020 (2020).
- World-Health-Organization. *2019-NCoV Outbreak Is an Emergency of International Concern*. <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf>. Accessed 1 Jan 2020 (2020).
- World-Health-Organization. More than 150 countries engaged in COVID-19 vaccine global access facility Geneva. <https://www.who.int/news/item/15-07-2020-more-than-150-countries-engaged-in-covid-19-vaccine-global-access-facility>. Accessed 15 July 2020 (2020).
- Shahzamani, K. *et al.* Vaccine design and delivery approaches for COVID-19. *Int. Immunopharmacol.* **100**, 108086 (2021).
- Tortorici, M. A. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **26**(6), 481–489 (2019).
- Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**(7798), 265–269 (2020).
- Hasan, A. *et al.* A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *J. Biomol. Struct. Dyn.* **39**(8), 3025–3033 (2021).
- Rafi, M. O. *et al.* Design of a multi-epitope vaccine against SARS-CoV-2: Immunoinformatic and computational methods. *RSC Adv.* **12**(7), 4288–4310 (2022).
- Rahman, N. *et al.* Virtual screening of natural products against type II transmembrane serine protease (TMPRSS2), the priming agent of coronavirus 2 (SARS-CoV-2). *Molecules (Basel, Switzerland)* **25**(10), 2271 (2020).
- Sutton, P. & Boag, J. M. Status of vaccine research and development for *Helicobacter pylori*. *Vaccine* **37**(50), 7295–7299 (2019).
- Sakaguchi, S., Miyara, M., Costantino, C. M. & Hafler, D. A. FOXP3+ regulatory T cells in the human immune system. *Nat. Rev. Immunol.* **10**(7), 490–500 (2010).
- Ahlers, J. D. & Belyakov, I. M. Molecular pathways regulating CD4(+) T cell differentiation, energy and memory with implications for vaccines. *Trends Mol. Med.* **16**(10), 478–491 (2010).
- Sakharkar, K. R., Sakharkar, M. K. & Chandra, R. *Post-Genomic Approaches in Drug and Vaccine Development* (ed. River Publishers), 1–452 (River Publishers, 2015).
- Singh, H., Jakhar, R. & Sehrawat, N. Designing spike protein (S-protein) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics. *Heliyon* **6**(11), e05528 (2020).
- Walls, A. C. *et al.* Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**(7592), 114–117 (2016).
- Weidle, U. H., Auer, J., Brinkmann, U., Georges, G. & Tiefenthaler, G. The emerging role of new protein scaffold-based agents for treatment of cancer. *CGP* **10**(4), 155–168 (2013).
- Silacci, M. *et al.* Linker length matters, fynomer-Fc fusion with an optimized linker displaying picomolar IL-17A inhibition potency. *J. Biol. Chem.* **289**(20), 14392–14398 (2014).
- Grabulovski, D., Kaspar, M. & Neri, D. A novel, non-immunogenic Fyn SH3-derived binding protein with tumor vascular targeting properties. *J. Biol. Chem.* **282**(5), 3196–3204 (2007).
- Yuan, F., Wang, L., Fang, Y. & Wang, L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transbound. Emerg. Dis.* **68**(6), 3288–3304 (2021).
- Liljeroos, L., Malito, E., Ferlenghi, I. & Bottomley, M. J. Structural and computational biology in the design of immunogenic vaccine antigens. *J. Immunol. Res.* **2015**, 156241 (2015).
- Kloetzel, P. M. The proteasome and MHC class I antigen processing. *Biochim. Biophys. Acta.* **1695**(1–3), 225–233 (2004).
- Larsen, M. V. *et al.* An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **35**(8), 2295–2303 (2005).
- Arrieta-Bolaños, E., Madrigal, J. A. & Shaw, B. E. Human leukocyte antigen profiles of Latin American populations: Differential admixture and its potential impact on hematopoietic stem cell transplantation. *Bone Marrow Res.* **2012**, 136087 (2012).
- Ong, E. *et al.* Vaxign-ML: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics* **36**(10), 3185–3191 (2020).
- Kathwate, G. H. In silico design and characterization of multi-epitopes vaccine for SARS-CoV-2 from its spike protein. *Int. J. Pept. Res. Ther.* **28**(1), 37 (2022).
- Dimitrov, I., Flower, D. R. & Doytchinova, I. AllerTOP—A server for in silico prediction of allergens. *BMC Bioinform.* **14**(Suppl 6), S4 (2013).
- Dimitrov, I., Naneva, L., Doytchinova, I. & Bangov, I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**(6), 846–851 (2014).
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium & Raghava, G. P. In silico approach for predicting toxicity of peptides and proteins. *PLoS one* **8**(9), e73957 (2013).
- Magnan, C. N., Randall, A. & Baldi, P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**(17), 2200–2207 (2009).
- Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein-Sol: A web tool for predicting protein solubility from sequence. *Bioinformatics* **33**(19), 3098–3100 (2017).
- D'Amico, C. *et al.* Development of vaccine formulations: Past, present, and future. *Drug Deliv. Transl. Res.* **11**, 353–372 (2021).
- Barbier, A. J. *et al.* The clinical progress of mRNA vaccines and immunotherapies. *Nat. Biotechnol.* **40**, 840–854 (2022).
- Ali, M. *et al.* Exploring dengue genome to construct a multi-epitope-based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* **7**(1), 9232 (2017).
- Kar, T. *et al.* A candidate multi-epitope vaccine against SARS-CoV-2. *Sci. Rep.* **10**(1), 10895 (2020).

37. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4), 404–405 (2000).
38. Geourjon, C. & Deléage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *CABIOS* **11**(6), 681–684 (1995).
39. Tahmoorespur, M., Nazifi, N. & Pirkhezranian, Z. In silico prediction of B-cell and T-cell epitopes of protective antigen of *Bacillus anthracis* in development of vaccines against anthrax. *Iran. J. Appl. Anim. Sci.* **7**(3), 429–436 (2017).
40. Haas, J. *et al.* Introducing “best single template” models as reference baseline for the Continuous Automated Model Evaluation (CAMEO). *Proteins* **87**(12), 1378–1387 (2019).
41. Hee Shin, W., Rie Lee, G., Heo, L., Hasup, L. & Chaok, S. Prediction of protein structure and interaction by GALAXY protein modeling programs. *Biodesign* **2**(1), 1–11 (2014).
42. Messaoudi, A., Belguith, H. & Ben Hamida, J. Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1  $\beta$ -lactamase. *Theor. Biol. Med. Model.* **10**, 22 (2013).
43. Lovell, S. C. *et al.* Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* **50**(3), 437–450 (2003).
44. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
45. de Vries, S. J., van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**(5), 883–897 (2010).
46. Tarrahimofrad, H., Rahimnahal, S., Zamani, J., Jahangirian, E. & Aminzadeh, S. Designing a multi-epitope vaccine to provoke the robust immune response against influenza A H7N9. *Sci. Rep.* **11**(1), 24485 (2021).
47. Mahmud, S. *et al.* Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach. *Sci. Rep.* **11**(1), 15431 (2021).
48. Minch, M. J. An introduction to hydrogen bonding (Jeffrey, G. A.). *J. Chem. Educ.* **76**(6), 759 (1999).
49. Spassov, D. S., Atanasova, M. & Doytchinova, I. A role of salt bridges in mediating drug potency: A lesson from the *N*-myristoyltransferase inhibitors. *Front. Mol. Biosci.* **9**, 1066029 (2023).
50. Khalid, K., Irum, S., Ullah, S. R. & Andleeb, S. In-silico vaccine design based on a novel vaccine candidate against infections caused by *Acinetobacter baumannii*. *Int. J. Pept. Res. Ther.* **28**(1), 16 (2022).
51. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
52. Sayed, S. B. *et al.* Exploring lassa virus proteome to design a multi-epitope vaccine through immunoinformatics and immune simulation analyses. *Int. J. Pept. Res. Ther.* **26**(4), 2089–2107 (2020).
53. Grote, A. *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**(Web Server issue), W526–W531 (2005).
54. Rapin, N., Lund, O., Bernaschi, M. & Castiglione, F. Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE* **5**(4), e9862 (2010).
55. Pollard, A. J. & Bijker, E. M. A guide to vaccinology: From basic principles to new developments. *Nat. Rev. Immunol.* **21**, 83–100 (2021).
56. Panahi, Y. *et al.* Fully understanding the efficacy profile of the COVID-19 vaccination and its associated factors in multiple real-world settings. *Front. Immunol.* **13**, 947602 (2022).
57. Morse, J. S., Lalonde, T., Xu, S. & Liu, W. R. Learning from the past: Possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *ChemBiochemistry* **21**(5), 730–738 (2020).
58. Wadman, M. Why obesity worsens COVID-19. *Science* **369**(6509), 1280–1281 (2020).
59. Goumari, M. M., Farhani, I., Nezafat, N. & Mahmoodi, S. Multi epitope vaccines (MEVs), as a novel strategy against infectious diseases. *Curr. Proteom.* **17**(5), 354–364 (2020).
60. Osamor, V. C., Ikekanam, E., Bishung, J. U., Abiodun, T. N. & Ekpo, R. H. COVID-19 vaccines: Computational tools and development. *Inf. Med. Unlocked.* **37**, 101164 (2023).
61. Sette, A. & Fikes, J. Epitope-based vaccines: An update on epitope identification, vaccine design and delivery. *Curr. Opin. Immunol.* **15**(4), 461–470 (2003).
62. Bayat, M., Asemani, Y. & Najafi, S. Essential considerations during vaccine design against COVID-19 and review of pioneering vaccine candidate platforms. *Int. Immunopharmacol.* **97**, 107679 (2021).
63. Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B. & Prado, E. O. SARS-CoV-2 vaccines strategies: A comprehensive review of phase 3 candidates. *NPJ Vaccines* **6**(1), 28 (2021).
64. Byrne, P. O. & McLellan, J. S. Principles and practical applications of structure-based vaccine design. *Curr. Opin. Immunol.* **77**, 102209 (2022).
65. 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**(Web Server issue), W407–W410 (2007).
66. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods Mol. Biol. (Clifton, N.J.)* **1607**, 595–610 (2017).
67. Noor, F. *et al.* Designing a multi-epitope vaccine against *Chlamydia pneumoniae* by integrating the core proteomics, subtractive proteomics and reverse vaccinology-based immunoinformatics approaches. *Comput. Biol. Med.* **145**, 105507 (2022).
68. Almofti, Y. A., Abd-Elrahman, K. A. & Eltilib, E. E. M. Vaccinomic approach for novel multi epitopes vaccine against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). *BMC Immunol.* **22**(1), 22 (2021).
69. Dong, R., Chu, Z., Yu, F. & Zha, Y. Contriving multi-epitope subunit of vaccine for COVID-19: Immunoinformatics approaches. *Front. Immunol.* **11**, 1784 (2020).
70. Oluwagbemi, O. O. *et al.* Bioinformatics, computational informatics, and modeling approaches to the design of mRNA COVID-19 vaccine candidates. *Computation* **10**(7), 117 (2022).
71. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5**, 172 (2014).
72. Ojha, R., Pareek, A., Pandey, R. K., Prusty, D. & Prajapati, V. K. Strategic development of a next-generation multi-epitope vaccine to prevent Nipah virus zoonotic infection. *ACS omega.* **4**(8), 13069–13079 (2019).
73. Antonelli, A. C. B. *et al.* In silico construction of a multi-epitope Zika virus vaccine using immunoinformatics tools. *Sci. Rep.* **12**(1), 53 (2022).
74. Dixit, N. K. Design of monovalent and chimeric tetravalent dengue vaccine using an immunoinformatics approach. *Int. J. Pept. Res. Ther.* **27**(4), 2607–2624 (2021).
75. Kamthania, M. *et al.* Immunoinformatics approach to design T-cell epitope-based vaccine against Hendra virus. *Int. J. Pept. Res. Ther.* **25**(3), 1627–1637 (2019).
76. Pandey, R. K., Bhatt, T. K. & Prajapati, V. K. Novel Immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci. Rep.* **8**(1), 1125 (2018).
77. Dey, J. *et al.* Designing of multi-epitope peptide vaccine against *Acinetobacter baumannii* through combined immunoinformatics and protein interaction-based approaches. *Immunol. Res.* **71**(4), 639–662 (2023).
78. Shams, M. *et al.* Engineering a multi-epitope vaccine candidate against *Leishmania infantum* using comprehensive immunoinformatics methods. *Biology* **77**(1), 277–289 (2022).

79. Solanki, V., Tiwari, M. & Tiwari, V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against *Pseudomonas aeruginosa*. *Sci. Rep.* **9**(1), 5240 (2019).
80. Dar, H. A. *et al.* Immunoinformatics-aided design and evaluation of a potential multi-epitope vaccine against *Klebsiella pneumoniae*. *Vaccines* **7**(3), 88 (2019).
81. Chauhan, V., Rungta, T., Goyal, K. & Singh, M. P. Designing a multi-epitope-based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci. Rep.* **9**(1), 2517 (2019).
82. Magnan, C. N. *et al.* High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics* **26**(23), 2936–2943 (2010).
83. Doytchinova, I. A. & Flower, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
84. Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* **8**, 424 (2007).
85. Moutaftsi, M. *et al.* A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat. Biotechnol.* **24**(7), 817–819 (2006).
86. Jurtz, V. *et al.* NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immun. J. Immun., Balt.* **199**(9), 3360–3368 (2017).
87. Nielsen, M. *et al.* NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* **2**(8), e796 (2007).
88. Singh, A., Thakur, M., Sharma, L. K. & Chandra, K. Designing a multi-epitope peptide-based vaccine against SARS-CoV-2. *Sci. Rep.* **10**(1), 16219 (2020).
89. Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**(3), 394–406 (2018).
90. Reynisson, B. *et al.* Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J. Proteome Res.* **19**(6), 2304–2315 (2020).
91. Shey, R. A. *et al.* In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* **9**(1), 4409 (2019).
92. Wang, P. *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinform.* **11**, 568 (2010).
93. Nagpal, G. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci. Rep.* **7**, 42851 (2017).
94. Dhandu, S. K., Vir, P. & Raghava, G. P. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct.* **8**, 30 (2013).
95. Saha, S. & Raghava, G. P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **65**(1), 40–48 (2006).
96. El-Manzalawy, Y., Dobbs, D. & Honavar, V. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit.* **21**(4), 243–255 (2008).
97. Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* **45**(W1), W24–W29 (2017).
98. Ponomarenko, J. *et al.* ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **9**, 514 (2008).
99. Yao, B., Zhang, L., Liang, S. & Zhang, C. SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE* **7**(9), e45152 (2012).
100. Emimi, E. A., Hughes, J. V., Perlow, D. S. & Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* **55**(3), 836–839 (1985).
101. Sanchez-Trincado, J. L., Gomez-Perosanz, M. & Reche, P. A. Fundamentals and methods for T- and B-cell epitope prediction. *J. Immunol. Res.* **2017**, 2680160 (2017).
102. Almof, Y., Abd-elrahman, K., Gassmallah, S. & Salih, M. Multi epitopes vaccine prediction against severe acute respiratory syndrome (SARS) coronavirus using immunoinformatics approaches. *Am. J. Microbiol. Res.* **6**(3), 94–114 (2018).
103. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinf.* **23**, 2947–2948 (2007).
104. Kumar, S., Stecher, G., Li, M., Nuyez, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**(6), 1547–1549 (2018).
105. Bui, H. H. *et al.* Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform.* **7**, 153 (2006).
106. Bui, H. H., Sidney, J., Li, W., Fusseder, N. & Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform.* **8**, 361 (2007).
107. Peters, B. *et al.* The design and implementation of the immune epitope database and analysis resource. *Immunogenetics* **57**(5), 326–336 (2005).
108. Tan, C., Zhu, F., Pan, P., Wu, A. & Li, C. Development of multi-epitope vaccines against the monkeypox virus based on envelop proteins using immunoinformatics approaches. *Front. Immunol.* **14**, 1112816 (2023).
109. González-Pech, R. A., Stephens, T. G. & Chan, C. X. Commonly misunderstood parameters of NCBI BLAST and important considerations for users. *Bioinformatics* **35**(15), 2697–2698 (2019).
110. Dorosti, H. *et al.* Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *J. Biomol. Struct. Dyn.* **37**(13), 3524–3535 (2019).
111. Nain, Z. *et al.* Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis* using immunoinformatic approaches. *J. Biomol. Struct. Dyn.* **38**(16), 4850–4867 (2020).
112. Ahmad, T. A., Eweida, A. E. & Sheweita, S. A. B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials Vaccinol.* **5**, 71–83 (2016).
113. Martin, W. R. & Cheng, F. A rational design of a multi-epitope vaccine against SARS-CoV-2 which accounts for the glycan shield of the spike glycoprotein. *J. Biomol. Struct. Dyn.* **40**(15), 7099–7113 (2022).
114. López-Laguna, H. *et al.* Insights on the emerging biotechnology of histidine-rich peptides. *Biotechnol. Adv.* **54**, 107817 (2022).
115. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*. 571–607 (2005).
116. Nielsen, H. Predicting secretory proteins with SignalP. *Methods Mol. Biol. (Clifton, N.J.)* **1611**, 59–73 (2017).
117. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
118. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**(W1), W402–W407 (2019).
119. Robin, X. *et al.* Continuous Automated Model Evaluation (CAMEO)-perspectives on the future of fully automated evaluation of structure prediction methods. *Proteins* **89**(12), 1977–1986 (2021).
120. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**(Web Server issue), W526–W531 (2004).
121. Ko, J., Park, H., Heo, L. & Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **40**(Web Server issue), W294–W297 (2012).
122. Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **41**(Web Server issue), W384–W388 (2013).



123. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**(9), 1511–1519 (1993).
124. Bowie, J. U., Lüthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**(5016), 164–170 (1991).
125. Pontius, J., Richelle, J. & Wodak, S. J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* **264**(1), 121–136 (1996).
126. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**(2), 283–291 (1993).
127. Hoof, R. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**(6580), 272 (1996).
128. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* **8**(4), 477–486 (1996).
129. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**(1), 129–134 (2018).
130. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**(1), 293–315 (2018).
131. Agnihotry, S., Kumar Pathak, R., Bukhsh Singh, D., Tiwari, A. & Hussain, I. Protein structure prediction. *Bioinformatics.* 177–188 (2022).
132. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**(1), 70–82 (2021).
133. de Vries, S. J. & Bonvin, A. M. CPORT: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE* **6**(3), e17695 (2011).
134. van Zundert, G. C. P. *et al.* The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**(4), 720–725 (2016).
135. Honorato, R. V. *et al.* Structural biology in the clouds: The WeNMR-EOSC ecosystem. *Front. Mol. Biosci.* **8**, 729513 (2021).
136. Xue, L. C., Rodrigues, J. P., Kastriitis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: A web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**(23), 3676–3678 (2016).
137. Vangone, A. & Bonvin, A. M. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife*, **4**, e07454 (2015).
138. Laskowski, R. A. *et al.* PDBsum: A web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* **22**(12), 488–490 (1997).
139. Lemkul, J. A. From proteins to perturbed Hamiltonians: A suite of tutorials for the GROMACS-2018 molecular simulation package [article v1.0]. *LiveCoMS.* **1**(1), 5068 (2018).
140. Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**(6), 1657–1666 (1988).
141. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**(12), 1463–1472 (1997).
142. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**(1), 014101 (2007).
143. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
144. Ke, Q., Gong, X., Liao, S., Duan, C. & Li, L. Effects of thermostats/barostats on physical properties of liquids by molecular dynamics simulations. *J. Mol. Liq.* **365**, 120116 (2022).
145. Chitongo, R., Obasa, A. E., Mikasi, S. G., Jacobs, G. B. & Cloete, R. Molecular dynamic simulations to investigate the structural impact of known drug resistance mutations on HIV-1C Integrase-Dolutegravir binding. *PLoS ONE* **15**(5), e0223464 (2020).
146. Soman, S. S., Sivakumar, K. C. & Sreekumar, E. Molecular dynamics simulation studies and in vitro site directed mutagenesis of avian beta-defensin Apl\_AvBD2. *BMC Bioinform.* **11**(Suppl 1), S7 (2010).
147. Samad, A. *et al.* Designing a multi-epitope vaccine against SARS-CoV-2: An immunoinformatics approach. *J. Biomol. Struct. Dyn.* **40**(1), 14–30 (2022).
148. Yang, Z., Bogdan, P. & Nazarian, S. An in silico deep learning approach to multi-epitope vaccine design: A SARS-CoV-2 case study. *Sci. Rep.* **11**(1), 3238 (2021).
149. Castiglione, F., Mantile, F., De Berardinis, P. & Prisco, A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med.* **2012**, 842329 (2012).

## Author contributions

J. S., designed the construct, performed analyses, and wrote the manuscript. B. B. K., supervised the study, provided specialized scientific and revised the paper. A. GH, D. SH. and H. O provided specialized scientific support. All authors have agreed with the manuscript and provided their consent for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61025-2>.

**Correspondence** and requests for materials should be addressed to B.B.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2024, corrected publication 2024