



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Design of a multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach

Samira Sanami^a, Milad Zandi^b, Behzad Pourhossein^b, Gholam-Reza Mobini^c, Mohsen Safaei^a, Atena Abed^a, Pooria Mohammadi Arvejeh^c, Fatemeh Amini Chermahini^c, Morteza Alizadeh^{d,*}

^a Department of Medical Biotechnology, School of Advanced Technologies, Shahrekord University of Medical Sciences, Shahrekord, Iran

^b Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

^c Cellular and Molecular Research Center, Basic Health Sciences Institute, Shahrekord University of Medical Sciences, Shahrekord, Iran

^d Department of Tissue Engineering, School of Advanced Technologies, Shahrekord University of Medical Sciences, Shahrekord, Iran

ARTICLE INFO

Article history:

Received 2 May 2020

Received in revised form 30 June 2020

Accepted 7 July 2020

Available online 15 July 2020

Keywords:

COVID-19

Vaccine

Epitopes

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused COVID-19 disease in China. So far, no vaccine has licensed to protect against infection with COVID-19, therefore an effective COVID-19 vaccine needed. The aim of this study was to predict antigenic peptides of SARS-CoV-2 for designing the COVID-19 vaccine using immunoinformatic analysis. In this study, T and B-cell epitopes of S protein were predicted and screened based on the antigenicity, toxicity, allergenicity, and cross-reactivity with human proteomes. The epitopes were joined by the appropriate linker. LT-IIc as an adjuvant was attached to the end of the structure. The secondary and 3D structure of the vaccine was predicted. The refinement process was performed to improve the quality of the 3D model structure; the validation process is performed using the Ramachandran plot and ProSA z-score. The proposed vaccine's binding affinity to the HLA-A11:01 and HLA-DRB1_01:01 molecule was evaluated by molecular docking. Using molecular dynamics, the stability of vaccine-HLA complexes was also evaluated. Finally, *in silico* gene cloning was performed in the pET30a (+) vector. The findings suggest that the current vaccine may be a promising vaccine to prevent SARS-CoV-2 infection.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

In December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused COVID-19 disease in China, which was associated with a cluster of respiratory infections [1]. Infection with SARS-CoV-2 has spread to various countries and created a significant threat to public health. Following the outbreak of SARS-CoV in 2002 and 2003 and MERS-CoV outbreak in 2012, the SARS-CoV-2 outbreak is a recurrence of the previous occurrence with new pressure, and the bat is highly probable to be the origin of all types of CoV [2]. Researchers have confirmed the transmission of this virus from human to human, which has led to a further outbreak of the disease [3]. Coronaviruses are classified into four genera: Alpha, Beta, Gamma, and Delta Coronavirus. Alpha and Beta coronaviruses infect bats and human [4,5], SARS-CoV-2 is a member of β coronavirus genera [6]. SARS-CoV-2 is a single strand, positive-sense RNA virus with a genome of around 30 Kb in length [7]. In all coronaviruses, the structural proteins comprise spike (S), envelope (E), membrane protein (M), and nucleoprotein (N) [8,9]. SARS-CoV-2 uses the S protein as the key to binding and entering host cells therefore

understanding its structure and function is important. The S protein is a type I transmembrane protein and highly glycosylated. Spike glycoproteins assemble into trimers on the viral particle surface. S glycoprotein has three domains including ectodomain, transmembrane, and endodomain. Ectodomain subdivided into S1 and S2 subunits, S1 is important for binding to host cell receptors, the amino acid sequences of S1 are more variable than S2 [10]. S protein is an attractive target for vaccine development for the following reasons: S protein is located at the virus surface and uses angiotensin-converting enzyme 2 receptor in respiratory epithelial cells to enter the host cells [11,12], it also can induce neutralizing antibodies [13].

The development and production of a vaccine are costly and will take many years to achieve this goal, the various strategies were used to reduce the time and costs of this process [14]. Reverse vaccinology or vaccinomics is a new method that combines immunogenomics and bioinformatics to develop novel vaccines [15]. This method has many benefits over conventional vaccinology, it decreases the vaccine development time and costs and it also makes possible to study non-cultivable or risky microorganisms [16].

Scientists are working very hard to combat this virus, but still, there is no drug or vaccine available to protect against infection with COVID-19 [17]. Thus, in this regard, we decided to design a new multi-epitope

* Corresponding author.

E-mail address: mor1361@gmail.com (M. Alizadeh).

vaccine against SARS-CoV-2 infection using reverse vaccinology approaches. In this research, first, the CTL, HTL, and B-cell epitopes of the S protein were predicted using ProPred-1, ProPred, and ABCPred servers, respectively, and then were selected based on antigenicity, toxicity, allergenicity, and cross-reactivity with human proteomes. The selected epitopes were joined together using the appropriate linker to construct the main core of the multi-epitope vaccine. Since the multi-epitope vaccine has poor antigenicity, LT-IIc as an adjuvant to overcome this weakness was attached to the end of the structure. LT-IIc is a type II subfamily of LTBs consisting of 121 amino acids. Moreover, when co-administered with a multi-epitope vaccine it shows an increase in CD8 + T cell immune responses [18]. Next, the physicochemical properties of the construct were investigated, the 3D structure of the protein was predicted, and finally, its affinity to the MHC I and II molecules was investigated through docking, following that, was performed the molecular dynamics (MD) simulation of docking complexes. Finally, the vaccine coding gene expression in *Escherichia coli* (*E. coli*) was evaluated, and *in silico* gene cloning was performed in the pET30a (+) vector.

2. Methodology

2.1. Amino acid sequence retrieval

The amino acid sequences of SARA-COV-2 S protein (QII57328.1) and LT-IIc (H6W8F2) were obtained in FASTA format from NCBI (<https://www.ncbi.nlm.nih.gov/>) and Uniprot (<https://www.uniprot.org/>) databases, respectively.

2.2. Prediction and screening of epitopes

The binding epitopes to MHC I alleles were predicted using the ProPred-1 server (<http://crdd.osdd.net/raghava/propred1/>). ProPred-1 server is based on a method of predicting epitopes that bind to 47 different alleles of MHC I [19]. To predict HTL epitopes was used the ProPred server (<http://crdd.osdd.net/raghava/propred/>) which uses quantitative matrices to predict T- cell epitopes binding to 51 different alleles of MHC II [20]. In this study, peptide fragments with default parameters for MHC I and II alleles with the highest coverage of Iran's population were predicted. Gamma interferon (IFN- γ) is a cytokine that plays a key role in both innate and adaptive immune response; IFNepitope server (<https://webs.iitd.edu.in/raghava/ifnepitope/predict.php>) determined the IFN- γ -inducing HTL epitopes. ABCpred server (<http://crdd.osdd.net/raghava/abcpred/>) using an artificial neural network predicted the B-cell epitopes. This is the first server developed using fixed-length patterns based on a recurrent neural network. The server predicts peptides 10–20 mer in length and the threshold is among 0.1 to 1. In this study, the server was predicted peptide fragments of 16 mer length, and the threshold was set at 0.89 [21].

Antigenicity, toxicity, and allergenicity were evaluated for predicted epitopes. The antigenicity of the epitopes were calculated by Vaxijen v2.0 server (<http://www.ddgpharmfac.net/vaxijen/Vaxijen/Vaxijen.html>), which is based on the transformation of the protein sequences auto cross-covariance (ACC) into uniform vectors of main amino acid properties. The performance accuracy of the Vaxijen server varied from 70 to 89%, based on the selected model organism. In this study, the Vaxijen v2.0 server determined the epitopes antigenicity based on a threshold of 0.4 and a virus model [22–24]. The epitope toxicity was determined using the ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred/>). This server predicts of toxic or non-toxic peptides from a wide variety of peptides, along with their important physical and chemical properties such as hydrophobicity, hydropathicity, amphipathicity, molecular weight, and pI charge [25]. The AllerTOP v. 2.0 Server (<https://www.ddg-pharmfac.net/AllerTOP/index.html>) predicted epitopic allergenicity. The approach used in this server is based on the automatic cross-covariance (ACC) transformation of protein sequences into uniform vectors of equal length [26]. The PIR

peptide matching program (<https://research.bioinformatics.udel.edu/peptidematch/index.jsp>) was used to investigate the presence or absence of similarity with the human proteome in predicted epitope sequences [27].

2.3. Vaccine design and prediction of features

The selected T and B-cell epitopes from the previous step were placed in the vaccine structure. KK and GPGPG linkers joined the screened T and B-cell epitopes in a structure of multi-epitope vaccine, respectively. The LT-IIc amino acid sequence was attached to the merged epitopes at the N-terminus using the EAAAK linker to boost the immunogenicity of the vaccine. An effective vaccine should not only cause strong immune responses but should also have appropriate physicochemical properties during the development cycle. ProtParam server (<https://web.expasy.org/protparam/>) calculated the amino acid content and some of the designed vaccine's physicochemical properties such as molecular weight, theoretical pI, half-life in the mammalian reticulocytes, instability index, aliphatic index, and grand average of hydropathy (GRAVY) [28]. A standard method for determining the molecular weight of a protein, given the number of amino acids that make it up, is to multiply that number by the average weight of an amino acid, which is 110 Da [29]. The isoelectric point is the pH at which amino acid or protein loses its charge and is, therefore, unable to move in a direct current electric field. Knowing the theoretical pI of a protein can be very useful for selecting and optimizing the methods used for protein purification, including ion-exchange chromatography and isoelectric focusing electrophoresis [30]. The half-life is the time it takes for half of the protein to disappear inside the cell after synthesis [31]. The instability index of proteins indicates their stability in the test tube so that proteins with an instability index of less than 40 are predicted as stable proteins [32]. The aliphatic index of a protein is the relative volume of protein occupied by aliphatic side chains, this factor indicates the stability of the protein against heat [33]. The GRAVY is computed by dividing the sum of the hydropathic values of all amino acids by the number of residues in the sequence [34]. The GRAVY demonstrates the amphipathic nature of the proteins, the negative and positive values suggest the hydrophilic and hydrophobic nature of the designed vaccine, respectively [35].

SOLpro server (<http://scratch.proteomics.ics.uci.edu/>) predicted the protein tends to be soluble when overexpression in *E. coli* [36]. ANTIGENpro server (<http://scratch.proteomics.ics.uci.edu/>) was used to estimate the antigenicity of the vaccine, the accuracy of the server was determined to be 76% based on cross-validation experiments using the combined dataset [37]. Allergenicity and toxicity of the vaccine were determined by AllerTOP v. 2.0 and ToxinPred servers, respectively.

2.4. Secondary structure prediction

Prabi server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) used GOR4 method to predict the secondary structure of the engineered vaccine. GOR IV with a mean accuracy of 64.4% uses all possible pair frequencies within a 17 amino acid residues window [38].

2.5. Vaccine 3D modeling, refinement, and validation

3Dpro server (<http://scratch.proteomics.ics.uci.edu/>) predicted the tertiary structure of the multi-epitope vaccine [36]. 3Drefine server (<http://sysbio.mnet.missouri.edu/3Drefine/>) improved the quality of the 3D model from the previous stage. 3Drefine refining protocol includes a two-step process as follows: (i) Optimization of the hydrogen bonding network and (ii) minimization of atomic energy by integrating physics with knowledge-based force fields [39]. Model validation is an important step to detect possible errors in crude structures and to

compare the quality of models before and after the refinement process [40], for this purpose, RAMPAGE (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) and ProSA (<https://prosa.services.came.sbg.ac.at/prosa.php>) servers were used in this study. Ramachandran plot was generated using the RAMPAGE server. The most important application of the Ramachandran plot is to predict the quality of different protein structures regarding experimental methods (X-ray crystallography, NMR, and Cryo-EM). Acceptable and nonacceptable quality protein structures consisted of a collection of torsional angles in the allowed and disallowed regions, respectively [41]. The ProSA server evaluated the overall quality of the protein models in the form of a z-score. When the predicted model's z-scores are outside the normal range of native proteins, it indicates an incorrect structure [42].

2.6. Molecular docking

ClusPro 2.0 server (<https://cluspro.org>) performed molecular docking between vaccine candidate and two of the most common binding alleles in the population of Iran including HLA-A11:01 with PDB entry of 1X7Q, and HLA-DRB1_01:01 with PDB entry of 1AQD. ClusPro can discriminate the hypothetical user-created structures and run any server-compatible docking algorithms. Additional functions allow the users to modify the structure, determine the attraction and repulsion, or identify pairwise distance restrictions [43–45].

2.7. Molecular dynamics (MD) simulation

MD simulation was performed using Gromacs 5.1.5 (Groningen Machine for Chemical Simulations) software to investigate the behavior of vaccine, HLA-A11:01, and HLA-DRB1_01:01 molecules in docking structure [46]. The input structures were prepared using the amber ff99SB force field. The correct hydrogen status of histidines was established for all the histidines presented in the structure, and the disulfide bonds were defined for the protein. Then, the surface charge of the structure was neutralized by adding several sodium and chlorine ions. The protein was inserted into a layer of TIP3P water molecules at 10 Å using gmx solvate software. Energy minimization was performed on structures with the 1000-step by steepest descent method to eliminate Van der Waals interactions and to form the hydrogen bonds between water and complex molecules. In the next step, the system temperature gradually increased from 0 up to 300 K for 200 ps in constant volume and was then equilibrated in constant pressure at 1000 ps. MD simulation was performed at 27 °C (300 K) for 40 ns. Non-bonding interactions with a distance of 10 Å were calculated using the PME method. To increase the computational speed, the SHAKE algorithm was used to limit the bonds involved in the hydrogen atom.

2.8. In silico cloning

Backtranslation of the amino acid sequence of the multi-epitope vaccine was conducted using the Gene infinity server (http://www.geneinfinity.org/sms/sms_backtranslation.html). The Codon Adaptation Index (CAI) and GC content of optimized DNA were assessed by the GenScript server (<https://www.genscript.com/tools/rare-codon-analysis>). Finally, the multi-epitope vaccine sequence was inserted into the pET-30a (+) vector by the SnapGene tool.

3. Result

3.1. Prediction and screening of epitopes

Using the ProPred-1 server, which was set to default parameters, were predicted 15 CTL epitopes (9 mer) for human MHC-I alleles including HLA-A*0201, HLA-A*1101, HLA-B*3501, HLA-B*5101 and HLA-B*5301 (Table 1). The ProPred server predicted 657 HTL epitopes (9 mer) for the alleles of MHC-II, namely, DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*11:01, DRB1*13:01 and DRB1*15:01. Next, 52 epitopes capable of binding to at least four types of MHC alleles were assessed using the IFNepitope server for their ability to induce IFN- γ , the result of this step suggested that 23 HTL epitopes were capable of inducing IFN- γ (Table 2). Using the default parameters of the ABCPred server were predicted 15 linear B-cell epitopes (Table 3). The predicted CTL, HTL, and B-cell epitopes were evaluated for overlapping, antigenicity, toxicity, allergenicity, and cross-reactivity with human proteomes. It is noteworthy that among the overlapped epitopes were selected epitopes that had a higher antigenicity and improved quality of the second and three-dimensional structure of the vaccine by being in the chimeric structure. Finally, 4, 8, and 6 epitopes were selected from CTL, HTL, and B-cell epitopes, respectively.

3.2. Vaccine design and prediction of features

The final vaccine structure consists of four domains: LT-IIc, Linear B-cell, CTL, and HTL epitopes, which were placed in the vaccine construct by different linkers (Fig. 1). The ProtParam results revealed that the multi-epitope vaccine comprised of 380 amino acids. The vaccine's molecular weight was 41.65 kDa and its theoretical pI value was 9.96. The half-life of the vaccine was 30 h in mammalian reticulocytes, more than 20 h in yeast, and more than 10 h in *E. coli*. The instability index, aliphatic index, and GRAVY were predicted 22.73, 91.08, and -0.035, respectively. The SOLpro server result showed the vaccine to be soluble. The antigenicity of the final sequence (including the adjuvant) and the main sequence (without adjuvant) of vaccine were predicted using the ANTIGENpro server to be 0.584792 and 0.512677,

Table 1

Results of CTL epitope prediction. Four CTL epitopes were selected (labeled with *) for incorporation in the proposed vaccine amino acid sequence.

Pos	CTL epitope	Vaxijen score	Allergenicity	Toxicity	Cross-reactivity with human proteomes	Final decision
8	LPLVSSQCV	0.8267	Yes	No	Zero	-
56	LPFFSNVTW	1.0808	Yes	No	Zero	-
89	GVYFASTEK	0.7112	No	No	Zero	*
142	GVYHKNK	0.8264	Yes	No	Zero	-
269	YLPRTFLL	0.4532	Yes	No	Zero	-
271	QPRTFLKY	-0.5552	No	No	Zero	-
321	QPTESIVRF	-0.5551	Yes	No	Zero	-
550	GVLTESNKK	0.8797	No	No	Zero	*
712	IAIPTNFTI	0.7052	No	No	Zero	*
861	LPPLLTDEM	0.4723	Yes	No	Zero	-
1000	RLQSLQTYV	-0.2167	No	No	Zero	-
1052	FPQSAPHGV	-0.0058	No	No	Zero	-
1060	VVFLHVTYV	1.5122	No	No	Zero	*
1161	SPDVLGDI	1.573	Yes	No	Zero	-
1212	WPWYIWLGF	1.4953	Yes	No	Zero	-

Table 2
Results of HTL epitope prediction. Eight HTL epitopes were selected (labeled with *) for incorporation in the proposed vaccine amino acid sequence.

Pos	IFN- γ -inducing HTL epitope	Vaxijen score	Allergenicity	Toxicity	Cross-reactivity with human proteomes	Final decision
1	FVFLVLLPL	0.8601	No	No	Zero	*
2	VFLVLLPLV	0.4651	No	No	Zero	–
3	FLVLLPLVS	0.4266	No	No	Zero	–
4	LVLLPLVSS	0.6523	Yes	No	Zero	–
41	VFRSSVLHS	–0.3392	Yes	No	Zero	–
117	LIVNNATNV	0.1942	No	No	Zero	–
266	VGYLQPRTF	1.2441	No	No	Zero	*
317	FRVQPTESI	0.9396	Yes	No	Zero	–
366	VLYNSASFS	0.4029	Yes	No	Zero	–
406	VRQIAPGQT	0.8675	Yes	No	Zero	–
448	YNYLYRLFR	–0.8692	Yes	No	Zero	–
488	YFPLQSYGF	0.5107	Yes	No	Zero	–
504	YQPYRVVVL	0.5964	No	No	Zero	*
510	VVLSFELLH	1.409	No	No	Zero	*
575	VRDPQTEI	0.7025	No	No	Zero	*
950	VVNQNAQAL	0.4749	No	No	Zero	*
1000	LQSLQTYVT	–0.0592	No	No	Zero	–
1017	IRASANLAA	0.4455	Yes	No	Zero	–
1128	VIGIVNNTV	0.7774	Yes	No	Zero	–
1213	WYIWLGFIA	1.0356	No	No	Zero	*
1216	WLGFIAGLI	0.1544	Yes	No	Zero	–
1217	LGFIAGLIA	0.5797	No	No	Zero	*
1220	IAGLIAIVM	0.4716	No	No	Zero	–

respectively. The vaccine sequence was predicted to be non-allergenic and non-toxic using AllerTOP v. 2.0 and ToxinPred servers, respectively.

3.3. Secondary structure prediction

The Prabi server predicted that 24.47, 25.26 and 50.26% of the total 380 amino acids of vaccine construct were alpha helix, extended strand, and random coil, respectively.

3.4. Vaccine 3D modeling, refinement, and validation

Vaccine primary 3D model was predicted using the 3Dpro server. The 3Drefine server performed the process of refining the protein model. Five parameters, including 3Drefine, GDT-TS, GDT-HA, RMSD, RWplus, and MolProbity, ranked the optimized models. The higher GDT-TS, RMSD and GDT-HA scores, and the low scores of 3Drefine, MolProbity, and RWplus showed a higher quality model. Based on these factors, refined model No. 5 was selected, which is indicated in the first row of Table 4, the raw 3D structure and refined model of the vaccine are depicted in Fig. 2A and B. The validation process was performed using the Ramachandran plot and the ProSA z-score.

Ramachandran plot result revealed that in the primary model, 93.4%, 3.2%, and 3.4% of residues were located in favored, allowed, and outlier regions, respectively (Fig. 3A), while in the refined model, 93.1%, 5%, and 1.9% of residues were found in favored, allowed, and outlier regions, respectively (Fig. 3B). Z-score of the initial model was calculated –0.82 (Fig. 4A), while the Z-Score of the refined model was estimated –1.08 (Fig. 4B).

3.5. Molecular docking

The interaction of the refined model of vaccine with HLA-I and HLA-II molecules was investigated using the ClusPro 2.0 server. Proteins with PDB entry of 1X7Q and 1AQD were identified separately as receptor and vaccine as a ligand to do molecular docking. The server provided 30 clusters from each complex and ranked the clusters based on the amount of energy. Cluster No. 0.00 in the results of the interaction between the HLA-A11:01 and HLA-DRB1_01:01 molecules with the multi-epitope vaccine showed the lowest energy weighted score of –1165.4 and –1279.1 Kcal/mol, respectively (Figs. 5A and 6A). The PDBsum server was used to obtain accurate information on the interaction between residues in docking complexes. A total of 41 vaccine residues coupled with 39 residues of chain A from HLA-A11:01 molecule. A

Table 3
Results of linear B-cell epitope prediction. Six linear B-cell epitopes were selected (labeled with *) for incorporation in the proposed vaccine amino acid sequence.

Pos	linear B-cell epitope	Vaxijen score	Allergenicity	Toxicity	Cross-reactivity with human proteomes	Final decision
236	TRFQTLALHRSYLTP	0.5115	No	No	Zero	*
245	HRSYLTPGDSSSGWTA	0.6017	No	No	Zero	*
257	GWTAGAAAYVGYLQP	0.6210	No	No	Zero	*
307	TVEKGIVQTSNFRVQP	0.6733	Yes	No	Zero	–
470	TEIYQAGSTPCNGVEG	–0.0105	Yes	No	Zero	–
476	GSTPCNGVEGFNCYFP	0.2489	Yes	No	Zero	–
492	LQSYGFQPTNGVGYQP	0.5258	Yes	No	Zero	–
594	GVSVITPGTNTSNQVA	0.4651	No	No	Zero	*
648	GCLIGAEHVNNSEYCD	0.848	Yes	Yes	Zero	–
879	AGTITSGWTFGAGAAL	0.2868	No	No	Zero	–
1058	HGVVFLHVTYVPAQEK	0.8847	No	No	Zero	*
1112	PQIITDNTFVSGNCD	0.2404	Yes	No	Zero	–
1180	QKEIDRLNEVAKNLNE	0.0684	Yes	No	Zero	–
1206	YEQYIKWPWYIWLGF	0.951	No	No	Zero	*
1247	CCSCGSCCKFDEDDSE	0.1273	Yes	Yes	Zero	–

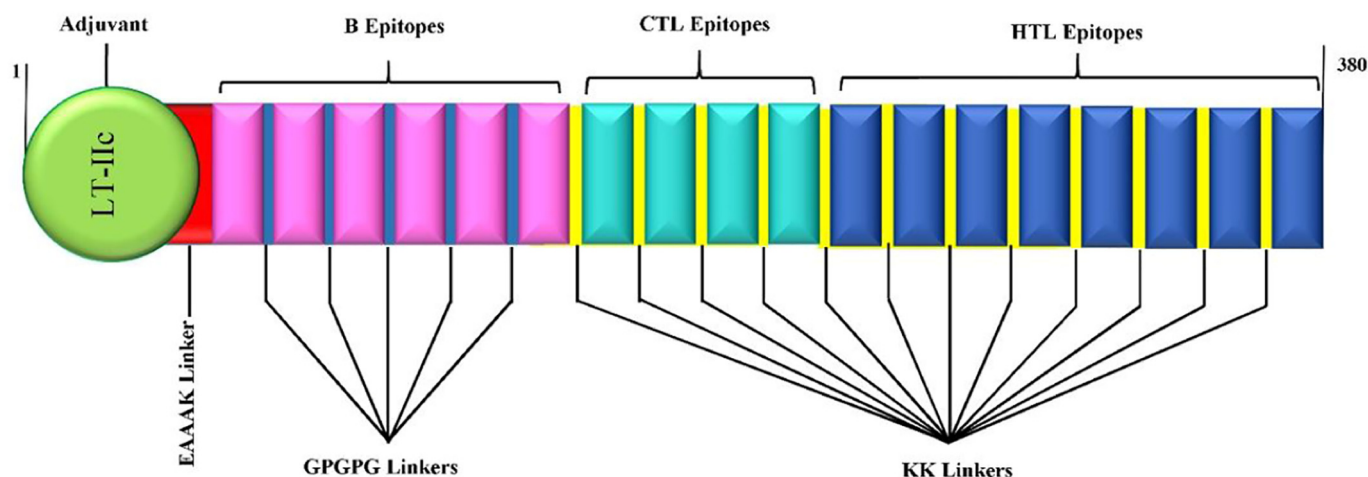


Fig. 1. Graphical visualization of the multi-epitope vaccine. The 380 amino acid sequence of vaccine contains an adjuvant (green) at the N-terminal of multi-epitope sequence, B-cell epitopes (violet), CTL epitopes (cyan), and HTL epitopes (blue). The used linkers for fuse mentioned domains together were EAAAK (red), GPGPG (blue), and KK (yellow).

Table 4

The model refinement results. The refined model No. 5 was selected based on the 3Drefine score, GDT-TS, GDT-HA, RMSD, MolProbity, and RWPlus parameters.

Model	3D ^{refine} score	GDT-TS	GDT-HA	RMSD (Å)	MolProbity	RWPlus
5	26,777.7	0.9980	0.9625	0.376	3.238	-49,406.905116
4	27,111.1	0.9993	0.9743	0.345	3.205	-49,297.629523
3	27,585.9	0.9993	0.9822	0.311	3.254	-49,016.949003
2	28,548.4	1.0000	0.9908	0.265	3.271	-48,826.479019
1	34,484.1	1.0000	0.9980	0.199	3.297	-48,423.046186

number of 18 hydrogen bonds were formed between the residues of the chain A from the HLA-A11:01 molecule and the residues of the vaccine, which included Pro105-His113, Glu53-Lys250, Glu53-His211, Asn174-Gln183, Glu166-Tyr109, Arg169-Tyr109, Lys146-Tyr361, Asp39-Lys259, Glu58-Lys27, Lys176-Gly376, Lys68-Asn112, Lys68-Ala379, Lys68-Leu377, Arg163-Ser18, Asp106-Asn98, Arg75-Ile367, Arg75-Ala368 and Thr80-Trp360 (Fig. 5B). In the vaccine-HLA-DRB1_01:01 complex 15 and 23 residues of the vaccine associated with 14 residues of the chain A and 24 residues of the chain B from the HLA-DRB1_01:01 molecule, respectively. Hydrogen bond of Thr162-Ser53 was established between the vaccine and chain A from HLA-DRB1 01:01 molecule and 14 hydrogen bonds, including Glu176-Lys4,

Glu128-Lys4, Gln92-Ala34, Gly84-Arg149, His81-Arg149, Glu87-Arg149, Glu87-Lys39, Thr21-Lys39, Glu22-Gln42, Asn19-Val41, Asn82-Gly145, Asn82-Thr141, Thr77-Thr141 and Gln70-Arg137 were formed between the chain B from HLA-DRB1 01:01 molecule and vaccine (Fig. 6B).

3.6. Molecular dynamics simulation

The best molecular docking complexes were used as inputs to MD simulation. At this stage, MD simulation of both complexes was performed for 40 ns and the results were evaluated as root mean square deviation (RMSD), Root Mean Square Fluctuation (RMSF), and radius of gyration (Rg).

RMSD between the created structures during MD simulation is an appropriate parameter to ensure the stability of the vaccine-HLA complexes. RMSD of the HLA-A11:01 molecule in the vaccine - HLA-A11:01 complex reached 0.2 nm after approximately 1000 ps and showed a slight variation around this amount until the end of the MD simulation. The RMSD of the vaccine showed a significant increase from the beginning of the MD simulation until 3500 ps, after that, there was a downward trend up to 5000, from this point to until 36,000 ps the trend was approximately increasing (Fig. 7A). The profile of RMSD changes of HLA-DRB1_01:01 molecule in the vaccine - HLA-DRB1_01:01 complex was similar to HLA-A11:01, while the rate of

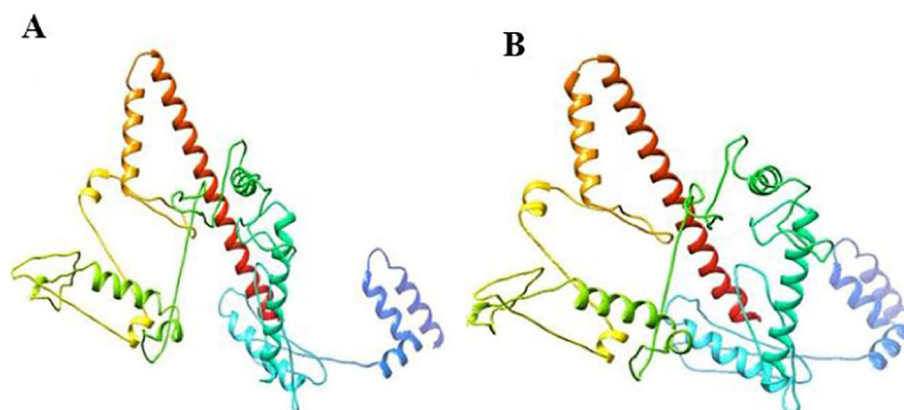


Fig. 2. Protein modeling and refinement. (A) The Three-dimensional structure of the proposed vaccine obtained from the 3Dpro server. (B) The final 3D model of the vaccine refined by the 3Drefine server.

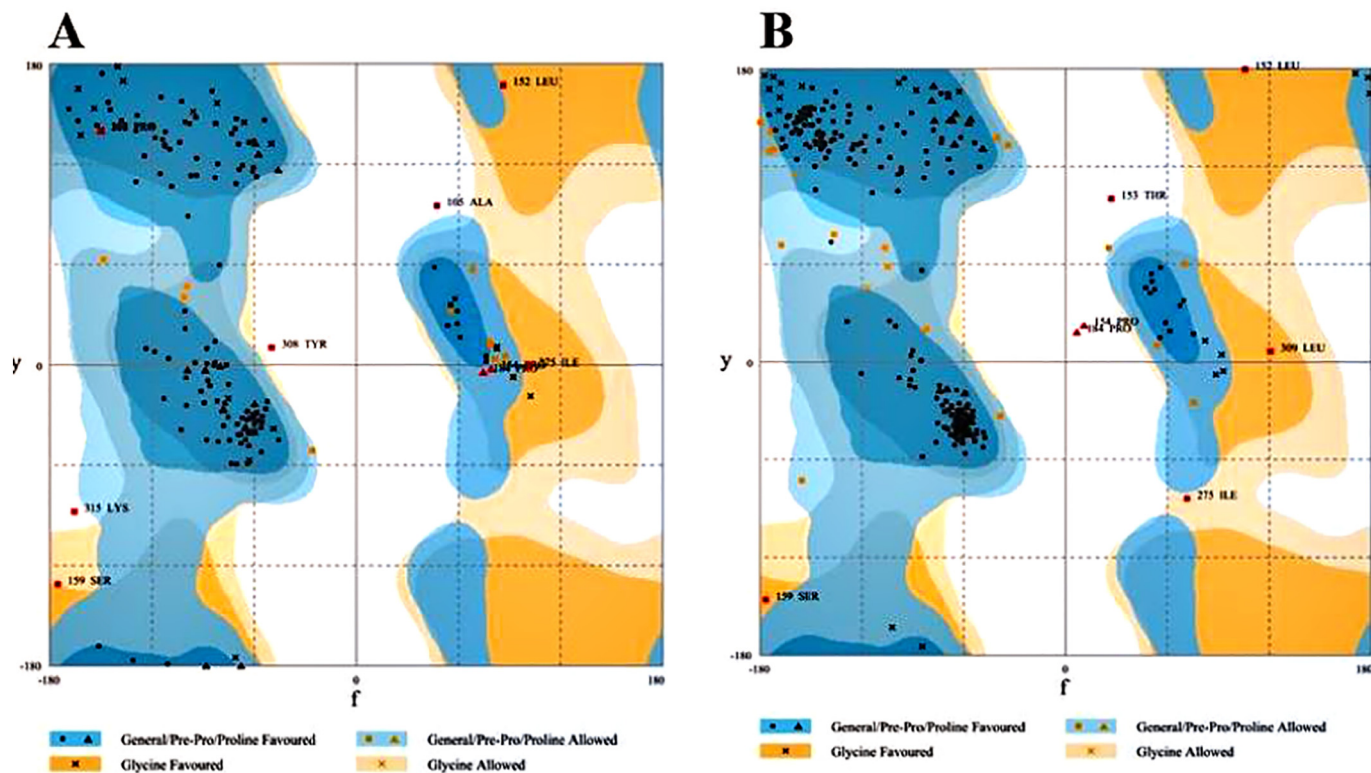


Fig. 3. 3D protein model validation using the Ramachandran plot before and after refining. (A) In the initial model, 93.4%, 3.2%, and 3.4% of residues were positioned in favored, allowed and outlier regions, respectively. (B) In the refined model 93.1%, 5%, and 1.9% of residues were observed in favored, allowed and outlier regions, respectively.

RMSD changes in the vaccine was relatively different. At the beginning of the MD simulation, the value of RMSD showed a sharp increase until 20,000 ps which reached 2 nm, of this point until the end of the MD simulation was observed no significant changes. (Fig. 7B).

RMSF determined the flexibility of any residue in the vaccine- HLA complexes. RMSF result showed that the RMSF of most amino acids in the chain A of the HLA-A11:01 molecule was high and the maximum

and minimum RMSF values were 0.17 and 0.05, respectively (Fig. 8A). Two regions on the chain B of the molecule showed relatively high fluctuations, which included regions 12–21 and 70–78. In the range of 12–21, the RMSF reached 0.23 nm (Fig. 8B). The study of vaccine structure in this complex showed high RMSF in most of its amino acids. The two regions of the vaccine showed very high levels of RMSF of 0.57 and 0.5 nm, which included amino acids in regions 65–80 and 300–320,

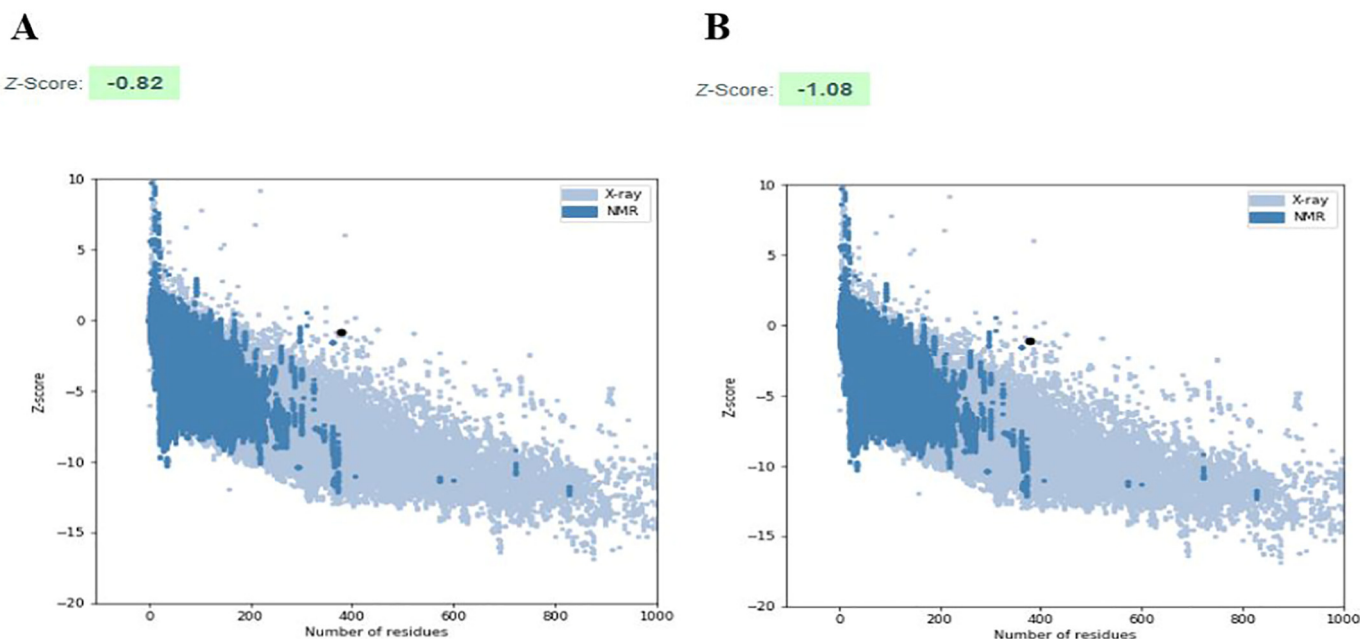


Fig. 4. ProSA-web z-score plot for primary and refined models. (A) The z-score in the primary model was -0.82 , (B) while this value was changed to -1.08 in the refined model.

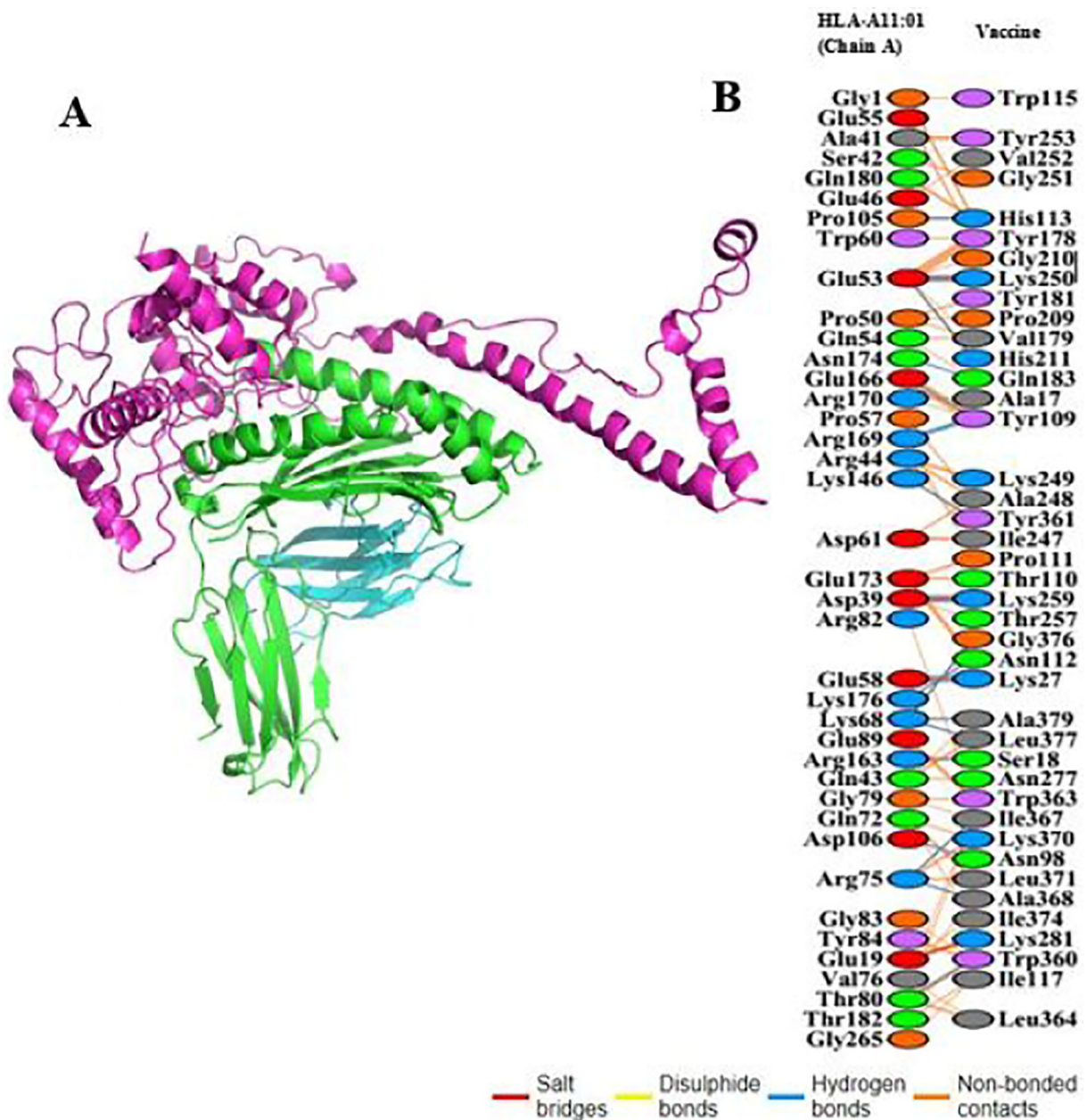


Fig. 5. Molecular docking of the vaccine with the HLA-A11:01 molecule (chain A). (A) The cartoon depictions of the vaccine-HLA-A11:01 complex illustrated using the PyMOL software. The vaccine, chain A and B of HLA-A11:01 molecule has been shown in magenta, green and cyan, respectively. (B) Bonds between residues vaccine with HLA-A11:01 molecule (chain A). A total of 41 residues of the vaccine associated with 39 residues of HLA-A11:01 molecular chain A. A number of 18 hydrogen bonds (blue line) were formed between the residues of the chain A from the HLA-A11:01 molecule and the residues of the vaccine.

respectively (Fig. 8C). Most parts in chain A of HLA-DRB1_01:01 molecule in the vaccine - HLA-DRB1_01:01 complex including amino acids in areas 32–38, 48–51, 92–100, 125–130, 154–157 and 167–171 showed high fluctuations (Fig. 9A). In chain B, the lowest and highest value of RMSF was 0.05 and 0.15 nm, respectively (Fig. 9B). The vaccine showed very high fluctuations in most areas and had a maximum RMSF of 0.55 nm (Fig. 9C).

Rg for the HLA-A11:01 molecule in the vaccine - HLA-A11:01 complex was 2.35 nm at the beginning of the simulation and showed no change from the beginning of the simulation to the end (Fig. 10A). While Rg of the HLA-DRB1_01:01 molecule in the vaccine - HLA-DRB1_01:01 complex was 2.4 nm at the beginning of the MD simulation and had a slight increase up to 1000 ps, then remained constant until the end of the MD simulation (Fig. 10B). In the presence of the HLA-A11:01 molecule, the Rg value of the vaccine was 3.7 at the beginning

of the simulation, and during the simulation period, a slight increase and decrease were observed in this parameter, and at the end of the simulation time, the Rg value reached 3.5 (Fig. 10A). The Rg level of the vaccine in the presence of the HLA-DRB1_01:01 molecule showed a relatively significant decrease from the beginning of the simulation to 17,500 ps then remained constant from that point until the end of the simulation period (Fig. 10B).

3.7. In silico cloning

The Gene infinity server conducted back translation of the vaccine construct. GenScript Tool was used to evaluating the key properties of the gene sequence including the Codon Adaptation Index (CAI) and GC content. The optimized nucleotide sequence CAI value was 1 and the GC content of the optimized sequence was 55.47% in *E. coli*.

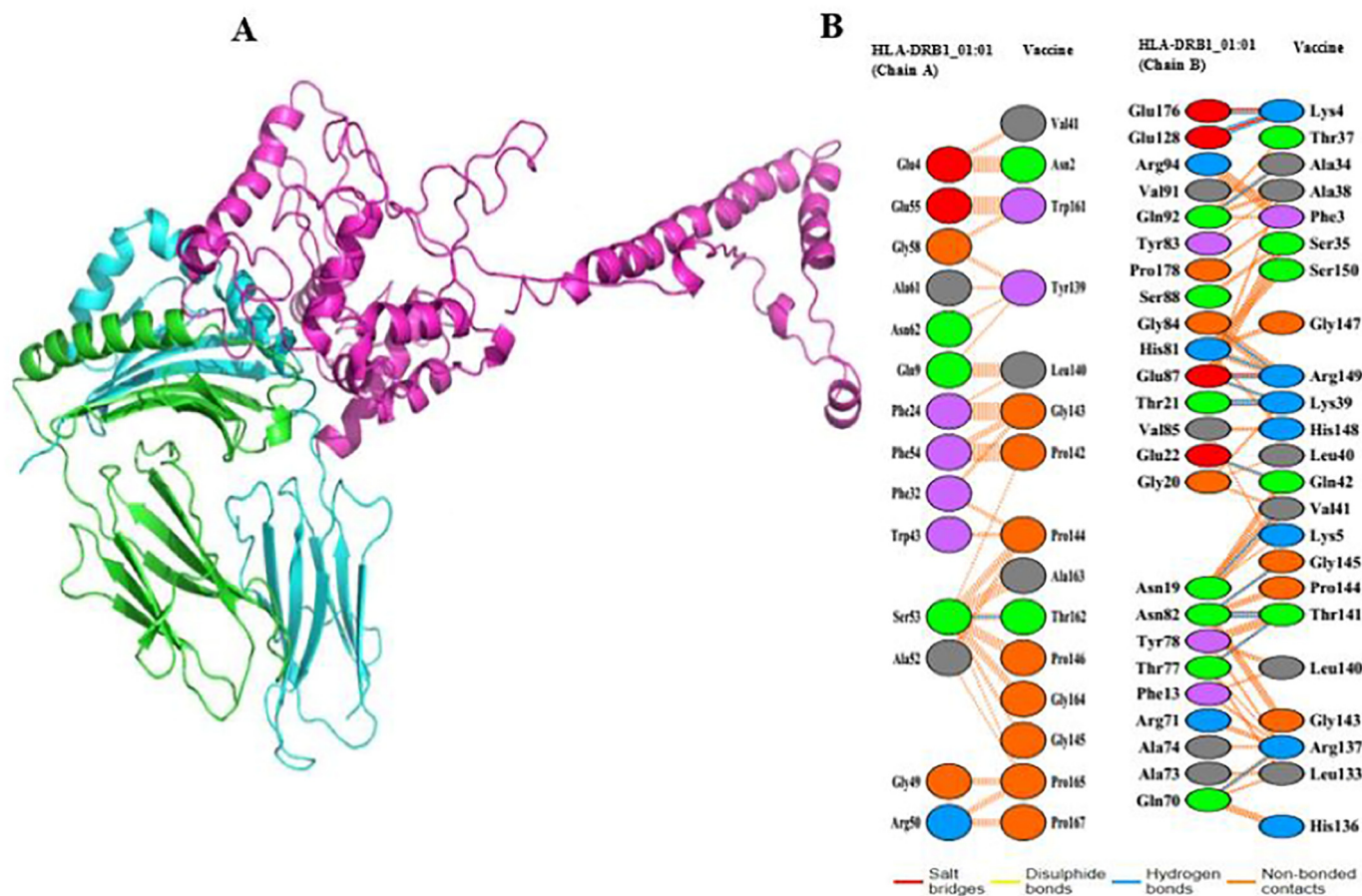


Fig. 6. Molecular docking of the vaccine with the HLA-DRB1_01:01 molecule (chain A and B). (A) The cartoon depictions of the vaccine- HLA-DRB1_01:01 complex illustrated using the PyMOL software. The vaccine, chain A and B of HLA-A11:01 molecule has been shown in magenta, green and cyan, respectively. (B) Bonds between residues vaccine with HLA-DRB1_01:01 molecule (chain A and B). A number of 15 and 23 residues of the vaccine paired with 14 residues from the chain A and 24 residues from the chain B of the HLA-DRB1_01:01 molecule, respectively. A total of 15 hydrogen bonds (blue line) were formed between the residues of the chain A and B from the HLA-DRB1_01:01 molecule and the residues of the vaccine.

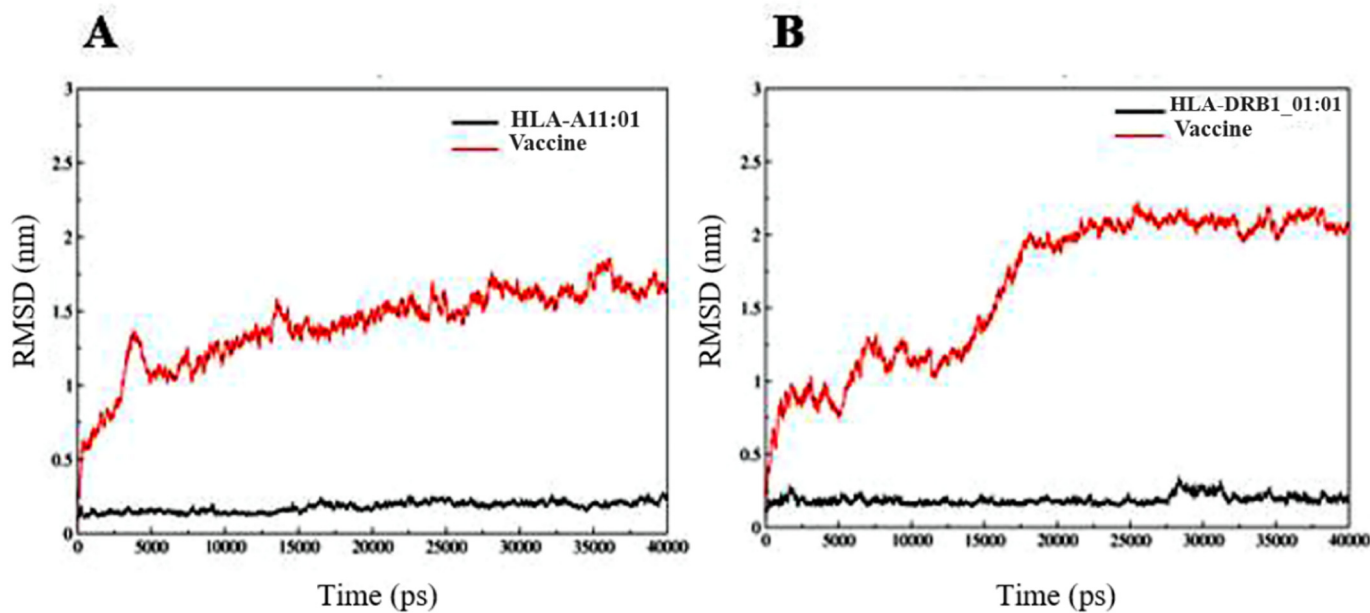


Fig. 7. RMSD plots. (A) RMSD analysis of vaccine - HLA-A11:01 complex indicated that during the MD simulation period, RMSD was approximately constant in the HLA-A11:01 molecule and increased in the vaccine. (B) RMSD analysis of vaccine - HLA-DRB1_01:01 complexes showed that RMSD value of HLA-DRB1_01:01 molecule did not change significantly until the end of the simulation time and RMSD of vaccine increased from the beginning of the simulation to 20,000 ps, then stabilized until the end of the simulation period.

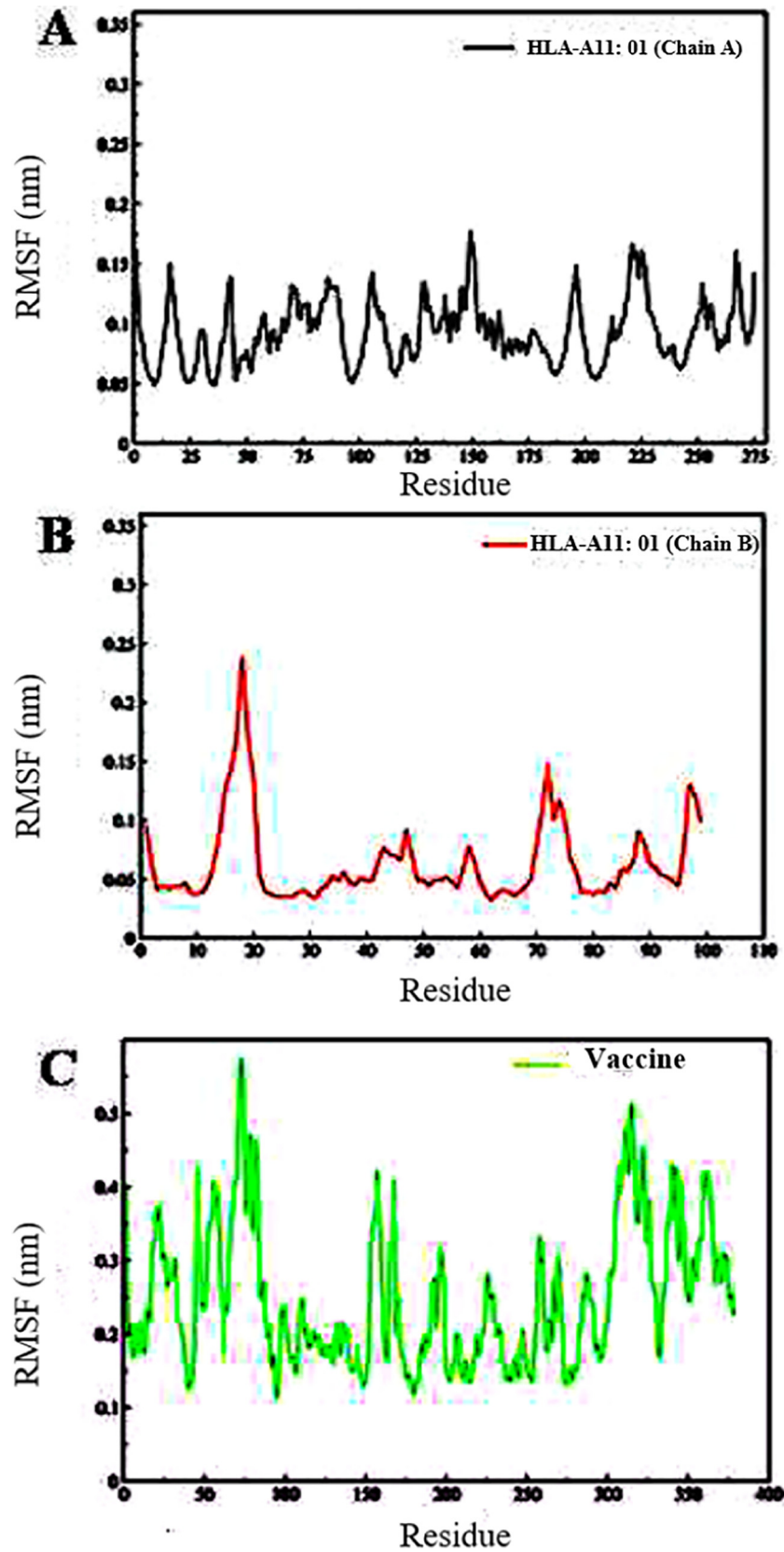


Fig. 8. RMSF diagram for amino acid (A) chain A and (B) chain B of HLA-A11:01 molecule, and (C) vaccine. There are areas with high flexibility in the peaks.

Recognition sites for *Bam*HI and *Xho*I restriction enzymes were added to 5' and 3' end of the optimized gene, respectively. The adapted codon sequence was inserted into the pET30a (+) vector using SnapGene software (Fig. 11).

4. Discussion

The COVID-19 outbreak demonstrates a pandemic threat to global public health [47]. The SARS-CoV-2 can be transmitted from person to

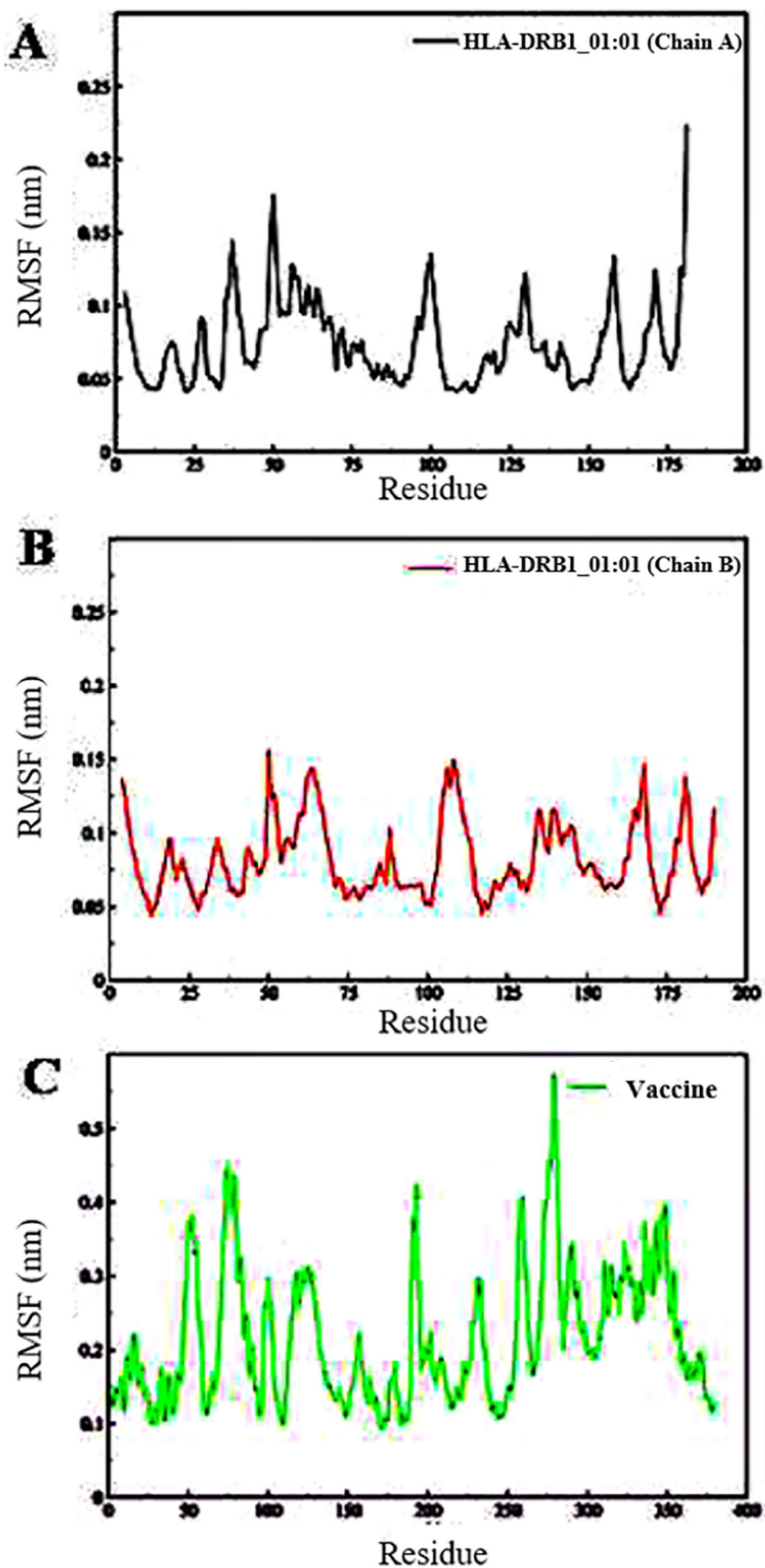


Fig. 9. RMSF diagram for amino acid (A) chain A and (B) chain B of HLA-DRB1_01:01 molecule, and (C) vaccine. There are areas with high flexibility in the peaks.

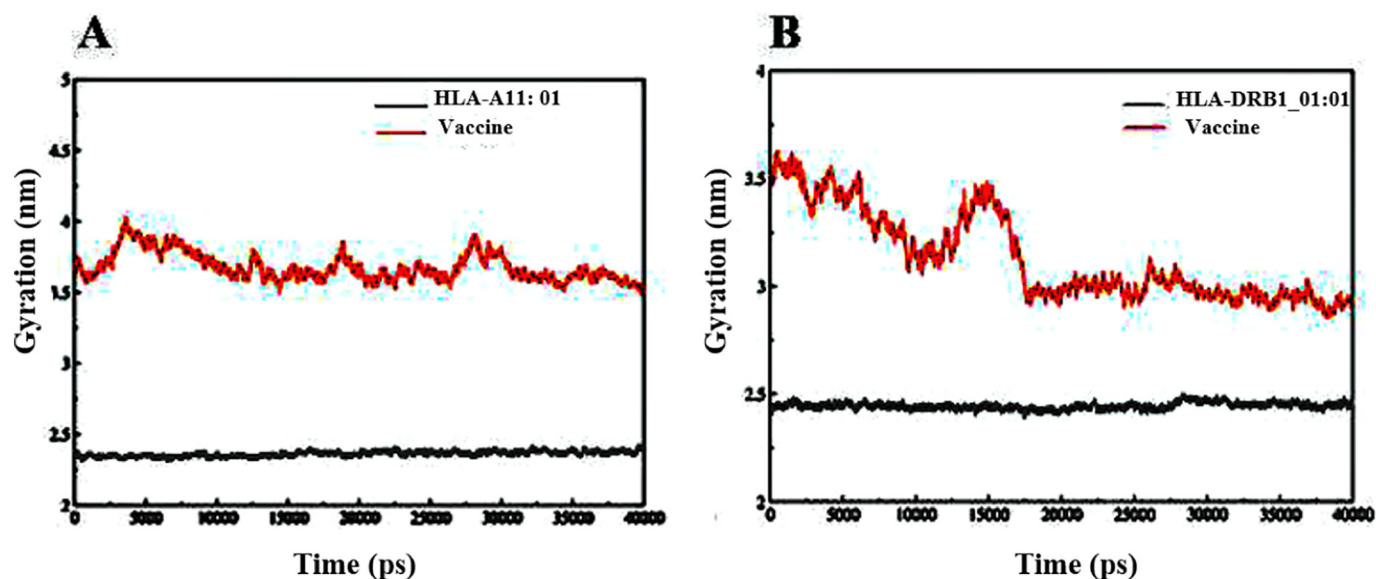


Fig. 10. Rg plot in during MD simulation. (A) The Rg value of the HLA-A11:01 molecule in the vaccine - HLA-A11:01 complex did not show much difference from the beginning to the end of the simulation. The Rg vaccine from the beginning of the simulation to its end, despite a series of small peaks during the simulation period, eventually reached the initial value. (B) The Rg value of the HLA-DRB1_01:01 molecule in the vaccine - HLA-A11:01 showed no significant change from the beginning to the end of the simulation and the Rg of vaccine revealed a relatively significant decrease from the beginning of the simulation to 17,500 ps then remained constant until the end of the simulation period.

person, even when the infected individual has no clinical symptoms, and sometimes cannot be detected in the infected individual for several days or even weeks. There are cases where a person has clinical symptoms of COVID-19, but the diagnosis test is negative, so the best way to prevent coronavirus from occurring is to get a vaccine [48], but unfortunately, because of the novelty of the virus, there is still no licensed vaccine for the disease. However, researchers and pharmacists around the world are working at an unprecedented pace to develop an effective virus vaccine, and Clinical trials of several candidate vaccines have begun, including mRNA-1273, INO-4800, and ChAdOx1 nCoV-19 [49]. Reverse vaccinology technology has been widely used to determine the epitopes for the development of multi-epitope vaccines in various organisms. In recent months, several reverse vaccinology studies have also been conducted to design the COVID-19 vaccine. In a study conducted by Enayatkhani et al. the B-cell and T epitopes of N, ORF3a, and M proteins were predicted, a multi-epitope vaccine candidate was constructed, then were performed bioinformatics evaluations [50]. Peele et al. designed a vaccine candidate using B-cell and T epitopes of S protein by *in silico* methods [51]. Kalita et al. selected N, M, and S proteins as the target antigen for the prediction of T and B-cell epitopes and designed a multi-epitope vaccine against SARS-CoV-2 [48]. Guglielmo Lucchese predicted S protein antigenic peptides from SARS-CoV-2 using immunoinformatics methods [52]. Ahmed et al. predicted a collection of B cell and T cell epitopes derived from the S and N proteins of SARS-CoV-2 [53].

In each of the studies mentioned above, different proteins have been used to predict epitopes, also, the type of bioinformatics softwares, Linkers, and many other factors are different in each study, therefore the laboratory evaluations of each of these designed vaccines show different results. This work focussed on the *in silico* design of a potential multi-epitope vaccine for COVID-19 using S protein. It is important to identify B-cell and T epitopes of the target antigens in order to develop a multi-epitope vaccine. In the current study, we predicted a large number of CTL, IFN- γ -inducing HTL, and B-cell epitopes for S protein then selected 18 high-ranked epitopes based on antigenicity, toxicity, allergenicity, and cross-reactivity with human proteomes. Linkers play a dual role in the multi-epitope vaccine structure: (i) the prevention of the generation of junctional epitopes (neoepitopes), which is the major concern in designing multi-epitope vaccines, and (ii) promotion

of the immune processing and presentation of HLA-II binding epitopes [40,54]. To design the vaccine construct, the selected T and B-cell epitopes were fused using the KK and GPGPG linkers, respectively. LT-IIc was attached to the N-terminal of the multi-epitope sequence by an EAAAK linker. The designed vaccine was evaluated for its physicochemical features, antigenicity, allergenicity, and toxicity. The vaccine's molecular weight was 41.65 kDa, which makes it an acceptable vaccine since proteins with a molecular weight of less than 110 kDa are considered to be more appropriate targets for the development of vaccines, as proteins with a lower molecular weight can be easily purified [29]. The vaccine construct was basic in nature and the total numbers of negatively and positively charged residues were 16 and 48, respectively. The vaccine structure has a half-life of 30, 20, and 10 h in mammalian, yeast, and *E. coli*, respectively, therefore the vaccine is exposed to the immune system for a relatively long time and causes more immune responses [28]. The recombinant vaccine's instability index was calculated at 22.73, and as it was less than 40, it was considered to be a stable protein [28]. The calculated aliphatic index and the GRAVY were 91.08 and -0.035 , respectively. The higher aliphatic index value suggests greater thermostability and the negative GRAVY value shows the vaccine's hydrophilic nature therefore it can show strong interaction with water molecules [55]. The sequences of vaccine (with and without adjuvant) were both antigenic, and adding the adjuvant to the N-terminal of the vaccine sequence increased the vaccine's antigenicity. Results also suggested that the vaccine was soluble, non-allergenic, and non-toxic.

To obtain a high-quality 3D structure, the initial 3D structure was subjected to the refinement process. The quality of the initial and refined models was assessed using Ramachandran plot and ProSA z-score. Ramachandran plot data showed that 3.4% of residues of the initial model was located in the outlier region while this value was reduced to 1.9% after the refinement process. In addition, the z-score of primary model was calculated -0.82 , while this parameter was determined -1.08 after refining. These results suggest that the three-dimensional structure of the vaccine has improved after refinement. ClusPro server conducted the docking process between the multi-epitope vaccine with HLA-A11:01 and HLA-DRB1_01:01 molecules, the value of lowest energy in the selected complexes of vaccine-HLA-A11:01 and vaccine-HLA-DRB1_01:01 were -1165.4 and -1279.1 , respectively, which

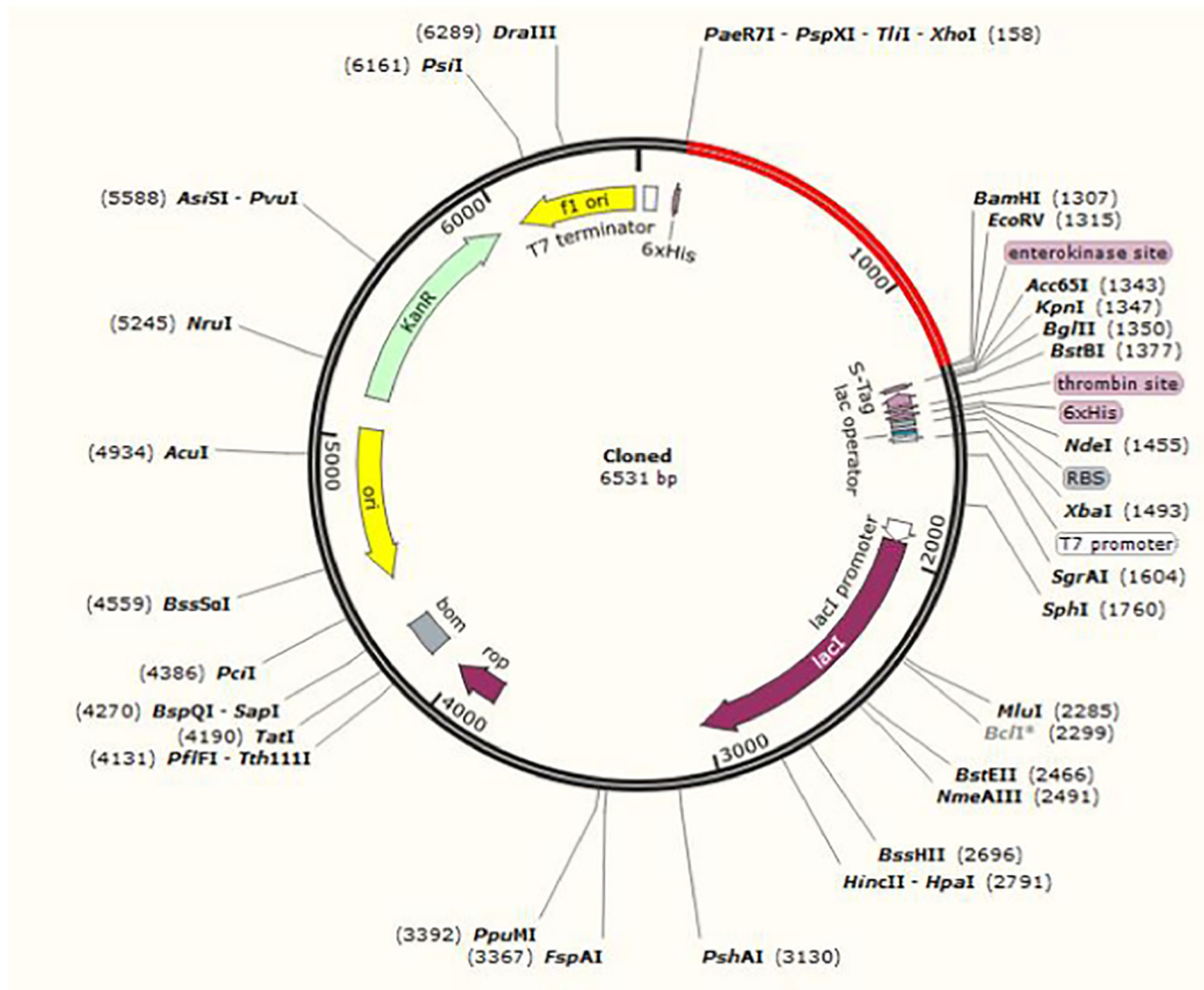


Fig. 11. *In silico* restriction cloning of the multi-epitope vaccine sequence into the pET30a (+) expression vector. The red region represents the vaccine coding gene and the black circle represents the vector backbone.

indicated that the interaction of the vaccine with the HLA-DRB1_01:01 to be stronger than its interaction with the HLA-A11:01. The results of the RMSD evaluation showed that the vaccine - HLA-DRB1_01:01 complex reached stability at 20 ns times, while the vaccine - HLA-A11:01 complex required more time to stabilize. The RMSF result suggested that the flexibility in the regions between vaccine-HLA in the both complex and the results of Rg showed that the vaccine structure in both complexes was densified during simulation, but gained more density in the vaccine-HLA-DRB1_01:01 complex. In practice, some essential factors including CAI and GC content of the gene need to be optimized to achieve an optimal level protein expression in *E. coli*. According to the results, the value of the CAI was 1, the CAI value was greater than 0.8, which suggests a higher probability of protein expression at a high level [56]. GC content was 55.47%, and the ideal percentage of GC content ranges from 30 to 70%. The results of bioinformatics evaluation of the vaccine showed that this vaccine candidate may have a high potency against SARS-CoV-2 but the *in vitro* and *in vivo* studies are needed to confirm these results.

5. Conclusion

The purpose of this research was to develop a COVID-19 vaccine using bioinformatics approaches. To achieve this aim were predicted T and B-cell epitopes of S protein and then screened for antigenicity, toxicity, allergenicity, and cross-reactivity with human proteomes. The eighteen high-ranked epitopes and LT-IIc were incorporated into the

final construct. After evaluating the physicochemical properties of the vaccine construct, its secondary and three-dimensional structure was predicted. Structural quality, binding affinity to MHC-I and II, and stability of the docked complexes were assessed. Finally, the expression level of the vaccine in *E. coli* strain K12 was evaluated. Significant results were reported in this research, experimental confirmation of the engineered vaccine required to further assess the efficacy of this vaccine candidate.

Funding

None.

Declaration of competing interest

The authors have declared no competing interest.

Acknowledgments

We would like to thank Dr. Shahram Tahmasebian and Dr. Sorayya Ghasemi for their valuable feedback and suggestions.

References

- [1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (10223) (2020) 497–506.

- [2] C. Chakraborty, A.R. Sharma, M. Bhattacharya, G. Sharma, S.-S. Lee, The 2019 novel coronavirus disease (COVID-19) pandemic: a zoonotic prospective, *Asian Pac. J. Trop. Biomed.* 13 (6) (2020) 242.
- [3] C. Chakraborty, A. Sharma, G. Sharma, M. Bhattacharya, S. Lee, SARS-CoV-2 causing pneumonia-associated respiratory disorder (COVID-19): diagnostic and proposed therapeutic options, *Eur. Rev. Med. Pharmacol. Sci.* 24 (7) (2020) 4016–4026.
- [4] A.H. de Wilde, E.J. Snijder, M. Kikkert, M.J. van Hemert, Host factors in coronavirus replication, *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*, Springer 2017, pp. 1–42.
- [5] P.C. Woo, S.K. Lau, C.S. Lam, C.C. Lau, A.K. Tsang, J.H. Lau, R. Bai, J.L. Teng, C.C. Tsang, M. Wang, Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus, *J. Virol.* 86 (7) (2012) 3995–4008.
- [6] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (2020) 727–733.
- [7] P. Yang, X. Wang, COVID-19: a new challenge for human beings, *Cell Mol Immunol* (2020) 1–3.
- [8] M.A. Marra, S.J. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S. Butterfield, J. Khattra, J.K. Asano, S.A. Barber, S.Y. Chan, The genome sequence of the SARS-associated coronavirus, *Science* 300 (5624) (2003) 1399–1404.
- [9] E.J. Snijder, P.J. Bredenbeek, J.C. Dobbe, V. Thiel, J. Ziebuhr, L.L. Poon, Y. Guan, M. Rozanov, W.J. Spaan, A.E. Gorbalenya, Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage, *J. Mol. Biol.* 331 (5) (2003) 991–1004.
- [10] S. Belouzard, J.K. Millet, B.N. Licitra, G.R. Whittaker, Mechanisms of coronavirus cell entry mediated by the viral spike protein, *Viruses* 4 (6) (2012) 1011–1033.
- [11] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* 367 (6483) (2020) 1260–1263.
- [12] C. Chakraborty, A.R. Sharma, M. Bhattacharya, G. Sharma, S.S. Lee, G. Agoramoorthy, Consider TLR5 for new therapeutic development against COVID-19, *J. Med. Virol.* (2020) <https://doi.org/10.1002/jmv.25997>.
- [13] H. Zhang, G. Wang, J. Li, Y. Nie, X. Shi, G. Lian, W. Wang, X. Yin, Y. Zhao, X. Qu, Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies, *J. Virol.* 78 (13) (2004) 6938–6945.
- [14] S. Plotkin, History of vaccination, *Proc. Natl. Acad. Sci.* 111 (34) (2014) 12283–12287.
- [15] G.A. Poland, I.G. Ovsyannikova, R.M. Jacobson, Application of pharmacogenomics to vaccines, *Pharmacogenomics* 10 (2009) 837–852.
- [16] R. Rappuoli, Reverse vaccinology, *Curr. Opin. Microbiol.* 3 (5) (2000) 445–450.
- [17] A. Saha, A.R. Sharma, M. Bhattacharya, G. Sharma, S.-S. Lee, C. Chakraborty, Probable molecular mechanism of remdesivir for the treatment of COVID-19: need to know more, *Arch. Med. Res.* (2020) <https://doi.org/10.1016/j.arcmed.2020.05.001>.
- [18] H.F. Nawar, N.D. King-Lyons, J.C. Hu, R.C. Pasek, T.D. Connell, LT-IIc, a new member of the type II heat-labile enterotoxin family encoded by an *Escherichia coli* strain obtained from a nonmammalian host, *Infect. Immun.* 78 (11) (2010) 4705–4713.
- [19] H. Singh, G. Raghava, ProPred1: prediction of promiscuous MHC class-I binding sites, *Bioinformatics* 19 (8) (2003) 1009–1014.
- [20] H. Singh, G. Raghava, ProPred: prediction of HLA-DR binding sites, *Bioinformatics* 17 (12) (2001) 1236–1237.
- [21] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, *Proteins* 65 (1) (2006) 40–48.
- [22] I.A. Doytchinova, D.R. Flower, Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC bioinformatics* 8 (1) (2007) 4.
- [23] I.A. Doytchinova, D.R. Flower, Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties, *Vaccine* 25 (5) (2007) 856–866.
- [24] I.A. Doytchinova, D.R. Flower, Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines, *Open Vaccine J* 1 (1) (2008) 4.
- [25] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P. Raghava, O.S.D.D. Consortium, *In silico* approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (9) (2013).
- [26] S. Wold, J. Jonsson, M. Sjöström, M. Sandberg, S. Rännar, DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures, *Anal. Chim. Acta* 277 (2) (1993) 239–253.
- [27] C. Chen, Z. Li, H. Huang, B.E. Suzek, C.H. Wu, U. Consortium, A fast peptide match service for UniProt knowledgebase, *Bioinformatics* 29 (21) (2013) 2808–2809.
- [28] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, *The Proteomics Protocols Handbook*, Springer 2005, pp. 571–607.
- [29] D. Barh, N. Barve, K. Gupta, S. Chandra, N. Jain, S. Tiwari, N. Leon-Sicaire, A. Canizalez-Roman, A.R. dos Santos, S.S. Hassan, Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by Piper betel derived compounds, *PLoS One* 8 (1) (2013), e52773.
- [30] M.R. Pergande, S.M. Cologna, Isoelectric point separations of peptides and proteins, *Proteomes* 5 (1) (2017) 4.
- [31] A. Ciechanover, A.L. Schwartz, How are substrates recognized by the ubiquitin-mediated proteolytic system, *Trends Biochem. Sci.* 14 (12) (1989) 483–488.
- [32] K. Guruprasad, B.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, *Protein Eng Des Sel* 4 (2) (1990) 155–161.
- [33] A. Ikai, Thermostability and aliphatic index of globular proteins, *J. Biochem.* 88 (6) (1980) 1895–1898.
- [34] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* 157 (1) (1982) 105–132.
- [35] P. Kalita, D.L. Lyngdoh, A.K. Padhi, H. Shukla, T. Tripathi, Development of multi-epitope driven subunit vaccine against *Fasciola gigantica* using immunoinformatics approach, *Int. J. Biol. Macromol.* 138 (2019) 224–233.
- [36] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, SCRATCH: a protein structure and structural feature prediction server, *Nucleic Acids Res.* 33 (suppl_2) (2005) W72–W76.
- [37] C.N. Magnan, M. Zeller, M.A. Kayala, A. Vigil, A. Randall, P.L. Felgner, P. Baldi, High-throughput prediction of protein antigenicity using protein microarray data, *Bioinformatics* 26 (23) (2010) 2936–2943.
- [38] J. Garnier, GOR secondary structure prediction method version IV, *Meth Enzym, RF Doolittle Ed* 266 (1998) 540–553.
- [39] N. Kalisman, A. Levi, T. Maximova, D. Reshef, S. Zafriri-Lynn, Y. Gleyzer, C. Keasar, MESH: a new library of Java classes for molecular modeling, *Bioinformatics* 21 (20) (2005) 3931–3932.
- [40] N. Nezafat, Z. Karimi, M. Eslami, M. Mohkam, S. Zandian, Y. Ghasemi, Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches, *Comput. Biol. Chem.* 62 (2016) 82–95.
- [41] C. Ramakrishnan, Ramachandran and his map, *Resonance* 6 (10) (2001) 48.
- [42] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (suppl_2) (2007) W407–W410.
- [43] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padohny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein–protein docking, *Nat. Protoc.* 12 (2) (2017) 255.
- [44] S. Vajda, C. Yueh, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, D. Kozakov, New additions to the ClusPro server motivated by CAPRI, *Proteins* 85 (3) (2017) 435–444.
- [45] D. Kozakov, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, S. Vajda, How good is automated protein docking? *Proteins* 81 (12) (2013) 2159–2166.
- [46] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1 (2015) 19–25.
- [47] J. Zheng, SARS-CoV-2: an emerging coronavirus that causes a global threat, *Int. J. Biol. Sci.* 16 (10) (2020) 1678.
- [48] P. Kalita, A. Padhi, K.Y. Zhang, T. Tripathi, Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2, *Microb. Pathog.* (2020) 104236.
- [49] Y.-F. Tu, C.-S. Chien, A.A. Yarmishyn, Y.-Y. Lin, Y.-H. Luo, Y.-T. Lin, W.-Y. Lai, D.-M. Yang, S.-J. Chou, Y.-P. Yang, A review of SARS-CoV-2 and the ongoing clinical trials, *Int. J. Mol. Sci.* 21 (7) (2020) 2657.
- [50] M. Enayatkhani, M. Hasaniyazad, S. Faezi, H. Guklani, P. Davoodian, N. Ahmadi, M.A. Einakian, A. Karmostaji, K. Ahmadi, Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an *in silico* study, *J. Biomol. Struct. Dyn.* (2020) 1–16.
- [51] K.A. Peele, T. Srihansa, S. Krupanidhi, A.V. Sai, T. Venkateswarulu, Design of multi-epitope vaccine candidate against SARS-CoV-2: a *in-silico* study, *J. Biomol. Struct. Dyn.* (2020) 1.
- [52] G. Lucchese, Epitopes for a 2019-nCoV vaccine, *Cell Mol Immunol* (2020) 1–2.
- [53] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, *Viruses* 12 (3) (2020) 254.
- [54] B. Meza, F. Ascencio, A.P. Sierra-Beltrán, J. Torres, C. Angulo, A novel design of a multi-antigenic, multistage and multi-epitope vaccine against *Helicobacter pylori*: an *in silico* approach, *Infect. Genet. Evol.* 49 (2017) 309–317.
- [55] M. Ali, R.K. Pandey, N. Khatoun, A. Narula, A. Mishra, V.K. Prajapati, Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection, *Sci. Rep.* 7 (1) (2017) 1–13.
- [56] A. Carbone, A. Zinovyev, F. Képes, Codon adaptation index as a measure of dominating codon bias, *Bioinformatics* 19 (16) (2003) 2005–2015.