

RESEARCH

Open Access



Development of a novel multi-epitope vaccine against the pathogenic human polyomavirus V6/7 using reverse vaccinology

Reza Salahlou^{1,2}, Safar Farajnia^{1,3*}, Nasrin Bargahi¹, Nasim Bakhtiyari^{2,3}, Faranak Elmi⁴, Mehdi Shahgolzari³, Steven Fiering⁵ and Srividhya Venkataraman⁶

Abstract

Background Human polyomaviruses contribute to human oncogenesis through persistent infections, but currently there is no effective preventive measure against the malignancies caused by this virus. Therefore, the development of a safe and effective vaccine against HPyV is of high priority.

Methods First, the proteomes of 2 polyomavirus species (HPyV6 and HPyV7) were downloaded from the NCBI database for the selection of the target proteins. The epitope identification process focused on selecting proteins that were crucial, associated with virulence, present on the surface, antigenic, non-toxic, and non-homologous with the human proteome. Then, the immunoinformatic methods were used to identify cytotoxic T-lymphocyte (CTL), helper T-lymphocyte (HTL), and B-cell epitopes from the target antigens, which could be used to create epitope-based vaccine. The physicochemical features of the designed vaccine were predicted through various online servers. The binding pattern and stability between the vaccine candidate and Toll-like receptors were analyzed through molecular docking and molecular dynamics (MD) simulation, while the immunogenicity of the designed vaccines was assessed using immune simulation.

Results Online tools were utilized to forecast the most optimal epitope from the immunogenic targets, including LTA_g, VP1, and VP1 antigens of HPyV6 and HPyV7. A multi-epitope vaccine was developed by combining 10 CTL, 7 HTL, and 6 LBL epitopes with suitable linkers and adjuvant. The vaccine displayed 98.35% of the world's population coverage. The 3D model of the vaccine structure revealed that the majority of residues (87.7%) were located in favored regions of the Ramachandran plot. The evaluation of molecular docking and MD simulation revealed that the constructed vaccine exhibits a strong binding (-1414.0 kcal/mol) towards the host's TLR4. Moreover, the vaccine-TLR complexes remained stable throughout the dynamic conditions present in the natural environment. The immune simulation results demonstrated that the vaccine design had the capacity to elicit robust immune responses in the host.

Conclusion The multi-parametric analysis revealed that the designed vaccine is capable of inducing sustained immunity against the selected polyomaviruses, although further in-vivo investigations are needed to verify its effectiveness.

*Correspondence:

Safar Farajnia
farajnias@tbzmed.ac.ir

Full list of author information is available at the end of the article



© The Author(s) 2024. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Human polyomavirus, Vaccine, Immunoinformatics, Molecular docking, Immunotherapy

Background

Polyomaviruses are small, nonenveloped DNA viruses, which are widespread in nature. Non-enveloped polyomaviruses are capable of infecting mammals and birds with their small circular double-stranded DNA genomes of approximately 5.0 kbp [1, 2]. Two major regulatory proteins are encoded by PyVs, the large tumor antigen (LT-ag) and the small tumor antigen (sT-ag), as well as several structural proteins (VP1 and VP2) [3]. Heterologous animal models indicate that PyVs may carry strong oncogenes, that contribute to cancer in humans. Regulatory proteins are important in viral replication and transcription early in the infection cycle, while structural proteins participate in capsid formation later [4].

Among 12 identified human polyomaviruses (HPyVs), six strains are involved in human diseases, especially in different human cancers. HPyV6, HPyV7, Merkel cell polyomavirus (MCPyV), trichodysplasiaspinulosa virus (TSPyV), HPyV9, MWPyV and BK virus (BKV) and JC virus (JCV), as well as newly identified viruses such as KI (KIPyV), WU (WUPyV), are the most commonly identified strains of human polyomavirus (HPyV) [5–7].

The World Health Organization reported increased incidence of skin cancer in the past few decades, with about 8,500 new cases of skin cancer reported daily in the United States. The relationship between HPyV6 and

HPyV7 in human skin cancer has recently been elucidated, in recent decade [7, 8]. Several studies have been conducted to confirm the presence of HPyV6- and HPyV7-DNA in cutaneous (Table 1) and non-cutaneous (Table 2) malignancies, including malignant melanoma (MM) [9], non-melanoma skin cancer tissues, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Studies on keratoacanthoma and trichoblastoma revealed the presence of HPyV6 in tumors [10]. HPyV7-DNA was found to be detected more frequently in non-cutaneous cancers compared to HPyV6-DNA, whereas HPyV6-DNA was more commonly observed in skin malignancies. A study concluded that all age groups and genders are infected with HPyV6 and 7, and 52–93% of humans are seropositive for HPyV6 while 33–84% are seropositive for HPyV7 [7]. PCR investigations demonstrated HPyV6 and HPyV7 DNA from the skin of both healthy individuals and those experiencing different types of skin tumors. Small and large T antigens are encoded by polyomaviruses, making them potentially oncogenic [11]. Severe cases of HPyV6 and 7 infections are associated with skin disorders, characterized by a significant increase in viral load, expression of dyskeratotic keratinocytes, and the presence of encapsidated virions observed through electron microscopy and sequencing [12]. Moreover, HPyV6 was identified in various forms of epithelial neoplasms,

Table 1 HPyV-6 and HPyV-7 seroprevalence in primary cutaneous malignancies

Cancer type	HPyV6 Positive	HPyV7 Positive	Reference
Verrucous keratosis	75%	75%	[14]
Squamous cell carcinoma in chronic lymphocytic leukemia cases who had bone marrow transplantation	33.3%	Not determined	[15]
Squamous cell carcinoma	36.4%	54.5%	[14]
Squamous cell carcinoma	100%	100%	[9]
Keratoacanthoma	42.3%	Not applicable	[10]
Trichoblastoma	22.2%	Not applicable	[10]
Cutaneous T-cell lymphomas	20%	8.6%	[16]

(Adapted from [7])

Table 2 HPyV6 and HPyV7 prevalence in non-cutaneous human malignancies

Tumor type	HPyV6 Positive	HPyV7 Positive	Reference
Laryngeal squamous cell carcinoma	14.3%	Not detected	[17]
Thymic epithelial tumors	Not detected	62.2%	[18]
Cholangiocarcinoma	9.5%	45.2%	[19]

(Adapted from [7])

while HPyV7 was observed in thymic epithelial tumors. These discoveries imply that HPyV6 and HPyV7 could be crucial contributors to the development of inflammatory skin conditions and may also possess oncogenic properties [13]. These findings indicate that HPyV6 and HPyV7 may have a crucial involvement in the development of inflammatory skin conditions and potentially possess oncogenic properties [13].

It has shown that HPyV6 and 7 bind, and inactivate p53 resulting in tumor progression. P53 suppresses tumor growth by regulating gene expression in response to stressors, such as DNA damage, leading to apoptosis and cell cycle arrest [20]. The transactivation domain of P53 is repressed by interaction with LT-antigen preventing its binding to DNA, eventually leading to cancer in humans [21, 22].

The lack of vaccine against HPyVs might be related to the complexity of the virus and the ability of HPyVs to evade the host immune system by various strategies, such as downregulating the expression of major histocompatibility complex (MHC) molecules, interfering with interferon signaling, and modulating the activity of immune cells [23]. Multi-epitope vaccines contain multiple antigenic fragments (epitopes) from different proteins of the target pathogen. This type of vaccine has several advantages over traditional vaccines, such as inducing a broader and more robust immune response, reducing the risk of antigenic escape and cross-reactivity, and facilitating the production and delivery of the vaccine. Hence, a multi-epitope vaccine against HPyVs could be beneficial for preventing or treating HPyV-associated diseases, especially in immunocompromised patients who are more susceptible to viral reactivation and complications [24].

In recent years, cancer vaccines have shown promising results against different cancers but vaccine development using traditional approaches is complex and requires a significant amount of effort [25]. Compared to traditional laboratory approaches, immunoinformatics enables rapid development of a multi-epitope vaccine, increasing efficiency and reducing costs [26, 27]. Epitope-based peptide vaccines demonstrated effectiveness in providing protective immunity against various viruses including Zika, dengue, SARS-CoV-2, and Coxsackie B viruses [28]. Hence, it is supposed that a peptide-based vaccine against oncogenic polyomavirus could provide an efficient protective vaccine against HPyV6/V7 oncogenic virus strains.

To develop a multi-epitope vaccine, CTL, HTL, and LBL epitopes of the HPyV6 and HPyV7 oncoproteins including large T antigen (LTAg), VP1, and VP2 were identified and the vaccine's stability and effectiveness were analyzed by immunoinformatics methods. The

study yielded compelling evidence supporting the likelihood that the multi-epitope vaccine can effectively initiate a strong anti-HyPV immune response.

Methods

Retrieval and analysis of protein sequences

Human polyomaviruses 6 and 7 were obtained from the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>) [29]. Complete amino acid sequence retrieval was performed from the UniProt database (<https://www.uniprot.org/>) [30] in FASTA format. In addition, the VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) was used to evaluate the antigenicity using the default threshold for viruses. VaxiJen 2.0 server is based on auto and cross-covariance (ACC) transformation methods with 70–89% accurate prediction [31, 32]. Then, the AllergenFP v1.0 server (<https://ddg-pharmfac.net/AllergenFP/>) was employed to find the allergenicity of the proteins. This server uses a novel alignment-free descriptor-based fingerprint approach that produces 88.9% accuracy in the prediction result [33]. TMHMM v2.0 server (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>), based on the hidden Markov model (HMM), was utilized for transmembrane (TM) helix prediction [34]. The next step of research included structural proteins that are non-allergenic, antigenic, and display less TM helices.

Identification and evaluation of Cytotoxic T-Lymphocyte (CTL) epitopes

The important role of cytotoxic T lymphocytes (CTLs) in the host defense mechanism is known [35]. CTLs have a receptor called CD8, which attaches to a molecule called MHC class I on the surface of infected cells. This enables them to deliver molecules that destroy the infection [36]. The NetCTL v1.2 server (<https://services.healthtech.dtu.dk/services/NetCTL-1.2/>) using weight matrix and artificial neural networks is highly efficient for epitopes prediction of 9-mer CTLs against 12 supertypes including A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62. Using a threshold value of 0.90 to maintain a specificity and sensitivity of 0.98 and 0.74 respectively, this server was used to predict CT epitopes with high combination scores among the obtained protein sequences [37]. The MHC-I binding tool of the IEDB resource was utilized to determine MHC-I binding alleles for each CTL epitope dependent on the CONSENSUS method (<http://tools.iedb.org/mhci/>) [38]. To characterize antigenicity for individual CTL epitopes of the VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [31], the allergenic profile of the AllerTOP v2.0 server (<https://www.ddg-pharmfac.net/AllerTOP/>)

[39], toxicity prediction of ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred/>) [40], and immunogenicity of IEDB Class I Immunogenicity tool (<http://tools.iedb.org/immunogenicity/>) [41] were used respectively. To separate allergens from non-allergens with a prediction accuracy of 85.3% in five-fold cross-validation, the AllerTop v2.0 server was used, which applies the amino acid descriptors, ACC transformation methods, and k-nearest neighbor (kNN) methods [39]. The ToxinPred server is used for evaluation properties of different peptides using support-vector machines (SVM), that is a machine learning approach combined with a quantitative matrix for toxicity prediction [40]. To confirm whether a specific epitope elicits an immune response or not, immunogenicity prediction was performed. CTL epitopes were utilized for the vaccine construction that were antigenic, non-allergenic, non-toxic, and immunogenic and showed high C-scores.

Identification and analysis of HTL (Helper T-Lymphocyte) epitopes

Antigen recognition by Helper T cells activates CTLs and B cells to eliminate infectious pathogens [35]. We used the MHC class II binding website of the IEDB resource (<http://tools.iedb.org/mhcii/>) to predict 15-mer HTL epitopes from the target protein sequences [42]. We also used the CONSENSUS method for predicting protein binding alleles with percentile rank threshold ≤ 2 to maintain consistency [43]. To predict characteristics of individual HTL epitopes we used the VaxiJen v2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [31], ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred/>) [40], and AllergenFP v1.0 server (<https://ddg-pharmfac.net/AllergenFP/>), respectively [33]. HTL epitopes were assessed based on their non-toxicity, antigenicity, and non-allergenicity, and considering their cytokine induction ability. IFN- γ plays a major role in inhibiting virus replication by stimulating immune responses of natural killer cells and macrophages, as well as by increasing T cell responses [44]. Interferon-gamma (IFN- γ) prediction was performed by the IFNepitope tool (<http://crdd.osdd.net/raghava/ifnepitope/predict.php>) with 81.39% precision. We applied the IL4pred tool (<https://webs.iitd.edu.in/raghava/il4pred/>) with a threshold score of 0.2 to evaluate the induction of interleukin 4 (IL-4). This operation was executed using SVM-based methods with 75.76% accuracy [45]. Then, the ability to induce both cytokines was prioritized in the selection of HTL epitopes for vaccine construction. For the proteins that were without cytokine-inducing functions, we prioritized the IFN- γ and IL4-inducing abilities of the HTL epitopes [46].

Identification and analysis of LBL (Linear B-Lymphocyte) epitopes

LBL epitopes are important for the induction of B lymphocytes to generate antibodies and has a important role in vaccine design. The ABCpred tool (<http://crdd.osdd.net/raghava/abcpred/>), based on recurrent neural network with a 0.51 threshold value, was utilized to estimate LBL from the selected protein sequences [47, 48]. LBLepitopes with scores > 0.8 were chosen as vaccine candidates. The AllerTOP v2.0 tool (<https://www.ddg-pharmfac.net/AllerTOP/>) [49], ToxinPred tool (<http://crdd.osdd.net/raghava/toxinpred/>) [40], and VaxiJen v2.0 tool (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [31] were used to assess the anticipated linear B-lymphocyte epitopes' allergenic, toxic, and antigenic profiles respectively.

Evaluation of the human homology and epitope conservancy

We used the "epitope conservancy analysis" server (<http://tools.iedb.org/conservancy/>) of the IEDB resource to analyze the conservation of selected MHC class I/II epitopes. This feature demonstrates the availability of the epitope in a range of various strains. In the conservancy analysis, epitopes with 100% maximum identity were selected for vaccine construct [50]. Epitope homology with the human proteome was investigated to avoid cross reaction with human proteins or weak response due to tolerance, and non-homologous epitopes were selected. Human homology was determined using the protein BLAST module of the BLAST server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) with Homo sapiens (taxid: 9606), and a threshold of 0.05. Non-homologous peptides where no hits were found below the threshold e-value were designated as epitopes [51, 52].

Molecular docking and peptide modeling analysis

To assess the binding ability, selective MHC-I epitopes were docked with related HLA alleles. For modeling the determined CTL epitopes, we used the PEP-FOLD v3.5 tool (<https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>). This tool utilizes the Taboo/Backtract sampling algorithm for the prediction of peptide 3D conformations with 5–50 residues [53]. After predicting five possible structures through this tool for any peptide sequence, the energy of each structure was determined using SWISS-PDB VIEWER, and the model with the least energy was selected for subsequent assessments [54]. The human alleles HLA-A*03:01 (PDB ID: 6O9B), HLA-B*18:01 (PDB ID: 6MT3), HLA-A*02:01 (PDB ID: 7RTD), HLA-A*24:02 (PDB ID: 7MJA), HLA-B*08:01 (PDB ID: 7NUI), HLA-B*07:02 (PDB ID: 7RZD), HLA-B*58:01

(PDB ID: 5VWJ), HLA-A*01:01 (PDB ID: 6AT9), HLA-B*44:03 (PDB ID: 4JQX), HLA-B*40:02 (PDB ID: 5IEK) were considered for MHC-I epitopes. Using the RCSB Protein Data Bank (<https://www.rcsb.org/>), the crystal structure of shortlisted HLA alleles was downloaded in pdb format [55]. Then, the protein preparation wizard of UCSF Chimera (version 1.11.2) was utilized to prepare proteins by removing structurally bound ligands [56]. The HADDOCK tool (<https://wenmr.science.uu.nl/haddock2.4/>) was used to estimate the interaction among the Alleles and CTL Epitopes [57, 58]. Molecular visualization of docking analysis was performed through Ligplot software, and images were obtained by UCSF Chimera and Microsoft PowerPoint 2019 [56, 59].

Population coverage analysis

Variations in the HLA allele distribution and expression in regions and races around the world may affect the response to vaccines based on epitope [60]. The population coverage of the candidate vaccine was estimated by implementing the IEDB population coverage server (<http://tools.iedb.org/population/>). To do this, the investigation of selected HTL and CTL epitopes coupled with their relevant HLA binding alleles in both MHC (I and II) classes were performed individually and in combination [61]. In this study, our emphasis was on the global coverage of alleles and parts of different continents.

MHC cluster analysis

The MHC gene family, as one of the most polymorphic genes in the various species' genomes, contains several thousand alleles in humans [62]. Cluster analysis of MHC alleles is used to identify two classes of MHC molecules with similar binding specificities. The MHCcluster 2.0 online tool (<https://services.healthtech.dtu.dk/services/MHCcluster-2.0/>) was utilized to provide highly instinctive heat maps and phylogenetic tree-based visualizations of the functional cluster between MHC variants based on the default parameters. During the MHC class I cluster analysis, the NetMHCpan-2.8 approach was utilized with an HLA-prevalent and -characterized module, while for the MHC class II cluster analysis, the relevant DRB allele modules were chosen [62, 63].

Designing and formulation of the multi-epitope vaccine

The vaccine was constructed from the selected HTL, CTL, and LBL epitopes of HPyV6 and HPyV7 proteins. Also, an adjuvant was attached to the vaccine structure using a suitable linker [64, 65]. We used TLR4 agonist as an adjuvant because viral glycoproteins were found to recognize TLR4 agonist [66, 67]. Therefore, 50S ribosomal protein L7/L12 was included as adjuvant (NCBI ID: P9WHE3) and attached to the N-terminal of the

vaccine peptides through a bifunctional linker (EAAAK). Contrastingly, the HTL, CTL, and LBL epitopes were connected using Gly-Pro-Gly-Pro-Gly (GPGPG), Ala-Ala-Tyr (AAY), and Lys-Lys (KK) linkers, respectively [64, 65]. The GPGPG linker inhibits the formation of the "junctional epitope" and aids in immune processing. The AAY linker improves epitope immunogenicity by affecting peptide stability. The KK linker improves the maintenance of independent immunogenic functions of the constructed vaccine [68, 69].

Antigenicity, allergenicity, solubility, and physicochemical property assessment

The ProtParam tool (<https://web.expasy.org/protparam/>) was applied to analyze the physicochemical profiles of the constructed vaccine [70]. We also used the Vaxijen v2.0 tool (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [31] and ANTIGENPro server (<https://scratch.proteomics.ics.uci.edu/>) of the Scratch protein forecast tool to predict antigenicity. ANTIGENPro demonstrated an accuracy of 76% with cross-validation tests on the combined dataset [32]. Three servers including AllergenFP v1.0 (<https://www.ddg-pharmfac.net/AllergenFP/>) [33], AllerTop v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) [39], and AlgPred (<http://crdd.osdd.net/raghava/algpred/>) [71] were used to predict the allergenic profile of the vaccine construct. The SOLpro online tool (<https://scratch.proteomics.ics.uci.edu/>) was applied to estimate the solubility of the proposed vaccine and a prediction score ≥ 0.5 , indicates the vaccine will be soluble [72]. In addition, we applied the Protein-Sol online tool (<https://protein-sol.manchester.ac.uk/>) to better understand the solubility [73]. By comparing the scaled solubility value (QuerySol) with the E. coli proteins' mean solubility from the experimental dataset (PopAvrSol), solubility is predicted, and a predicted score greater than 0.45 is considered to be soluble [74]. In order to predict the number of transmembrane helices, we used the TMHMM v2.0 tool (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>) [34]. Also, possible signal peptides were investigated using the application of the SignalP 4.1 tool (<https://services.healthtech.dtu.dk/services/SignalP-4.1/>) in the ultimate designed vaccine [75].

BLAST homology assessment

The PSI-BLAST algorithm of the NCBI Protein BLAST (BLASTp) module was used to determine the homology between the vaccine construct and the human proteome [76, 77]. Cross-checking study was performed to avoid autoimmune reactions through molecular imitation. The BLASTp search limited the results to records from *H. sapiens* (taxid: 9606) only. In order to be valid, the query

coverage must not show more than 40% homology to the human proteome [78].

Prediction of the secondary structure

The secondary structural configurations were identified using two servers PSIPRED v4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>) and Prabi (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) with default parameters [79, 80]. The percentage of 2D features, i.e. α -helix, β -turns, and random coils was calculated on the vaccine construct by both servers. The precision of prediction results with the Prabi tool was reported to be a mean of 64.4% [80]. The PSIPRED server, as the most accurate predictive generator of secondary structure, displays an accuracy of 78.1% [81].

Homology modeling and 3D structure refinement and validation

I-TASSER tool (<https://seq2fun.dcmf.med.umich.edu/I-TASSER/>) was used to generate the three-dimensional model of the multi-epitope vaccine. Using threading templates as templates, this server produces a 3D structure based on the amino acids sequence, and it estimates the C-score to assess the validity of the predicted models. The C-score of a model typically falls within the range of -5 to 2, with a higher C-score indicating greater confidence [82–84]. Tertiary structure vaccine model refinement was performed via the GalaxyRefine tool (<https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). Various parameters including GDT-HA, rmsd, poor rotations, Molprobit, clash score, and Rama-favored are produced in the output result of five refined models [85]. After validation of the models through the ProSA-web online tool (<https://prosa.services.came.sbg.ac.at/prosa.php>), the estimation of the Z-score and analysis of the stereochemical quality of each protein structure were performed by assessing the residue by residue geometry and overall structure geometry [86]. In order to further analyze the Ramachandran plot, the Procheck web server (<https://saves.mbi.ucla.edu/>) was used to determine the overall quality of the refined 3D structure of the vaccine. The Ramachandran plot is a plot of the dihedral angles phi (ϕ) and psi (ψ) of amino acids to visualize the percentage of amino acids in the generously allowed, disallowed, most favorite, and additional allowed areas [87].

Identification of discontinuous B-Cell epitopes

The ElliPro server (<http://tools.iedb.org/elliopro/>) was used to estimate conformational B-cell epitopes in the designed vaccine utilizing default parameters (minimum score: 0.5; maximum distance: 6 Å). The improved Thornton's technique using residue clustering algorithms is the basis of the results. Prediction is done based on the

neighbor residue clustering, protein form, and residual protein index (PI) [88].

Molecular docking of the immune receptor (TLR4) and designed vaccine

The protein data bank (RCSB) at 2.4 Å resolution was used to retrieve TLR4 complexes (PDB ID: 4G8A) [55]. Heteroatoms and three chains B, C, and D were deleted in the UCSF Chimera software (version 1.11.2) [56]. Energy minimization of protein was carried out using the Swiss-PDB Viewer with the GROMOS 43B1 force field [54]. Molecular docking of vaccine-TLR4 complexes was performed in Cluspro (<https://cluspro.bu.edu/login.php>) online tools [89]. It's an automated, web-based program for the docking of peptide–protein or protein–protein. The server executes a series of three computational procedures in the following manner: firstly, the process of rigid body docking is carried out employing PIPER; secondly, the 1000 docked structures with the lowest energy are subjected to clustering using pairwise IRMSD as the distance metric; and finally, the forecasted complex structures positioned at the cluster centers are refined by minimizing their energy. Also, We used the balanced coefficient to obtain the best protein–protein binding results [90]. The output of this server is a short list of putative complexes ranked according to their clustering properties.

In silico immune simulation

The immune simulation study was conducted using the C-ImmSim server (<https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php?page=1>) to understand and investigate the immunogenicity and immune response profile. Using the position-specific scoring matrix (PSSM), this server employs real lifelike immune responses and interactions, and machine learning [91]. The time steps in the CImmSim web tool (with default parameters) were set to 1, 42, and 84, each time step is equal to 8 h and time step 1 is injection at time=0. The time interval between two injections (a total of 3 injections) was considered 4 weeks [92].

Molecular dynamics simulation

Molecular dynamics (MD) simulation was applied to refine the TLR-vaccine complex structures using GROMACS 2018 [93, 94]. The structures were centered in a dodecahedron box and filled with water using tip3 water model. To neutralize systems some molecules of water were randomly replaced by Cl⁻ or Na⁺. After neutralization, the energy minimization was done using steepest descent algorithm. Equilibrating the systems was performed under 100 ps NVT at temperature of 298 K followed by 100 ps NPT ensembles at pressure of

1 bar. Electrostatic interactions were calculated by PME (Flores-Canales and Kurnikova, 2015) and the LINCS procedure was applied to constrain all bonds connecting hydrogen atoms. The Final MD simulation was run for 100 ns with no restraint.

In silico cloning and codon optimization of the final vaccine protein.

The online Java Codon Compatibility Tool (JCAT) web server (<http://www.jcat.de/>) was utilized for codon optimization and reverse translation of the ultimate vaccine protein [95]. To express the final construct in *E. coli*, the K12 strain was used as the host. Using this server, important parameters such as GC content and codon adaptive index (CAI) were calculated for assessment of protein expression levels. After introducing sites for BamHI and XhoI restriction enzymes within 3' and 5' ends of the designed vaccine sequence, respectively, this sequence was transformed into the pET30 (+) vector through the SnapGene software.

Results

Vaccine design

Retrieval and evaluation of protein sequences

The reference sequence for HPyV6 and HPyV7 large T antigen (LTag), and viral proteins 1/2 (VP1, VP2) were obtained from UniProt Proteome database. VaxiJen v2.0 tool was utilized to determine the subjected protein sequences. TMHMM v2.0 was employed to anticipate the number of TM helices. The antigenicity of candidate proteins varies from 0.4281 to 0.4996, hence the proteins of interest have sufficient predicted antigenic properties. In addition, the AllergenFP server suggested the proteins are non-allergenic. To construct a multi-epitope-based vaccine, the large T, VP1 and VP2 were included. The quantity of transmembrane helices is zero. Table 3 displays the sequences of these proteins, their UniProt entries, and allergenicity, antigenicity, and TM helices.

Identification and validation of CTL epitopes

Based on the specified selection range, 52 potential CTL epitopes of the candidate proteins were identified as non-allergenic, non-toxic, antigenic, as well as immunogenic (Table S1). The epitopes were calculated with NetCTL 1.2 using a combinatorial approach. Just ten predicted CTL epitopes were selected for peptide-based vaccine design. The list of the chosen CTL epitopes with their characteristics in the final vaccine is shown in Table 4.

Identification and validation of HTL epitopes

Overall, 55 potential HTL epitopes under specified selection range as non-allergenic, non-toxic, and antigenic were identified (Table S2). Evaluation of expected cytokine induction capability was done on selected HTL epitopes and based on those results 7 peptides were selected to include in the final vaccine. The list of the chosen HTL peptides with their characteristics in the final vaccine is shown in Table 5.

Identification and validation of linear B-cell epitope (LBL)

Overall, 46 LBL epitopes from the target proteins were identified by evaluating potential toxicity, immunogenicity and antigenic characteristics (Table S3). One LBL epitope was chosen from each of the 6 protein components for usage in the ultimate vaccine. The list of the six LBL epitopes with their characteristics in the final vaccine is shown in the Table 6.

Vaccine evaluation

Evaluation of human homology and epitope conservancy

A lack of homology to normal human proteins was assessed for each of the shortlisted epitopes in both MHC classes and no homologies were identified in the human proteome, suggesting that responses against these peptides are not likely to cause response against a normal protein. Tables 4 and 5 have incorporated conservancy and human homology analyses of selected epitopes.

Table 3 Details of the selected proteins and Their Selection Criteria

Species	UniProt entry	protein name	abbreviation	Antigenicity (Probable antigen)	allergenicity	no. of TM helices
HPyV6	D6QWG6	Large T antigen	LTag	0.4928	Probable non-allergen	0
	D6QWG0	Viral Protein 1	VP1	0.4555	Probable non-allergen	0
	D6QWF3	Viral Protein 2	VP2	0.4288	Probable non-allergen	0
HPyV7	D6QWJ6	Large T antigen	LTag	0.4996	Probable non-allergen	0
	D6QWJ5	Viral Protein 1	VP1	0.4281	Probable non-allergen	0
	D6QWJ3	Viral Protein 2	VP2	0.4323	Probable non-allergen	0

Table 4 A brief list of CTL Epitopes to form the final structure of the vaccine

specie	epitopes	protein	C-score	immunogenicity score	antigenicity	allergenicity	toxicity	epitope conservancy hit (maximum identity %)	human homology	interacting MHC-I alleles
HPV6	LSHATLGNK	LTag	1.2320	0.10205	1.5211	probable non-allergen	non-toxic	100	non-homologue	HLA-A*11:01, HLA-A*03:01, HLA-A*30:01
	FERWVSFGM	LTag	1.3474	0.24276	1.2787	probable non-allergen	non-toxic	100	non-homologue	HLA-B*18:01, HLA-B*40:01, HLA-B*40:02
	WLLFVLEEL	VP2	1.1038	0.28696	0.7654	probable non-allergen	non-toxic	100	non-homologue	HLA-A*02:01, HLA-A*02:06
HPV7	IKYVEAILL	VP1	1.5128	0.21171	0.4174	probable non-allergen	non-toxic	100	non-homologue	HLA-A*24:02, HLA-B*35:03
	MELTDVLLI	LTag	1.2565	0.08008	1.3253	probable non-allergen	non-toxic	100	non-homologue	HLA-B*44:02, HLA-B*44:03, HLA-B*40:01, HLA-B*18:01, HLA-B*53:01, HLA-B*40:02, HLA-B*51:01
	TPKRRNLLF	LTag	1.3998	0.01055	1.7138	probable non-allergen	non-toxic	100	non-homologue	HLA-B*08:01, HLA-B*07:02
	GPRIGSTTM	VP2		0.08343	0.7918	probable non-allergen	non-toxic	100	non-homologue	HLA-B*07:02
	LWLPQAWPW	VP2	1.0974	0.08973	0.5232	probable non-allergen	non-toxic	100	non-homologue	HLA-A*24:02, HLA-B*58:01, HLA-B*57:01, HLA-A*23:01, HLA-B*53:01
	DTMWWWEAY	VP1	2.0643	0.43299	0.5828	probable non-allergen	non-toxic	100	non-homologue	HLA-A*26:01, HLA-A*25:01, HLA-A*01:01, HLA-B*35:01, HLA-B*18:01, HLA-A*29:02, HLA-A*30:02, HLA-B*46:01
	TELLFAPQM	VP1	1.0319	0.05903	0.8241	probable non-allergen	non-toxic	100	non-homologue	HLA-B*18:01, HLA-B*44:02, HLA-B*40:02, HLA-B*44:03, HLA-B*40:01

Table 5 A brief list of HTL Epitopes to form the final structure of the vaccine

specie	epitopes	protein	antigenicity	allergenicity	IFN-γ	IL4	toxicity	epitope conservancy hit (maximum identity %)	human homology	interacting MHC-II alleles
HPV6	VAAQRRVLMLESTRQ	LTag	0.8394	probable non-allergen	positive	inducer	non-toxic	100	non-homologue	HLA-DRB1*04:10, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*04:08
	GLVNLVYAVVYNRQ	VP2	0.7212	probable non-allergen	positive	inducer	non-toxic	100	non-homologue	HLA-DRB1*08:06, HLA-DRB1*01:02, HLA-DRB1*08:13, HLA-DRB1*13:04, HLA-DRB1*08:01, HLA-DRB1*04:23
	GRFFRVHFRQRVKH	VP1	0.4931	probable non-allergen	positive	inducer	non-toxic	100	non-homologue	HLA-DRB1*11:20, HLA-DRB1*11:14, HLA-DRB1*13:23, HLA-DRB1*08:04, HLA-DRB5*01:01, HLA-DRB5*01:05
HPV7	SAGYIRAOGTPAGVE	VP1	0.8598	probable non-allergen	positive	inducer	non-toxic	100	non-homologue	HLA-DRB1*09:01
	DWNFVADFAADMELT	LTag	1.2109	probable non-allergen	positive	inducer	non-toxic	100	non-homologue	HLA-DRB3*01:01, HLA-DRB1*03:09, HLA-DRB1*03:05, HLA-DRB1*04:21
	FEPGGWMYDTQNLP	VP2	0.9125	probable non-allergen	negative	inducer	non-toxic	100	non-homologue	HLA-DRB1*15:06, HLA-DRB1*04:02, HLA-DRB1*08:06, HLA-DRB1*15:02, HLA-DRB1*08:04, HLA-DRB1*08:13, HLA-DRB1*04:10, HLA-DRB1*08:01
	FFRVHCRQRRIKHPY	VP1	0.4266	probable non-allergen	negative	inducer	non-toxic	100	non-homologue	HLA-DRB5*01:05, HLA-DRB1*11:20, HLA-DRB1*11:14, HLA-DRB1*13:23

Table 6 A brief list of LBL Epitopes to form the final structure of the vaccine

specie	epitopes	protein	protein probability score	antigenicity	allergenicity	toxicity	human homology
HPyV6	SSEVRPPPQYGSPGWE	LTA _g	0.84	1.0311	probable non-allergen	non-toxic	non-homologue
	PSKENQPSVAGIKATR	VP1	0.88	1.3222	probable non-allergen	non-toxic	non-homologue
	SNKKRRSSGGYGN SATF	VP2	0.80	0.7625	probable non-allergen	non-toxic	non-homologue
HPyV7	DSKYSATPPKQKPNP	LTA _g	0.90	1.1870	probable non-allergen	non-toxic	non-homologue
	PATIPPTVEGGLGFAP	VP1	0.88	1.1122	probable non-allergen	non-toxic	non-homologue
	DQRGGFHDEGTWWSFQ	VP2	0.81	1.1021	probable non-allergen	non-toxic	non-homologue

Molecular docking analyses of the CTL epitopes and HLA alleles

Utilizing molecular docking simulations, we assessed CTL epitope binding to alleles of HLA. We have elected the HLA-A*03:01 allele for LSHATLGNK epitope, HLA-B*18:01 allele for FERWVSFGM epitope, HLA-A*02:01 allele for WLLFVLEEL epitope, HLA-A*24:02 allele for IYKVEAILL epitope, HLA-B*08:01 allele for TPKRRNLLF epitope, HLA-B*07:02 allele for GPRIG-STTM epitope, HLA-B*58:01 allele for LWLPQAWPW epitope, HLA-A*01:01 allele for DTMIVWEAY epitope, HLA-B*44:03 allele for MELTDVLLI epitope, and HLA-B*40:02 allele for TELLFAPQM epitope. The more negative the z-score indicates that the cluster has a high level of reliability. Based on the docking parameters, the selected CTL epitopes exhibited excellent binding interactions with the active site of HLA alleles (Fig. 1). The docking statistics are demonstrated in Table 7.

Population coverage analysis

The analysis of MHC I and MHC II epitopes demonstrated that 97.74% of the global population is covered by MHC I epitopes, while MHC II epitopes cover 26.99% of the global population. Since a multi-epitope vaccine protein includes both MHC epitope classes, a combined estimate of their population coverage was used. Overall, 98.35% of the world's population was covered. Combining MHC class-I and class-II epitope coverage in Europe was 99.55%. followed by West Indies (98.25%), North America (98.19%), East Asia (96.82%), Oceania (95.27%), Southeast Asia (94.27%), North Africa (92.84%), West Africa (92.69%), Northeast Asia (91.63%), South Africa (91.08%), Southwest Asia (89.37%), East Africa (88.85%), South Asia (88.69%), Central Africa (84.65%), South America (80.56%) and Central America (9.07%). Comparison of population coverage between the epitopes of MHC class I/II and mixed MHC epitopes are shown in Fig. 2, Table 8, and Figures S1-S3.

MHC cluster analysis

MHC cluster v2.0 server was exploited in order to cluster MHC classes I and II alleles that interact by selected structural protein epitopes. In this study, 25 alleles from the MHC class I, and 22 alleles from the MHC class II were analyzed. MHC I and MHC II molecules Cluster analysis is shown in Fig. 3A, C, respectively. A tree map showing the cluster analysis of MHC I and MHC II is also shown in Fig. 3B, D. The red zones on the heat map were associated with stronger interactions, whereas the yellow zones were associated with feeble interactions among clusters of both MHC molecules.

Formulation of the vaccine construct

In order to formulate the vaccine construct, we assembled the most favorable CTL, HTL, and LBL epitopes using AAY, GPGPG, and KK linkers, respectively. Furthermore, 50S ribosomal protein L7/L12 adjuvant (NCBI ID: P9WHE3) was attached using EAAAK linker to the N-terminal region of the vaccine. The structure of the vaccine consists of seven epitopes of HTL, ten epitopes of CTL, and six epitopes of LBL from the target protein sequences of polyomaviruses 6 and 7. After suitable evaluation and comparison of different structures, we determined the final structure of the vaccine with 501 amino acids. The final recombinant vaccine was analyzed for subsequent evaluations (Fig. 4).

Vaccine assessment

Assessment of the antigenicity, allergenicity, and physicochemical properties of the final vaccine protein

The physicochemical characteristics of the formulated construct were evaluated. The vaccine's chemical formula is $C_{2455}H_{3814}N_{662}O_{694}S_{11}$. The molecular weight of the vaccine construct was estimated to be 54059.97 Da. It was calculated that the protein has a theoretical pI value of 9.41. This value represents that the protein is highly basic. The grand average of hydropathicity (GRAVY) property demonstrated the hydrophilic nature of protein, as it was -0.317. Furthermore, the instability index

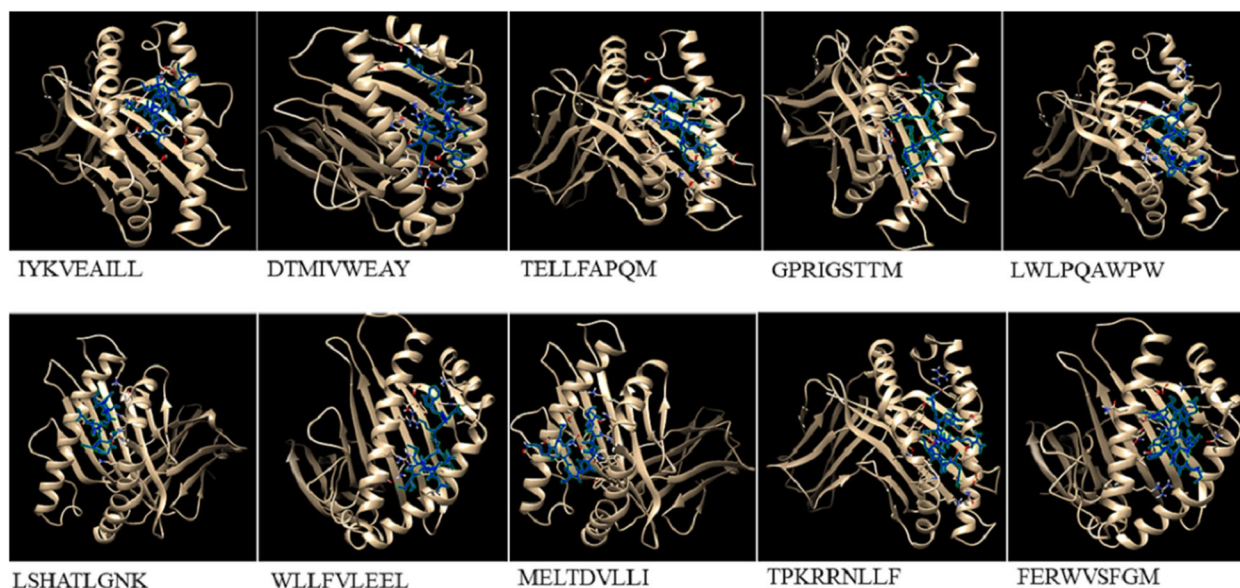


Fig. 1 Molecular docking of the selected CTL epitopes with their respective HLA alleles as indicated in Table 5

Table 7 Data of the molecular docking between CTL Epitopes and HLA Alleles

alleles	epitopes	HADDOCK score	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	RMSD from the overall lowest-energy structure	Z-Score
HLA-A*03:01	LSHATLGNK	-66.5 ± 3.3	-28.3 ± 3.5	-206.2 ± 23.6	0.2 ± 1.9	28.5 ± 24.3	1006.2 ± 77.1	1.2 ± 0.1	-1.7
HLA-B*18:01	FERWVSFGM	-101.3 ± 6.3	-45.5 ± 6.9	-112.0 ± 25.7	-39.3 ± 3.5	58.6 ± 38.3	1442.7 ± 27.8	0.5 ± 0.3	-1.6
HLA-A*02:01	WLLFVLEEL	-85.0 ± 0.4	-39.3 ± 3.4	-97.8 ± 4.0	-29.8 ± 1.6	36.2 ± 21.4	1408.9 ± 38.7	2.2 ± 0.1	-1.4
HLA-A*24:02	IYKVEAILL	-72.3 ± 2.8	-39.3 ± 5.0	-69.2 ± 14.9	-22.7 ± 1.9	35.1 ± 19.9	1143.8 ± 17.1	1.5 ± 0.1	-1.1
HLA-B*08:01	TPKRRNLLF	-99.3 ± 2.1	-43.4 ± 2.8	-220.7 ± 11.5	-12.0 ± 4.6	2.8 ± 1.2	1447.9 ± 117.0	0.6 ± 0.4	-1.6
HLA-B*07:02	GPRIGSTTM	-62.2 ± 0.8	-35.0 ± 1.6	-169.5 ± 9.4	5.9 ± 1.6	8.0 ± 7.7	1242.8 ± 24.3	0.2 ± 0.1	-1.7
HLA-B*58:01	LWLPQAWPW	-93.9 ± 1.8	-48.4 ± 2.6	-29.3 ± 8.5	-43.7 ± 2.3	40.4 ± 34.9	1296.0 ± 22.3	0.6 ± 0.1	-1.5
HLA-A*01:01	DTMIVWEAY	-108.1 ± 3.0	-52.8 ± 5.8	-181.1 ± 25.2	-19.3 ± 2.8	3.4 ± 3.5	1485.3 ± 18.1	0.4 ± 0.3	-2.3
HLA-B*44:03	MELTDVLLI	-70.2 ± 3.1	-47.4 ± 3.9	-63.0 ± 25.0	-10.5 ± 1.0	2.9 ± 1.7	1331.2 ± 24.4	0.3 ± 0.2	-1.2
HLA-B*40:02	TELLFAPQM	-82.9 ± 2.5	-46.1 ± 3.2	-87.9 ± 12.5	-20.9 ± 2.4	17.2 ± 13.6	1362.8 ± 30.6	0.4 ± 0.3	-1.5

score was 34.33, and the score of the aliphatic index was calculated as 72.95. The subjected protein’s half-life was evaluated 30 h in mammalian reticulocytes in vitro and exceeds 20 h in yeast also approximated 10 h in *E. coli* in vivo. Vaccine allergenicity and antigenicity were also assessed using multiple servers. According to ANTI-GENPro and Vaxijen 2.0 tools, the score of antigenicity was determined to be 0.897722 and 0.6224, respectively. We also employed several tools to evaluate the solubility of the vaccine sequence. In Solpro and Protein-Sol tools, the score of solubility was estimated to be 0.777639 and 0.504, respectively. Also, the final proposed vaccine did

not indicate any signal peptides and Transmembrane helices based on the prediction data (Table 9 and Fig. 5A).

BLAST homology assessment

Homo sapiens proteins were found to be 24% homologous with the protein vaccine sequence based on sequence homology between the designed vaccine and the proteome sequence. These results confirmed that chimeric vaccine construct cannot prompt autoimmune responses in the host based on BLAST homology assessment. This study focused on the Homo sapiens species (taxid:9606).

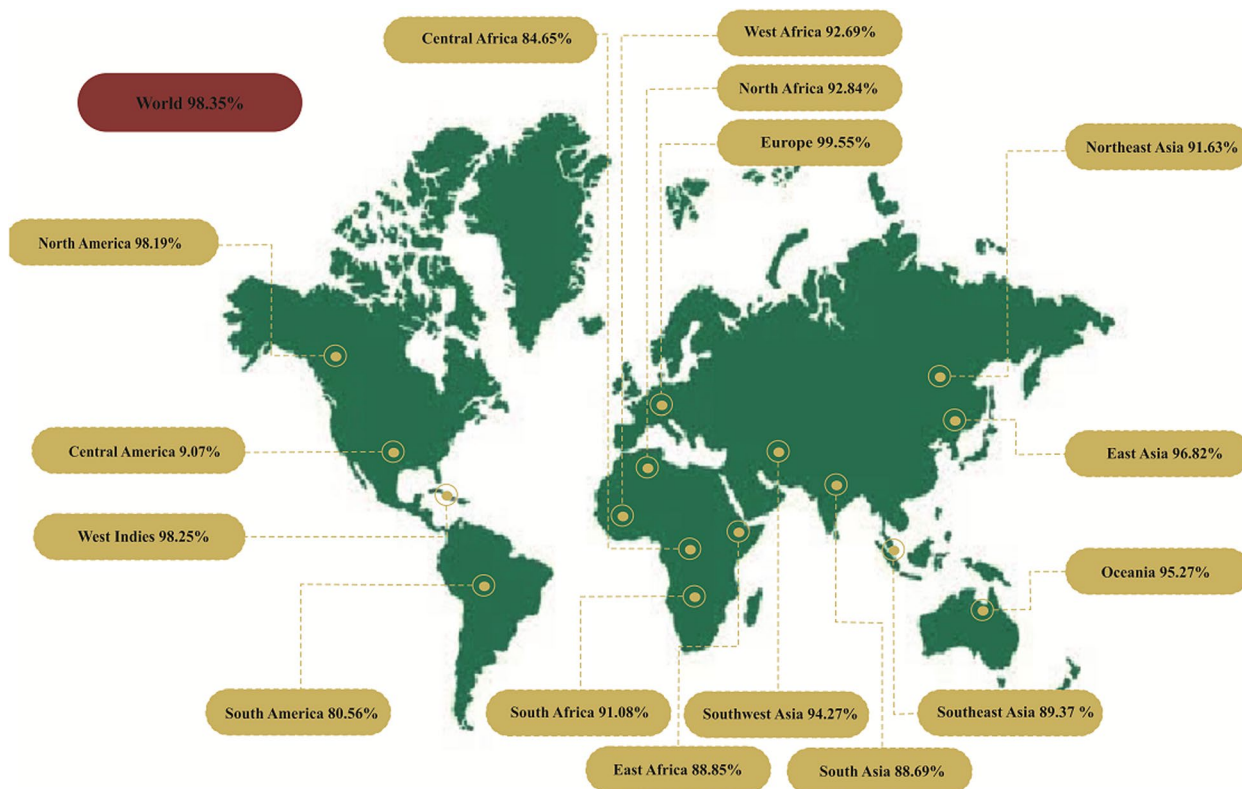


Fig. 2 Analysis of population coverage of alleles worldwide

Table 8 Analysis of MHC restriction data for worldwide population coverage

Population/Area	MHC I			MHC II			MHC I and MHC II Combined		
	Coverage(%) ^a	Average hit ^b	PC90 ^c	Coverage(%) ^a	Average hit ^b	PC90 ^c	Coverage(%) ^a	Average hit ^b	PC90 ^c
World	97.74%	3.2	1.55	26.99%	0.34	0.14	98.35%	3.54	1.77
East Asia	96.82%	3.29	1.62	53.39%	0.66	0.21	96.82%	3.29	1.62
Northeast Asia	91.63%	2.34	1.06	33.41%	0.39	0.15	91.63%	2.34	1.06
South Asia	88.69%	1.96	0.88	24.5%	0.29	0.13	88.69%	1.96	0.88
Southeast Asia	94.27%	3.05	1.26	41.54%	0.46	0.17	94.27%	3.05	1.26
Southwest Asia	89.37%	2.16	0.94	18.19%	0.21	0.12	89.37%	2.16	0.94
Europe	99.55%	3.86	2.18	20.76%	0.28	0.13	99.55%	3.86	2.18
East Africa	88.85%	2.18	0.9	28.6%	0.39	0.14	88.85%	2.18	0.9
West Africa	92.69%	2.38	1.11	52.54%	0.66	0.21	92.69%	2.38	1.11
Central Africa	84.65%	1.94	0.65	24.28%	0.35	0.13	84.65%	1.94	0.65
North Africa	92.84%	2.38	1.11	37.7%	0.51	0.16	92.84%	2.38	1.11
South Africa	91.08%	2.16	1.04	1.79%	0.02	0.1	91.08%	2.16	1.04
West Indies	98.25%	3.4	1.77	27.52%	0.37	0.14	98.25%	3.4	1.77
North America	98.19%	3.37	1.7	24.95%	0.32	0.13	98.19%	3.37	1.7
Central America	9.07%	0.14	0.11	16.53%	0.2	0.12	9.07%	0.14	0.11
South America	80.56%	2.05	0.51	27.31%	0.35	0.14	80.56%	2.05	0.51
Oceania	95.27%	3.08	1.35	35.77%	0.4	0.16	95.27%	3.08	1.35

^a Projected population coverage

^b Average number of epitope hits/HLA combinations recognized by the population

^c Minimum number of epitope hits/HLA combinations recognized by 90% of the population

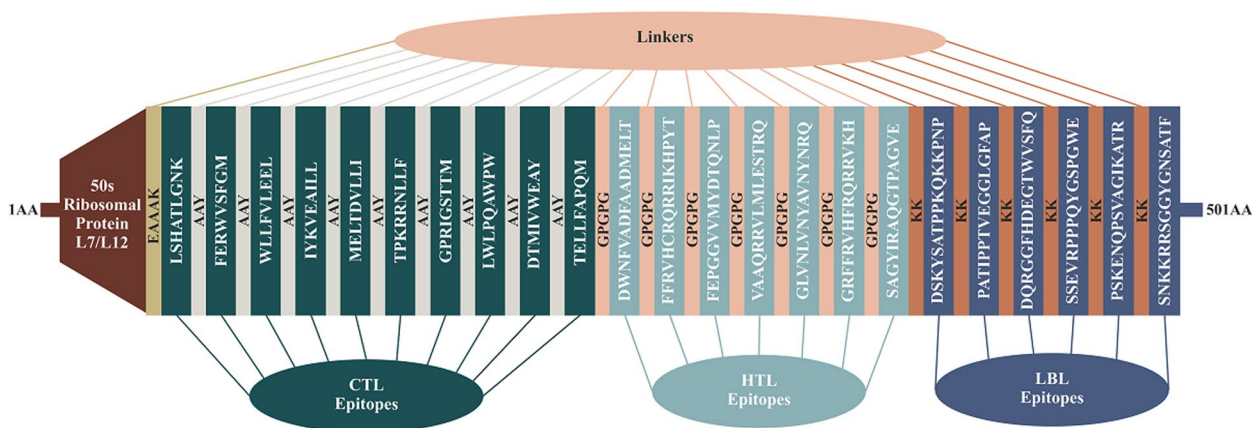


Fig. 4 Schematic presentation of the formulated multi-epitope vaccine construct. The multi-epitope vaccine constructs included (left to right) an adjuvant and CTL, HTL, and LBL epitopes, indicated in the Brown, navy, light blue, and violet rectangular boxes, respectively

refinement, Procheck and ProSA and tools were utilized to assessment this model. In the current survey, model 1 showed a z-score of -2.12 and 73.4% of the residues in the most favored regions. The GalaxyRefine tool was used to refine the 3D structure of the submitted model. This generated five refined structures for the raw model (Table 10). After refinement, all structures show the regions favored by Rama more than the submitted originally raw model. Model 3 was determined to be the best refined structure among the generated models. It displayed goodRama-favored (90.6), poor rotamers (0.8), MolProbity (2.195), clash score (14.4), rmsd (0.517), and GDT-HA (0.9037) scores. ProSA and

SAVES v6.0 online tools were employed to validate the refined structure. Based on Ramachandran plot of the selected model indicated that 87.7% of amino acids in favored regions, 9.6% additional allowed, 1.0% generously allowed, and 1.7% disallowed regions were found. The Z-score value for the refined model was estimated -2.61 (Fig. 6). For further analysis, we have selected model 3 in this study.

Screening for conformational B-Cell epitopes

The ElliPro server identified thirteen Conformational B-cell epitopes in the vaccine construct sequence (Fig. 7). A total of 243 residues were found in these epitopes

Table 9 Allergenicity, antigenicity, and physicochemical properties of the final structure of the vaccine

characteristics	findings	remarks
Number of amino acids	501	suitable
molecular weight	54,059.97 Da	average
Theoretical pI	9.41	basic
chemical formula	C ₂₄₅₅ H ₃₈₁₄ N ₆₆₂ O ₆₉₄ S ₁₁	--
extinction coefficient (at 280 nm in H2O)	77,810 M ⁻¹ cm ⁻¹	
estimated half-life (mammalian reticulocytes, in vitro)	30 h	satisfactory
estimated half-life (yeast cells, in vivo)	> 20 h	satisfactory
estimated half-life (<i>E. coli</i> , in vivo)	> 10 h	satisfactory
instability index of vaccine	34.63	Stable
aliphatic index of vaccine	72.95	thermostable
grand average of hydropathicity (GRAVY)	-0.317	hydrophilic
antigenicity	0.6455 (Vaxijen v2.0) 0.897722 (ANTIGENPro)	antigenic
allergenicity	No (AllergenFP v1.0) No (AlgPred) No (AllerTop v2.0)	non-allergen
solubility	0.766196 (SolPro) 0.504 (Protein-Sol)	soluble
TM helices	No	suitable

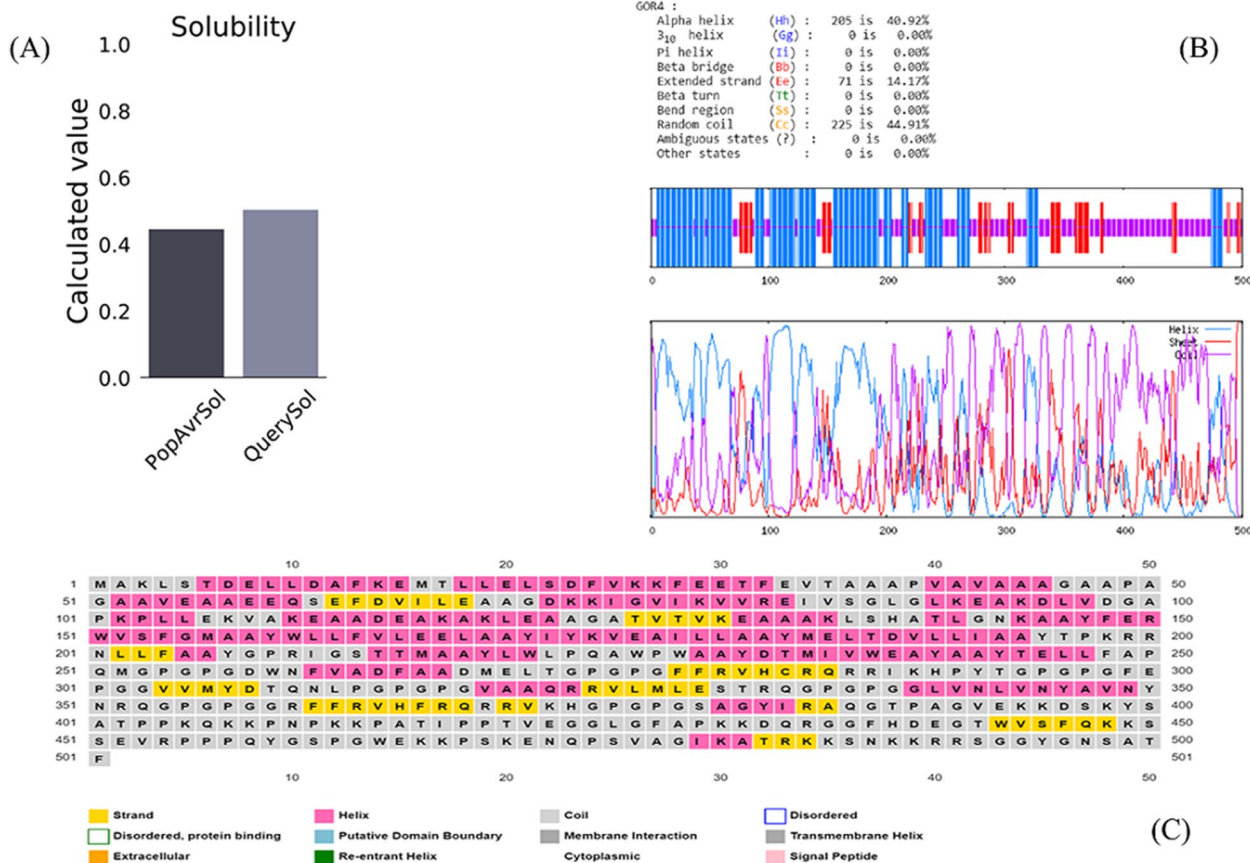


Fig. 5 Protein-Sol server prediction of vaccine protein Solubility (A), Prabi server Prediction of vaccine secondary structure (B), PSIPRED server Prediction of vaccine secondary structure (C)

ranging in size from 3 to 98, shown in Table 11. Moreover, B—cell epitope scores ranged from 0.975 to 0.53 for the prediction of conformational B-cell epitopes.

Molecular docking of the vaccine protein and TLR complex

Immune cells and vaccine constructs must interact in order to produce a stable and efficient immune response. Molecular Docking of the designed vaccine with TLR4 was carried out by the ClusPro 2.0 server. In the current study, the program produced 30various

clusters and ranked them by energy level. There were -1414.0, -1406.2, -1372.0, -1350.9, -1341.0, -1339.5, -1327.2, and -1321.0 kcal/mol of energy in the eight top clusters. The best group with the minimum energy of -1414.0 kcal/mol was selected. The Chimera 1.15rc program was applied to visualize the docked complex (Fig. 8). Using the LigPlot v1.4.5 software, we have generated a map with the hydrophobic interactions and hydrogen bonds between the protein vaccine and TLR4 (Fig. 9). The vaccine and chain A of TLR4 formed 20

Table 10 Models of vaccines refined by Galaxy Refine

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.074	5.4	18.2	71.5
MODEL 1	0.9097	0.506	2.216	14.1	0.8	89.6
MODEL 2	0.9087	0.504	2.240	15.4	0.3	90.0
MODEL 3	0.9037	0.517	2.195	14.4	0.8	90.6
MODEL 4	0.9142	0.498	2.338	15.3	1.3	89.6
MODEL 5	0.9077	0.515	2.230	14.2	0.5	89.2

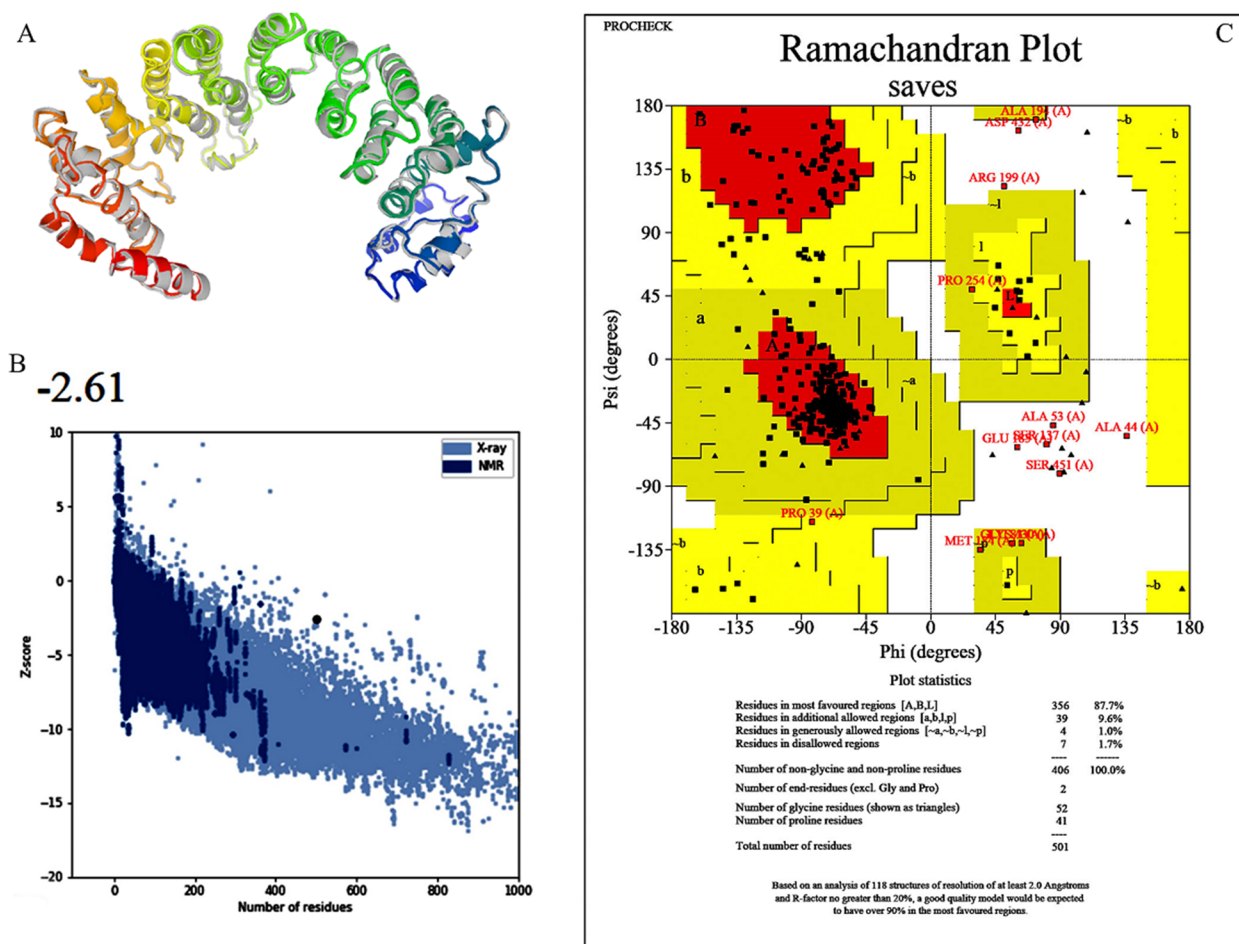


Fig. 6 The evaluation of the ultimate 3D structure. 3D structure refined by the GalaxyRefine server (A), a Z-score calculated using the PROSA server for the vaccine construct (B), and an analysis of the vaccine construct using Ramachandran plot (C)

hydrogen bonds. These hydrogen bonds are formed by amino acids along with their lengths, as shown in Table 12.

Vaccine immune simulation

In silico immune simulation

Immune simulator C-ImmSim was employed to provide simulations of the immune responses associated with the final chimeric vaccine construct. The secondary and tertiary responses were clearly indicated by the anticipated elevated levels of IgM + IgG, IgM, IgG1 + IgG2, and IgG1 antibodies, subsequently followed by a reduction in antigen concentration (Fig. 10A). Results indicated a variety of long-lasting B-cell isotopes. B-cell isotype switching and memory formation may be involved in this process (Fig. 10B). In addition, T helper (helper) and TC (cytotoxic) cells are showing a clear increase with memory growth (Fig. 10C and D). There is also a clear increase

in IFN-γ production and the growth of dendritic cells after immunization (Fig. 10E and F). These data represent that After successive exposures to the target antigen, robust and significant secondary immune response, antigen clearance enhancement, and production of vigorous immune memory occur.

Evaluation of MD simulations

The global structural stability of proteins was evaluated using Root Mean Square deviation (RMSD) of the backbone atoms. This plot shows how much the protein conformation has changed during MD simulation from initial structure. The TLR showed RMSD value in the range of 0.25 to 0.35 nm. The RMSD after 20 ns reached stability. Furthermore, the RMSD of vaccine was plotted and was in the range of 0.45 to 1. The root-mean-square fluctuation (RMSF) indicates the fluctuation of protein residues over time from a reference position during simulation. In current simulations, no unusual fluctuation

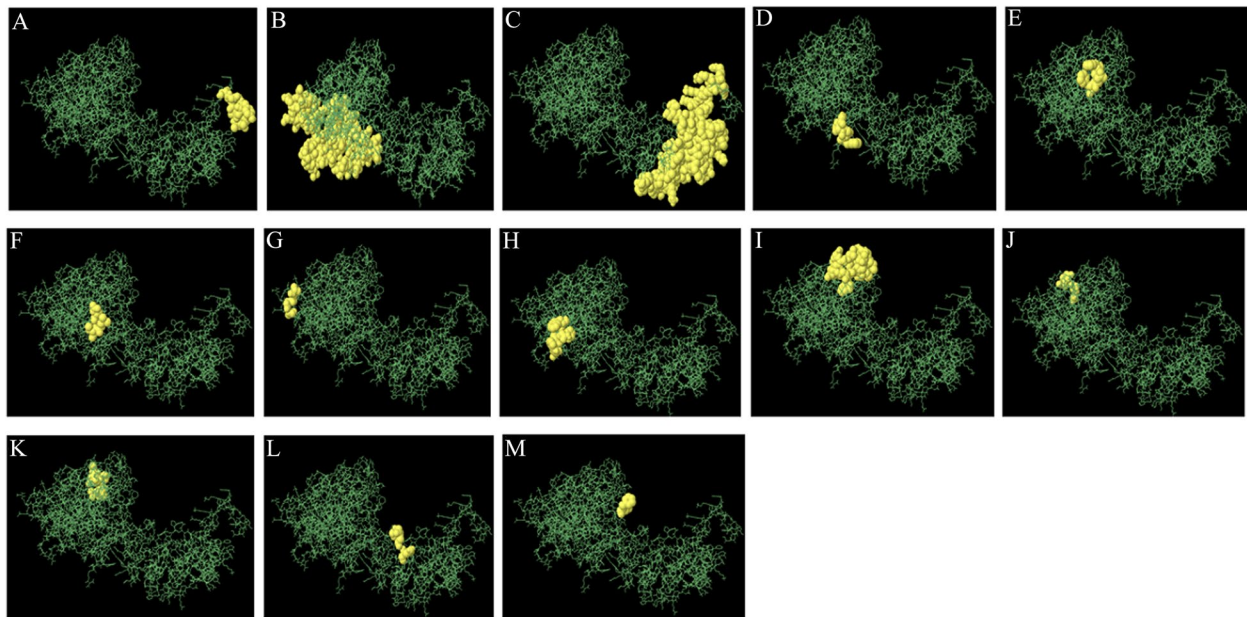


Fig. 7 3D structure Conformational B-cell epitopes existing in the protein vaccine (A-N). Green rods and yellow domain show the protein construct and Conformational B- cell epitopes, in order

Table 11 Shortlist of Conformational Epitopes of the ultimate designed vaccine

no	residues	no. of residues	score
1	B:S492, B:G493, B:G494, B:Y495, B:G496, B:N497, B:S498, B:A499, B:T500, B:F501	10	0.975
2	B:M1, B:A2, B:K3, B:L4, B:S5, B:T6, B:D7, B:E8, B:L9, B:L10, B:D11, B:A12, B:F13, B:K14, B:E15, B:M16, B:T17, B:L18, B:L19, B:E20, B:L21, B:S22, B:D23, B:F24, B:V25, B:K26, B:K27, B:F28, B:E29, B:E30, B:T31, B:F32, B:E33, B:V34, B:T35, B:A36, B:A37, B:A38, B:P39, B:V40, B:A41, B:V42, B:A43, B:A44, B:A45, B:G46, B:A47, B:P49, B:A50, B:G51, B:A52, B:A53, B:V54, B:E58, B:S61, B:E62, B:F63, B:D64, B:V65, B:I66, B:L67, B:E68, B:A69, B:A70, B:G71, B:D72, B:K73, B:K74, B:I75, B:G76, B:V77, B:I78, B:K79, B:V80, B:V81, B:R82, B:E83, B:I84, B:V85, B:S86, B:G87, B:L88	82	0.773
3	B:Q353, B:G354, B:P355, B:P357, B:G378, B:S379, B:A380, B:G381, B:R384, B:A385, B:T388, B:P389, B:G391, B:V392, B:K394, B:K395, B:D396, B:K398, B:Y399, B:S400, B:A401, B:T402, B:P403, B:K407, B:I417, B:P419, B:T420, B:V421, B:E422, B:G423, B:G424, B:L425, B:G426, B:F427, B:A428, B:P429, B:K430, B:K431, B:D432, B:Q433, B:R434, B:G435, B:G436, B:F437, B:H438, B:D439, B:E440, B:G441, B:T442, B:W443, B:V444, B:S445, B:F446, B:Q447, B:K448, B:K449, B:S450, B:S451, B:E452, B:V453, B:R454, B:P455, B:P456, B:P457, B:Q458, B:Y459, B:G460, B:S461, B:P462, B:G463, B:W464, B:E465, B:K466, B:K467, B:P468, B:S469, B:K470, B:E471, B:N472, B:Q473, B:P474, B:S475, B:V476, B:A477, B:G478, B:I479, B:K480, B:A481, B:T482, B:R483, B:K484, B:K485, B:S486, B:N487, B:K488, B:K489, B:R490, B:R491	98	0.737
4	B:G275, B:P276, B:G277, B:F278	4	0.696
5	B:F261, B:D264, B:F265	3	0.639
6	B:D268, B:L271, B:T272	3	0.632
7	B:A122, B:A123, B:G124, B:A125	4	0.582
8	B:W221, B:P223, B:Q224, B:A225, B:W228	5	0.568
9	B:L186, B:R199, B:R200, B:N201, B:L202, B:L203, B:Y207, B:R210, B:P250, B:Q251, B:M252, B:G253, B:P254, B:G255, B:P256, B:G257, B:D258, B:W259	18	0.563
10	B:K107, B:V108, B:E111	3	0.561
11	B:A100, B:P101, B:K102, B:P103, B:L104, B:L105	6	0.555
12	B:P337, B:G338, B:G339, B:N342	4	0.545
13	B:P297, B:G298, B:F299	3	0.53

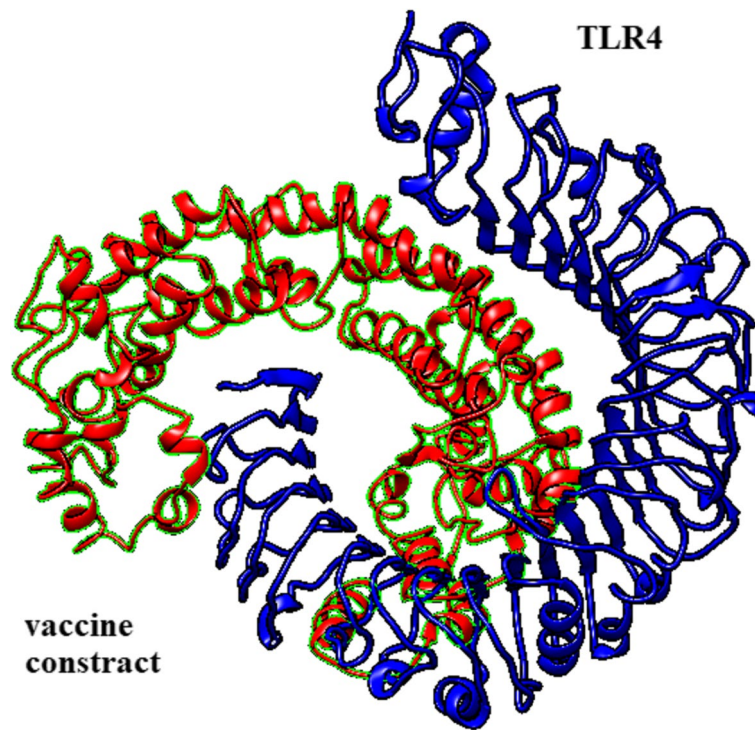


Fig. 8 Three-dimensional representation of molecular docking of the vaccine construct and TLR complex

was observed in protein structure. The compactness of TLR and vaccine was evaluated using Radius of gyration (Rg) plot. The Rg of TLR was in the range of 3.25 to 3.35. The Rg of vaccine was in the range of 0.3 to 0.35. A stability in compactness of each protein was observed. The hydrogen bond between receptor and peptide and was

calculated. About 15 H bonds were formed between TLR and vaccine (Fig. 11).

Codon optimization and in silico cloning of the designed vaccine
Codon optimization is used to ameliorate gene expression and translation precision of the recombinant protein

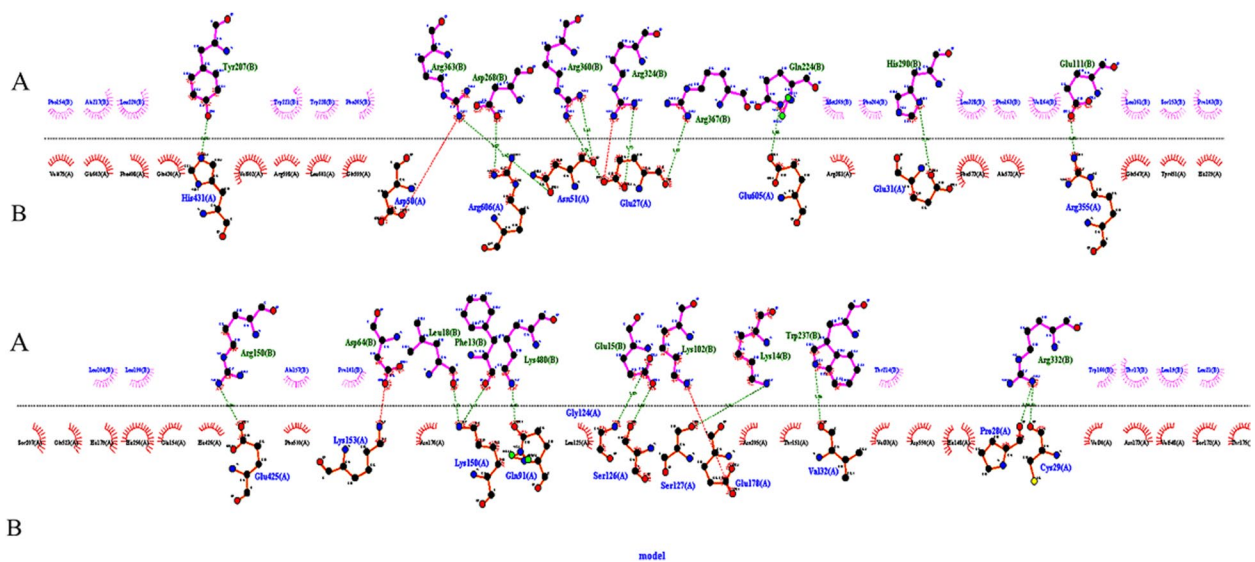


Fig. 9 Schematic of the interaction between TLR4 and the vaccine construct. **A** Amino acids involved in hydrogen bonding from Chain A of TLR4 **(B)** Amino acids of vaccine construct

Table 12 Hydrogen bonding interactions between TLR4 and vaccine amino acids

TLR4 (chain A)	Vaccine	Bond length (Å)
Tyr207	His431	3.02
Arg363	Asn51	2.67
Asp268	Arg606	2.87
Arg360	Asn51	2.62
Arg360	Glu27	2.76
Arg324	Glu27	2.73
Arg367	Glu27	2.87
Gln224	Glu605	2.88
His290	Glu31	3.14
Glu111	Arg355	2.72
Arg150	Glu425	2.74
Leu18	Lys150	2.53
Phe13	Lys150	2.67
Lsy480	Gln91	2.50
Glu15	Gly124	2.89
Glu15	Ser126	2.85
Lys14	Ser127	2.76
Trp237	Val32	2.86
Arg332	Pro28	2.86
Arg332	Cys29	2.81

by adapting to the target host’s codon bias. Reverse translation of the predicted vaccine was performed to achieve maximum expression in *Escherichia coli* strain K12 by Jcat server. After codon optimization, CAI score and GC content in improved protein sequence were estimated at 1.0 and 52.96, respectively. The data demonstrate that the improved protein sequence could be expressed sustainably in the *E. coli* system. Finally, the improved sequence was successfully integrated into the pET30a (+) vector by SnapGene program (Fig. 12).

Discussion

Conventional techniques for vaccine development involve the use of whole organisms, which can lead to undesired exposure to antigens and may trigger allergic responses. To prevent allergic responses, peptide-based vaccines that included short peptide fragments derived from immunogenic proteins have been used to produce strong and targeted immune reactions. Rabies, rubella, yellow fever, smallpox, hepatitis A/B, chickenpox, polio, influenza, Human Papillomavirus, and Japanese encephalitis are some of the infectious diseases which vaccines are highly effective against [96, 97]. The development of vaccines involves complex, time-consuming, and expensive in vitro and in vivo assays to

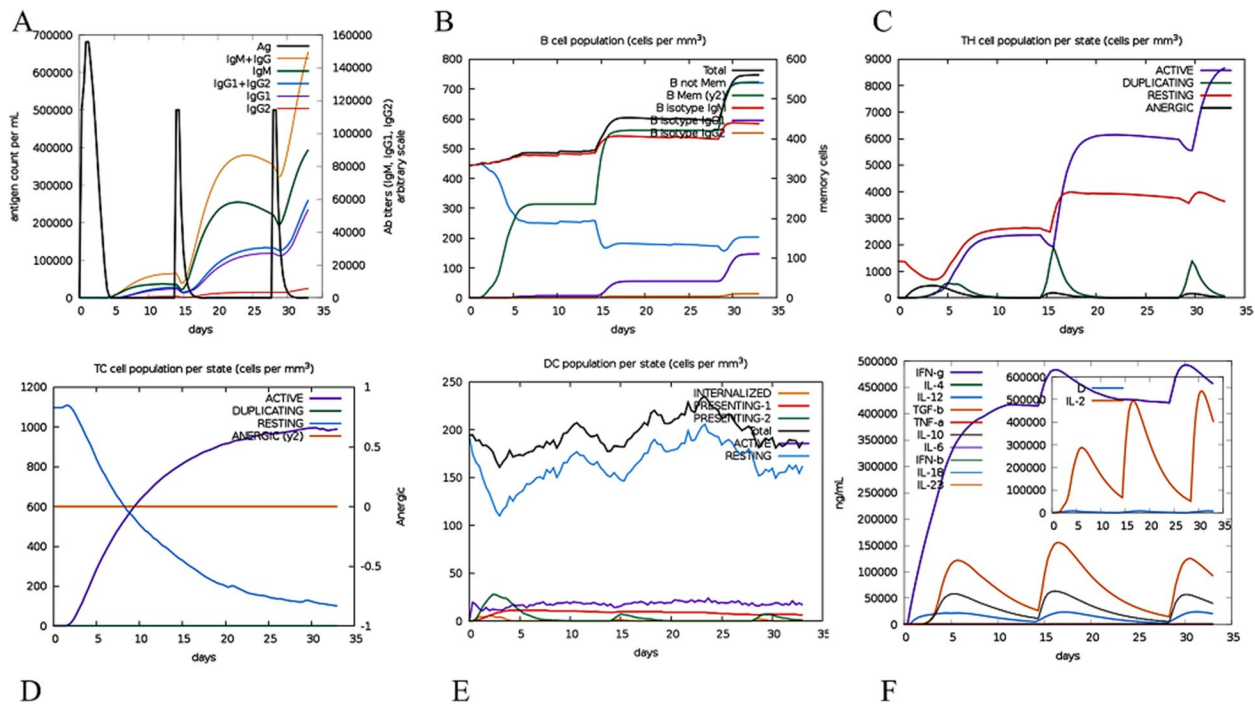


Fig. 10 Results of the in silico immune simulation using the C-ImmSim server for the designed vaccine. **A** the generation of immune complex and immunoglobulin as a result of response to designed vaccine injections, **B** B lymphocyte total count after the three injections, **C** growth of CD4 T-helper lymphocytes after the three injections including active, duplicating, resting, anergic, **D** Increasing the number of cytotoxic CD8 lymphocytes after injection of the designed vaccine, **E** Proliferation of dendritic cells after immunization, **F** Stimulation of cytokines and interleukins after vaccine administration

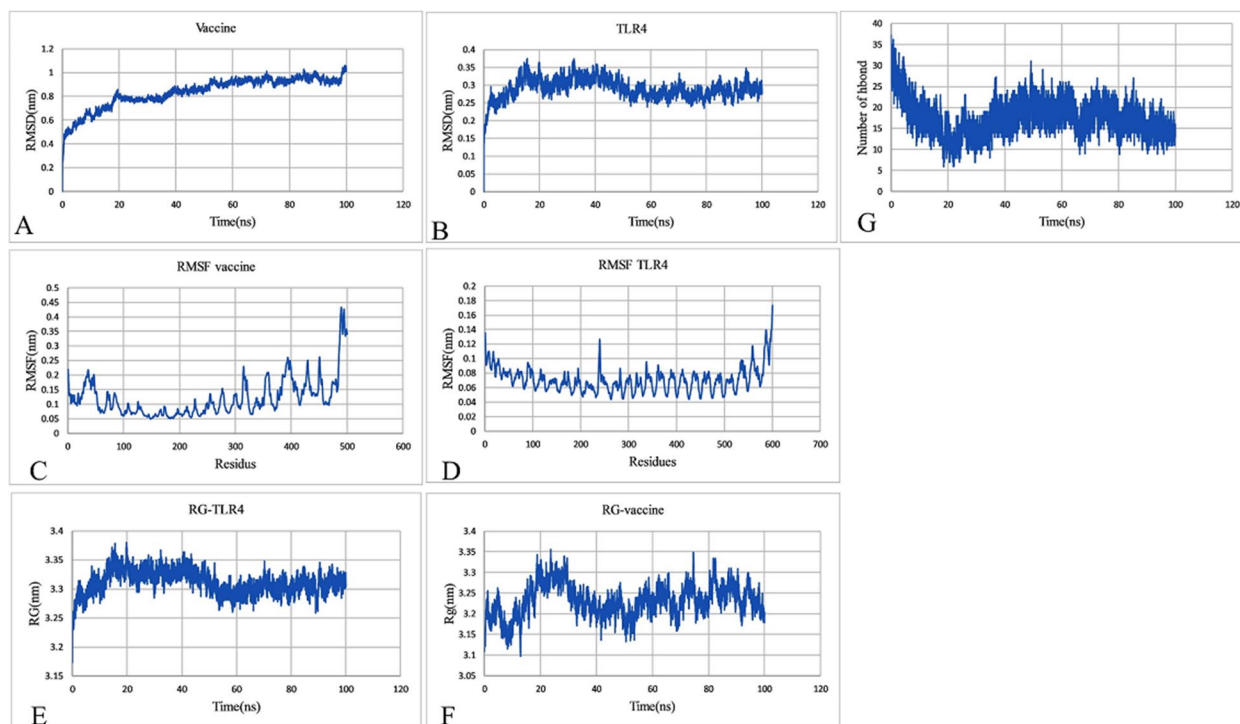


Fig. 11 Molecular dynamics simulation of the TLR4 complex and vaccine construct. **A** RMSD plot of the vaccine construct, **B** RMSD plot of the TLR4, **C** RMSF plot of the vaccine, **D** RMSF plot of the TLR4, **E** and **F** radius of gyration of the vaccine-TLR4 complex, and **G** hydrogen bond analysis from the simulation system

ensure vaccine effectiveness [98]. Current advances in immunoinformatics and computational biology allow the design of effective vaccines *in silico* and reduce the number of *in vitro* experiments [99]. Using an *in vitro* study, an experimentally validated multiepitope vaccine was designed against *Clostridium perfringens* [100]. With this method, a wide range of vaccine candidates can be identified without the requirement of cultivating pathogenic organisms [98].

Human polyomavirus 6 (HPyV6) and HPyV7 are polyomaviruses species initially discovered in the skin of healthy people [101]. The role of HPyV6 and 7 proteins in binding and inactivating p53 has been documented, suggesting its oncogenic role [102]. The incidence of malignant skin tumors has increased over recent decades, chiefly as a result of alterable exposures. The World Health Organization reported that about 8500 new cases of skin tumors are diagnosed every day in the U.S. [7, 8]. Several studies have shown the prevalence of HPyV6- and 7 in primary cutaneous malignancies, including actinic keratosis, basal cell carcinoma, bone marrow transplantation, neuroendocrine, and lymphoid skin cancers [7].

Therefore, in the present study, the POLY capsid protein VP1, POLY minor capsid protein VP2, and POLY large T antigen from HPyV6 and HPyV7 were examined

as candidate antigens for epitope identification. The allergenicity, toxicity, and antigenicity of the identified epitopes were assessed. There are a number of factors to consider when making peptide-based vaccines, including the intrinsic properties of the selected epitopes, adjuvant, and linker, and their arrangement and location within the protein. Based on the findings from the studies conducted by Olugbenga et al. [103], Mahnoor Majid et al. [104], and Sami et al. [99], we used KK, GPGPG, and AAY linkers to fuse LBL, HTL, and CTL epitopes, respectively. Epitope presentation is promoted with AAY and GPGPG linkers, while junctional epitopes are reduced with these linkers [105, 106]. The KK linker, a bi-lysine basic linker, preserved the immunogenic properties of B cell epitopes while keeping the pH near physiological levels [107, 108].

Compared with live attenuated vaccines, computational vaccines have relatively low immunogenicity. In order to address this problem, adjuvants are routinely employed. Hence, adjuvants have been widely used to increase vaccine effectiveness. Adjuvants generally function by activating innate immune cells through pathogen associated molecular pattern receptors. Adjuvants can also improve vaccines by stabilizing the epitope structure of the vaccine antigen, creating a suitable source for the gradual release of the antigen, better presenting the antigen to the

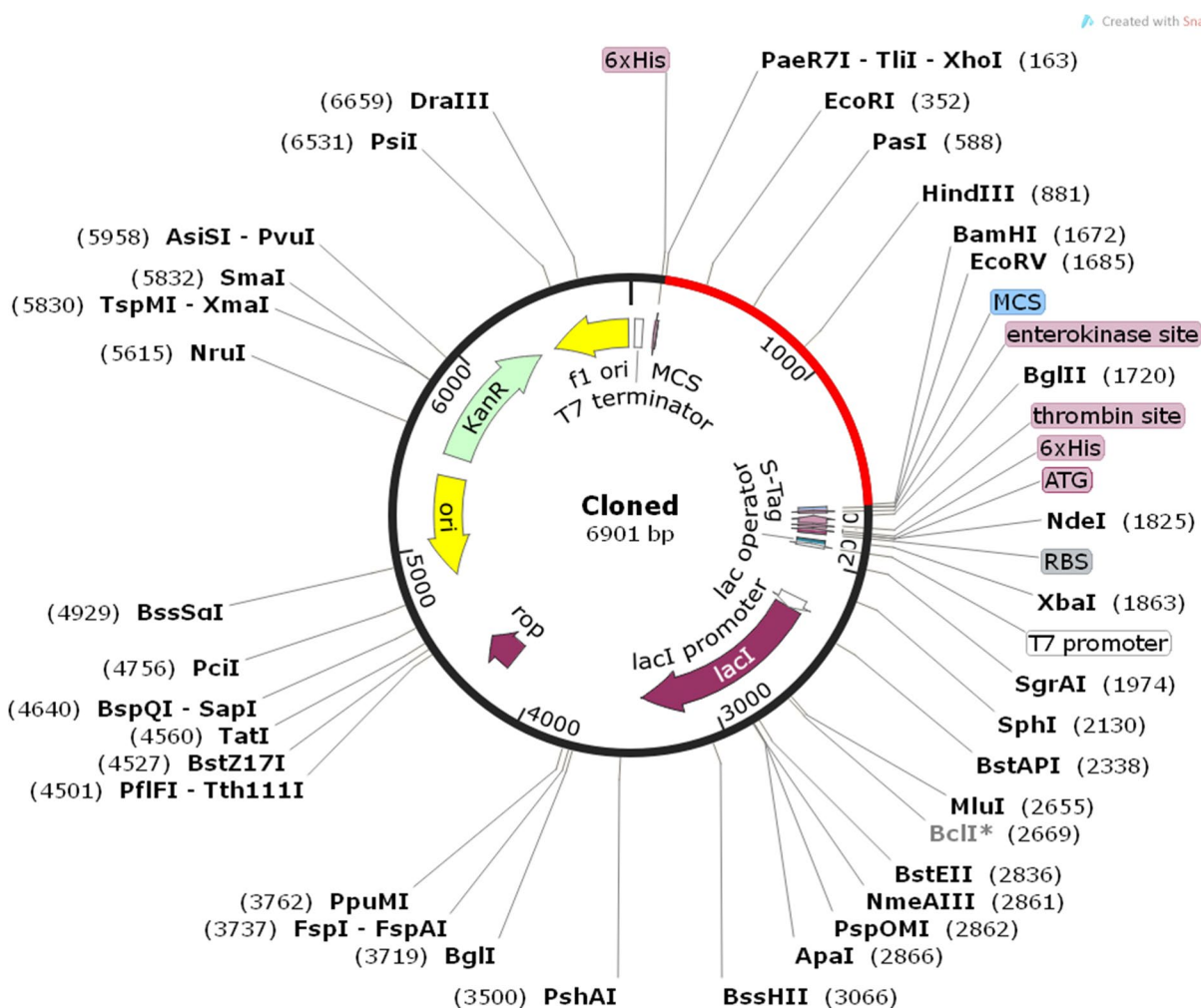


Fig. 12 The map of the in silico cloning of the vaccine construct into the pET30a (+) vector using SnapGene software. The black segment indicates the backbone of the vector and the red segment shows the vaccine construct. This vaccine construct contains restriction sites for XhoI and BamHI restriction enzymes at the 5' and 3' ends, respectively

antigen-presenting cells (APC), increasing the absorbing molecules of these cells at the site of the vaccine, and the proper binding of the antigen to these cells improves the vaccine performance [109]. The 50S ribosomal protein L7/L12 (Locus RL7_MYCTU) from Mycobacterium tuberculosis is a TLR4 agonist [110]. Thus, in order to enhance the immunogenicity of vaccine, we used it as an adjuvant. EAAAK, an empirical α -helical linker, reduces the connection with other protein regions while providing rigidity and improving chimeric protein durability [111]. Multiple servers determined that the construct vaccine was non-allergenic and highly antigenic, demonstrating that triggers robust immune responses without inducing unwanted allergies.

The final vaccine had a theoretical pI of 8.3, indicating its alkaline nature. Furthermore, the vaccine construct exhibited an average molecular weight of 54.05

KDa, indicating its favorable antigenic characteristics [112]. According to standards, proteins with a molecular weight below 110 kDa are deemed appropriate vaccine candidates [113]. The instability index of the vaccine was measured as 34.63, Values less than 40 are considered as a stable protein in biological environments [114]. The constructed vaccine has an indicated average half-life of above 20, 10, and 30 h in yeast cells (in vivo), E. coli (in vivo), and mammalian reticulocytes (in vitro), respectively. On the basis of previous findings, these half-life results are acceptable [99, 115]. The aliphatic index was 72.95, indicating the constructed vaccine would be thermostable at natural human body temperature [116]. The GRAVY value of the protein was -0.317, indicating the hydrophilic nature of the vaccine. Vaccine formulation and purification are made easier by the strong affinity for water molecules [117, 118].

After making the 3D model of the vaccine, the refinement system is employed to enhance its quality, both in terms of global and local structures. Validation of the model is necessary to accurately compare the unrefined structure with the refined structure. The Ramachandran plot indicated that 73.4% of the amino acid residues in the unrefined structure were detected in the desired region, while 87.7% of the amino acid residues in the refined structure were placed in the desired region, demonstrating improvement in the refined structure. Assessment of the immune response induced by an antigen is one of the primary characteristics in the validation of an introduced vaccine [119]. Molecular analyzes were employed to investigate the molecular connection between the formulated vaccine and TLR4, and suitable interactions were detected with a strong affinity score of -1414.0 kcal/mol. This relationship of the engineered vaccine with TLR-4 demonstrated that the recombinant protein vaccine has the capacity to stimulate an innate and adaptive immune response. To investigate the stability and dynamic efficiency of the vaccine/TLR4 complex, MD simulation was performed and the RMSD diagram confirmed the stable binding of this compound.

An appropriate host is required for the expression of recombinant protein. *E. coli* expression systems is the most common host for expressing recombinant proteins [120, 121]. To enable the recombinant vaccine to be expressed at high levels in *E. coli* (K12 strain) codon optimization was performed. An analysis of the designed vaccine indicated a CAI score of 1.0 and GC content of 52.96. CAI values of more than 0.8 and GC content of 30–70% have been reported to favor high expression in the *E. coli* host [122, 123].

Conclusion

Human polyomaviruses (HPyVs) infect a wide range of tissues such as skin, kidney and respiratory tract and often lead to persistent and asymptomatic infection, while these infections can lead to cancer. Currently, no significant therapeutic vaccine is available for HPyV. In this study, immunoinformatics techniques were applied to identify and refine candidate vaccines against HPyV. The highly immunogenic T and B cell epitopes were identified and used for vaccine design. The proposed vaccine is projected to produce robust immune reactions, including cytokines, and interferons. The binding analysis confirmed the vaccine binding to the immune receptor TLR4 that was dynamically stable. Although experimental trials in appropriate animal models is necessary to test the potency of the engineered vaccine, analysis using different bioinformatics tools indicated the high immunogenicity and preventive potential of the developed vaccine.

Abbreviations

HPyV6	Human polyomaviruses 6
HPyV7	Human polyomaviruses 7
CTLs	Cytotoxic T-lymphocytes
HTLs	Helper T cells
LBL	Linear B-Lymphocyte
LTA _g	Large T antigen
VP1	POLY capsid protein
VP2	POLY minor capsid protein
MHC I	Major histocompatibility complex (MHC) class I
MHC II	Major histocompatibility complex (MHC) class II
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
R _g	Radius of gyration
TLR4	Toll like receptor 4
CAI	Codon Adaptation Index
GRAVY	Grand average of hydropathicity
pI	Isoelectric point
MD	Molecular dynamics

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-024-09046-0>.

Additional file 1.

Acknowledgements

The authors acknowledge support provided by the Tabriz University of Medical Sciences, Tabriz, Iran.

Authors' contributions

R.S: Data modeling, interpretation of data, data curation, and manuscript writing; S. F: have designed and supervised the work; N. B: Manuscript reviewing and editing; N. B: wrote the manuscript; F. E: wrote the manuscript; M. S: Analyzed the data. S. F. and S.V: edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by a grant from Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran (Grant no. 68980).

Availability of data and materials

Datasets used in the experiments are listed as follows: (1) NCBI: <https://www.ncbi.nlm.nih.gov/> (2) UniProt database: <https://www.uniprot.org/> (3) VaxiJen server: <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html> (4) AllergenFP v1.0 server: <https://ddg-pharmfac.net/AllergenFP/> (5) TMHMM v2.0 tool: <https://services.healthtech.dtu.dk/services/TMHMM-2.0/> (6) NetCTL 1.2 server: <https://services.healthtech.dtu.dk/services/NetCTL-1.2/> (7) IEDB server: <http://tools.iedb.org/mhci/> (8) AllerTOP v2.0 server: <https://www.ddg-pharmfac.net/AllerTOP/> (9) ToxinPred server: <http://crdd.osdd.net/raghava/toxinpred/> (10) Class I Immunogenicity: <http://tools.iedb.org/immunogenicity/> (11) IEDB resource <http://tools.iedb.org/mhcii/> (12) IFNepitope tool: <http://crdd.osdd.net/raghava/ifnepitope/design.php> (13) IL4pred tool: <https://webs.iitd.edu.in/raghava/il4pred/> (14) ABCpred tool: <http://crdd.osdd.net/raghava/abcpred/> (15) epitope conservancy analysis: <http://tools.iedb.org/conservancy/> (16) BLAST server: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (17) PEP-FOLD v3.5 tool: <https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/> (18) HADDOCK tool: <https://wenmr.science.uu.nl/haddock2.4/> (19) RCSB Protein Data Bank: <https://www.rcsb.org/> (20) IEDB population coverage server: <http://tools.iedb.org/population/> (21) MHCcluster 2.0 tool: <https://services.healthtech.dtu.dk/services/MHCcluster-2.0/> (22) ExPASy ProtParam tool: <https://web.expasy.org/protparam/> (23) ANTIGENPro server: <https://scratch.proteomics.ics.uci.edu/> (24) AlgPred: <http://crdd.osdd.net/raghava/algpred/> (25) SOLpro tool: <https://scratch.proteomics.ics.uci.edu/> (26) SignalP 4.1 tool: <https://services.healthtech.dtu.dk/services/SignalP-4.1/> (27) PSIPRED v4.0: <http://bioinf.cs.ucl.ac.uk/psipred/> (28) Prabi: https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html (29) I-TASSER tool: <https://seq2fun.dcm.b.med.umich.edu/I-TASSER/> (30) GalaxyRefine tool: <https://>

galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE (31) ProSA-web: <https://prosa.services.came.sbg.ac.at/prosa.php> (32) Procheck server: <https://saves.mbi.ucla.edu/> (33) ElliPro server: <http://tools.iedb.org/ellipro/> (34) ClusPro: <https://cluspro.bu.edu/login.php> (35) JCat: <http://www.jcat.de/> (36) C-ImmSim server: <https://kraken.iac.rm.cnr.it/C-IMMSIM/index.php?page=1>

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Declarations

Ethics approval and consent to participate

This work was approved by Ethical committee of Tabriz University of Medical Sciences, Tabriz, Iran (IR.TBZMED.AEC.1400.016).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran. ²Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran. ³Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran. ⁴Department of Medical Biotechnology, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran. ⁵Department of Microbiology and Immunology, Geisel School of Medicine, and Dartmouth Cancer Center, Lebanon, NH, USA. ⁶Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada.

Received: 30 October 2023 Accepted: 22 January 2024

Published online: 09 February 2024

References

- Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, et al. Taxonomical developments in the family Polyomaviridae. *Arch Virol*. 2011;156(9):1627–34.
- Condez AC, Nunes M, Filipa-Silva A, Leonardo I, Parreira R. Human Polyomaviruses (HPyV) in Wastewater and Environmental Samples from the Lisbon Metropolitan Area: Detection and Genetic Characterization of Viral Structural Protein-Coding Sequences. *Pathogens*. 2021;10(10).
- van der Meijden E, Feltkamp M. The human polyomavirus middle and alternative T-Antigens; thoughts on roles and relevance to cancer. *Front Microbiol*. 2018;9:398.
- Prado JCM, Monezi TA, Amorim AT, Lino V, Paladino A, Boccardo E. Human polyomaviruses and cancer: an overview. *Clinics (Sao Paulo)*. 2018;73(suppl 1):e558s.
- Torres C. Evolution and molecular epidemiology of polyomaviruses. *Infect Genet Evol*. 2020;79:104150.
- Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, et al. The ancient evolutionary history of polyomaviruses. *PLoS Pathog*. 2016;12(4):e1005574.
- Klufah F, Mobaraki G, Liu D, Alharbi RA, Kurz AK, Speel EJM, et al. Emerging role of human polyomaviruses 6 and 7 in human cancers. *Infect Agent Cancer*. 2021;16(1):35.
- Higgins S, Nazemi A, Chow M, Wysong A. Review of nonmelanoma skin cancer in African Americans, Hispanics, and Asians. *Dermatol Surg*. 2018;44(7):903–10.
- Schrama D, Groesser L, Ugurel S, Hafner C, Pastrana DV, Buck CB, et al. Presence of human polyomavirus 6 in mutation-specific BRAF inhibitor-induced epithelial proliferations. *JAMA Dermatol*. 2014;150(11):1180–6.
- Beckervordersandforth J, Pujari S, Rennspiess D, Speel EJ, Winne-penninckx V, Diaz C, et al. Frequent detection of human polyomavirus 6 in keratoacanthomas. *Diagn Pathol*. 2016;11(1):58.
- Ehlers B, Wieland U. The novel human polyomaviruses HPyV6, 7, 9 and beyond. *APMIS*. 2013;121(8):783–95.
- Nguyen KD, Lee EE, Yue Y, Stork J, Pock L, North JP, et al. Human polyomavirus 6 and 7 are associated with pruritic and dyskeratotic dermatoses. *J Am Acad Dermatol*. 2017;76(5):932–40.e3.
- Hashida Y, Higuchi T, Matsuzaki S, Nakajima K, Sano S, Daibata M. Prevalence and genetic variability of human polyomaviruses 6 and 7 in healthy skin among asymptomatic individuals. *J Infect Dis*. 2018;217(3):483–93.
- Purdie KJ, Proby CM, Rizvi H, Griffin H, Doorbar J, Sommerlad M, et al. The role of human papillomaviruses and polyomaviruses in BRAF-inhibitor induced cutaneous squamous cell carcinoma and benign squamoproliferative lesions. *Front Microbiol*. 2018;9:1806.
- Hampras SS, Locke FL, Chavez JC, Patel NS, Giuliano AR, Miller K, et al. Prevalence of cutaneous viral infections in incident cutaneous squamous cell carcinoma detected among chronic lymphocytic leukemia and hematopoietic stem cell transplant patients. *Leuk Lymphoma*. 2018;59(4):911–7.
- Du-Thanh A, Foulongne V, Guillot B, Dereure O. Recently discovered human polyomaviruses in lesional and non-lesional skin of patients with primary cutaneous T-cell lymphomas. *J Dermatol Sci*. 2013;71(2):140–2.
- Poluschkin L, Rautava J, Turunen A, Wang Y, Hedman K, Syrjänen K, et al. Polyomaviruses detectable in head and neck carcinomas. *Oncotarget*. 2018;9(32):22642–52.
- Rennspiess D, Pujari S, Keijzers M, Abdul-Hamid MA, Hochstenbag M, Dingemans AM, et al. Detection of human polyomavirus 7 in human thymic epithelial tumors. *J Thorac Oncol*. 2015;10(2):360–6.
- Klufah F, Mobaraki G, Chteinberg E, Alharbi RA, Winne-penninckx V, Speel EJM, et al. High Prevalence of Human Polyomavirus 7 in Cholangiocarcinomas and Adjacent Peritumoral Hepatocytes: Preliminary Findings. *Microorganisms*. 2020;8(8).
- Muller PA, Vousden KH. p53 mutations in cancer. *Nat Cell Biol*. 2013;15(1):2–8.
- Moens U, Prezioso C, Pietropaolo V. Functional domains of the early proteins and experimental and epidemiological studies suggest a role for the novel human polyomaviruses in cancer. *Front Microbiol*. 2022;13:834368.
- Borchert S, Czech-Sioli M, Neumann F, Schmidt C, Wimmer P, Dobner T, et al. High-affinity Rb binding, p53 inhibition, subcellular localization, and transformation by wild-type or tumor-derived shortened Merkel cell polyomavirus large T antigens. *J Virol*. 2014;88(6):3144–60.
- White MK, Gordon J, Khalili K. The rapidly expanding family of human polyomaviruses: recent developments in understanding their life cycle and role in human pathology. *PLoS Pathog*. 2013;9(3):e1003206.
- Hirsch HH, Babel N, Comoli P, Friman V, Ginevri F, Jardine A, et al. European perspective on human polyomavirus infection, replication and disease in solid organ transplantation. *Clin Microbiol Infect*. 2014;20(Suppl 7):74–88.
- Nabel GJ. HIV vaccine strategies. *Vaccine*. 2002;20(15):1945–7.
- Oli AN, Obialor WO, Ifeanyi-chukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, et al. Immunoinformatics and vaccine development: an overview. *Immunotargets Ther*. 2020;9:13–30.
- Ahmad B, Ashfaq UA, Rahman MU, Masoud MS, Yousof MZ. Conserved B and T cell epitopes prediction of ebola virus glycoprotein for vaccine development: an immuno-informatics approach. *Microb Pathog*. 2019;132:243–53.
- Huang S, Zhang C, Li J, Dai Z, Huang J, Deng F, et al. Designing a multi-epitope vaccine against coxsackievirus B based on immunoinformatics approaches. *Front Immunol*. 2022;13:933594.
- Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(Web Server issue):W5–9.
- UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158–d69.
- Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.
- Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*. 2010;26(23):2936–43.

33. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*. 2014;30(6):846–51.
34. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305(3):567–80.
35. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. 2020;63(3):457–60.
36. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer*. 2021;124(2):359–67.
37. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 2007;8:424.
38. Moutafsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui HH, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol*. 2006;24(7):817–9.
39. Dimitrov I, Flower DR, Doytchinova I. AllerTOP—a server for in silico prediction of allergens. *BMC Bioinformatics*. 2013;14 Suppl 6(Suppl 6):S4.
40. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE*. 2013;8(9):e73957.
41. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013;9(10):e1003266.
42. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*. 2010;11:568.
43. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol*. 2008;4(4):e1000048.
44. Tahir UI Qamar M, Rehman A, Tusleem K, Ashfaq UA, Qasim M, Zhu X, et al. Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. *PLoS One*. 2020;15(12):e0244176.
45. Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. *Clin Dev Immunol*. 2013;2013:263952.
46. Bhuiyan MA, Quayum ST, Ahammad F, Alam R, Samad A, Nain Z. Discovery of potential immune epitopes and peptide vaccine design—a prophylactic strategy against Rift Valley fever virus. *F1000Research*. 2020;9:999.
47. Saha S, Raghava GP. Prediction methods for B-cell epitopes. *Methods Mol Biol*. 2007;409:387–94.
48. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006;65(1):40–8.
49. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP vol 2—a server for in silico prediction of allergens. *J Mol Model*. 2014;20(6):2278.
50. Bui HH, Sidney J, Li W, Füsseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2007;8:361.
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
52. Mehla K, Ramana J. Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a comparative genomics and immunoinformatics approach. *Mol Biosyst*. 2016;12(3):890–901.
53. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res*. 2016;44(W1):W449–54.
54. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18(15):2714–23.
55. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res*. 2000;28(1):235–42.
56. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–12.
57. Honorato RV, Koukos PI, Jiménez-García B, Tsaregorodtsev A, Verlatto M, Giachetti A, et al. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front Mol Biosci*. 2021;8:729513.
58. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastriitis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol*. 2016;428(4):720–5.
59. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 1995;8(2):127–34.
60. Adhikari UK, Tayebi M, Rahman MM. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus. *J Immunol Res*. 2018;2018:6718083.
61. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*. 2006;7:153.
62. Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules. *Immunogenetics*. 2013;65(9):655–65.
63. Hasan M, Ghosh PP, Azim KF, Mukta S, Abir RA, Nahar J, et al. Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. *Microb Pathog*. 2019;130:19–37.
64. Dorostii H, Eslami M, Negahdaripour M, Ghoshon MB, Gholami A, Heidari R, et al. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *J Biomol Struct Dyn*. 2019;37(13):3524–35.
65. Nain Z, Abdulla F, Rahman MM, Karim MM, Khan MSA, Sayed SB, et al. Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis* using immunoinformatic approaches. *J Biomol Struct Dyn*. 2020;38(16):4850–67.
66. Pandey RK, Bhatt TK, Prajapati VK. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating *Anopheles salivary* protein. *Sci Rep*. 2018;8(1):1125.
67. Olejnik J, Hume AJ, Mühlberger E. Toll-like receptor 4 in acute viral infection: too much of a good thing. *PLoS Pathog*. 2018;14(12):e1007390.
68. Abdellrazeq GS, Fry LM, Elnaggar MM, Bannantine JP, Schneider DA, Chamberlin WM, et al. Simultaneous cognate epitope recognition by bovine CD4 and CD8 T cells is essential for primary expansion of antigen-specific cytotoxic T-cells following ex vivo stimulation with a candidate *Mycobacterium avium* subsp. *paratuberculosis* peptide vaccine. *Vaccine*. 2020;38(8):2016–25.
69. Borthwick N, Silva-Arrieta S, Llano A, Takiguchi M, Brander C, Hanke T. Novel Nested Peptide Epitopes Recognized by CD4(+) T Cells Induced by HIV-1 Conserved-Region Vaccines. *Vaccines (Basel)*. 2020;8(1).
70. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the Expasy server. *The proteomics protocols handbook*. 2005:571–607.
71. Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34(Web Server issue):W202–9.
72. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics*. 2009;25(17):2200–7.
73. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*. 2017;33(19):3098–100.
74. Niwa T, Ying BW, Saito K, Jin W, Takada S, Ueda T, et al. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc Natl Acad Sci U S A*. 2009;106(11):4201–6.
75. Nielsen H. Predicting Secretory Proteins with SignalP. *Methods Mol Biol*. 2017;1611:59–73.
76. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
77. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. *Febs j*. 2005;272(20):5101–9.
78. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. *J Autoimmun*. 2018;95:100–23.
79. Buchan DW, Minnici F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res*. 2013;41(Web Server issue):W349–57.

80. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* 1996;266:540–53.
81. Montgomerie S, Sundararaj S, Gallin WJ, Wishart DS. Improving the accuracy of protein secondary structure prediction using structural alignment. *BMC Bioinformatics.* 2006;7:301.
82. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725–38.
83. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015;12(1):7–8.
84. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008;9:40.
85. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* 2012;40(Web Server issue):W294–7.
86. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35(Web Server issue):W407–10.
87. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins.* 2003;50(3):437–50.
88. Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Structural Biology.* 7(1):1–19.
89. Kozakov D, Hall DR, Xia B, Porter KA, Padhorna D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12(2):255–78.
90. Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. Performance and its limits in rigid body protein-protein docking. *Structure.* 2020;28(9):1071–81.e3.
91. Del Tordello E, Rappuoli R, Delany I. Reverse vaccinology: exploiting genomes for vaccine design. 2017:65–86.
92. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem.* 1989;182(2):319–26.
93. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1:19–25.
94. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29(7):845–54.
95. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 2005;33(Web Server issue):W526–31.
96. Plotkin SA. Vaccines: past, present and future. *Nat Med.* 2005;11(4 Suppl):S5–11.
97. Kash N, Lee MA, Kollipara R, Downing C, Guidry J, Tyring SK. Safety and efficacy data on vaccines and immunization to human papillomavirus. *J Clin Med.* 2015;4(4):614–33.
98. Mugunthan SP, Harish MC. Multi-epitope-based vaccine designed by targeting cytoadherence proteins of mycoplasma gallisepticum. *ACS Omega.* 2021;6(21):13742–55.
99. Sami SA, Marma KKS, Mahmud S, Khan MAN, Albogami S, El-Shehawi AM, et al. Designing of a multi-epitope vaccine against the structural proteins of Marburg virus exploiting the immunoinformatics approach. *ACS Omega.* 2021;6(47):32043–71.
100. Katalani C, Nematzadeh G, Ahmadian G, Amani J, Kiani G, Ehsani P. In silico design and in vitro analysis of a recombinant trivalent fusion protein candidate vaccine targeting virulence factor of *Clostridium perfringens*. *Int J Biol Macromol.* 2020;146:1015–23.
101. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe.* 2010;7(6):509–15.
102. Prezioso C, Van Ghelue M, Moens U, Pietropaolo V. HPyV6 and HPyV7 in urine from immunocompromised patients. *Viral J.* 2021;18(1):24.
103. Onile OS, Musaigwa F, Ayawei N, Omoboyede V, Onile TA, Oghenevwo E, et al. Immunoinformatics Studies and Design of a Potential Multi-Epitope Peptide Vaccine to Combat the Fatal Visceral Leishmaniasis. *Vaccines (Basel).* 2022;10(10).
104. Majid M, Andleeb S. Designing a multi-epitopic vaccine against the enterotoxigenic *Bacteroides fragilis* based on immunoinformatics approach. *Sci Rep.* 2019;9(1):19780.
105. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, et al. A rational strategy to design multi-epitope immunogens based on multiple Th lymphocyte epitopes. *J Immunol.* 2002;168(11):5499–506.
106. Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y. Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. *Comput Biol Chem.* 2016;62:82–95.
107. Negahdaripour M, Nezafat N, Eslami M, Ghoshon MB, Shoolian E, Najafipour S, et al. Structural vaccinology considerations for in silico designing of a multi-epitope vaccine. *Infect Genet Evol.* 2018;58:96–109.
108. Gu Y, Sun X, Li B, Huang J, Zhan B, Zhu X. Vaccination with a par-amyosin-based multi-epitope vaccine elicits significant protective immunity against trichinella spiralis infection in mice. *Front Microbiol.* 2017;8:1475.
109. Awate S, Babiuk LA, Mutwiri G. Mechanisms of action of adjuvants. *Front Immunol.* 2013;4:114.
110. Lee SJ, Shin SJ, Lee MH, Lee MG, Kang TH, Park WS, et al. A potential protein adjuvant derived from *Mycobacterium tuberculosis* Rv0652 enhances dendritic cells-based tumor immunotherapy. *PLoS ONE.* 2014;9(8):e104351.
111. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013;65(10):1357–69.
112. Khatoun N, Pandey RK, Prajapati VK. Exploring Leishmania secretory proteins to design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Sci Rep.* 2017;7(1):8285.
113. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, et al. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infect Genet Evol.* 2015;32:280–91.
114. Guruprasad K, Reddy BV, Pandit MW. 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.* 1990;4(2):155–61.
115. Shah SZ, Jabbar B, Mirza MU, Waqas M, Aziz S, Halim SA, et al. An Immunoinformatics Approach to Design a Potent Multi-Epitope Vaccine against Asia-1 Genotype of Crimean-Congo Haemorrhagic Fever Virus Using the Structural Glycoproteins as a Target. *Vaccines (Basel).* 2022;11(1).
116. Chaudhri G, Quah BJ, Wang Y, Tan AH, Zhou J, Karupiah G, et al. T cell receptor sharing by cytotoxic T lymphocytes facilitates efficient virus control. *Proc Natl Acad Sci U S A.* 2009;106(35):14984–9.
117. Chang KY, Yang JR. Analysis and prediction of highly effective antiviral peptides based on random forests. *PLoS ONE.* 2013;8(8):e70166.
118. Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol.* 1982;157(1):105–32.
119. Gori A, Longhi R, Peri C, Colombo G. Peptides for immunological purposes: design, strategies and applications. *Amino Acids.* 2013;45(2):257–68.
120. Chen R. Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnol Adv.* 2012;30(5):1102–7.
121. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 2014;5:172.
122. Morla S, Makhija A, Kumar S. Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene.* 2016;584(1):1–6.
123. Ali M, Pandey RK, Khatoun N, Narula A, Mishra A, Prajapati VK. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci Rep.* 2017;7(1):9232.

Publisher's Note

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