VIM-2020-0306-ver9-Ahmad_1P
Type: research-article

VIRAL IMMUNOLOGY
Volume 00, Number 00, 2021

Pp. 1–13
DOI: 10.1089/vim.2020.0306

AU1 ▶

Computational Modeling of Immune Response Triggering Immunogenic Peptide Vaccine Against the Human Papillomaviruses to Induce Immunity Against Cervical Cancer

AU2 ▶ Namra Ahmad,[1] Syed Shujait Ali,[1] Sajjad Ahmad,[2] Zahid Hussain,[1] Muhammad Qasim,[3]
AU3 ▶ Muhammad Suleman,[1] Shahid Ali,[1] N. Nizam-Uddin,[4] Abbas Khan,[5,*] and Dong-Qing Wei[5–7]

Abstract

Papillomaviruses are placed within the family Papillomaviride, and the members of this family have a double-stranded circular DNA genome. Every year, ~30% of cancers are reported to be human papillomavirus (HPV) related, which represents 63,000 cancers of all infectious agent-induced cancers. HPV16 and HPV18 are reported to be associated with 70% of cervical cancers. The quest for an effective drug or vaccine candidate still continues. In this study, we aim to design B cell and T cell epitope-based vaccine using the two structural major capsid protein L1 and L2 as well as other three important proteins (E1, E2, and E6) against HPV strain 16 (HPV16). We used a computational pipeline to design a multiepitope subunit vaccine and tested its efficacy using *in silico* computational modeling approaches. Our analysis revealed that the multiepitope subunit vaccine possesses antigenic properties, and using *in silico* cloning method revealed proper expression and downstream processing of the vaccine construct. Besides this, we also performed *in silico* immune simulation to check the immune response upon the injection. Our results strongly suggest that this vaccine candidate should be tested immediately for the immune response against the cervical cancer-causing agent. The safety, efficacy, expression, and immune response profiling makes it the first choice for experimental and *in vivo* setup.

Keywords: HPV, vaccine, docking, *in silico* cloning, immune simulation

Introduction

Papillomaviruses are placed within the family Papillomaviride, and the members of this family have a double-stranded circular DNA genome. Proteins E1, E2, E4, E5, E6, E7, E8, E2C, L1, and L2 are involved in viral gene expression, replication, and survival. However, E1, E2, L1, and L2, are found in the entire papillomaviruses identified to date (39) and are essential for processes like replication and shedding of the virus. Papillomaviruses are believed to be host restricted; however, some cases of cross-species transmission are also reported (6,11,12).

International human papillomavirus (HPV) reference center (www.hpvcenter.se) has reported >200 HPVs until 2013, and the list is still expanding (5,30). Alpha HPVs are considered as high-risk HPV types (HR HPVs) and are

[1]Center for Biotechnology and Microbiology, University of Swat, Swat, Pakistan.
[2]National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan.
[3]Department of Environmental and Conservation Sciences, University of Swat, Swat, Pakistan.
[4]Department of Biomedical Engineering, HITEC University, Taxila, Pakistan.
[5]Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, P.R. China.
[6]State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, P.R. China.
[7]Peng Cheng Laboratory, Shenzhen, P.R China.
*ORCID ID (https://orcid.org/0000-0002-4288-7602).

involved in anogenital, head, and neck cancers (4,15). Whereas beta HPVs have a possible role in nonmelanoma skin cancers, in addition to UV radiations (1).

Recent studies have suggested the role of E6 and E7 proteins of mucosal HR HPV types in the increase of cervical cancer by altering the pathways that regulate the human immune system response to promote cellular transformation and cause persistent infection (28). However, its cutaneous expression is also important for cellular transformation initiation and maintenance; meanwhile, for skin cancer, the expression of E6 and E7 is not crucial. Beta HPVs exacerbate the accumulation of somatic mutation and DNA breaks induced by ultraviolet radiations. Hence, it plays a vital role by acting as a facilitator in skin carcinogenesis (11).

Every year ~30% of cancer cases are caused by HPV (4). HPV16 and HPV18 are reported to be associated with 70% of cervical cancers (41). In this study, we aim to design B cell and T cell epitope-based vaccine using the two structural major capsid protein L1 and L2 as well as other three important proteins (E1, E2, and E6) against HPV strain 16 (HPV16). We used a computational pipeline to design a multiepitope subunit vaccine and tested its efficacy using *in silico* computational modeling approaches. Our analysis revealed that the multiepitope subunit vaccine possesses antigenic properties and using *in silico* cloning method revealed proper expression and downstream processing of the vaccine construct. Besides this, we also performed *in silico* immune simulation to check the immune response upon the injection. Our results strongly suggest that this vaccine candidate needs further scientific consensus by testing in experimental setup to be used against the cervical agents.

Materials and Methods

Retrieval of HPV16 polyprotein

Using the GenBank database of the NCBI (https://www.ncbi.nlm.nih.gov/), the entire structural polyprotein for HPV16 was recovered in FASTA format under the Accession number: NC_001526.4 (3). Figure 1 represents the overall methodology of the analysis.

Cytotoxic T lymphocyte epitope prediction

NetCTL 1.2 (www.cbs.dtu.dk/services/NetCTL/), an online web server, was used for the prediction of Cytotoxic T lymphocyte epitope (CTL epitopes) prediction of the HPV16. The prediction of epitopes is performed based on three key components that are Proteasomal C terminal cleavage, transportation efficiency transport associated with antigen processing (TAP), and MHCI-binding peptide prediction. The threshold was set to 0.75 to carry out the CTL epitope prediction (37).

Helper T lymphocyte epitope prediction

IEDB tool (www.iedb.org/) server was used to predict the helper T-lymphocyte epitopes (HTL epitopes) for all the proteins. The set of seven reference alleles provided by IEDB was used for the analysis. The peptide affinity for each of the receptors is based on the IC50 score given to each and every predicted epitope. Peptides with higher binding affinity are required to have an IC50 value <50 nM. Meanwhile, the IC50 score <500 nM points to an intermediate and <5,000 nM points to the epitopes' low binding affinity, respectively. The binding affinity and percentile rank of the predicted epitopes are linked inversely: lower

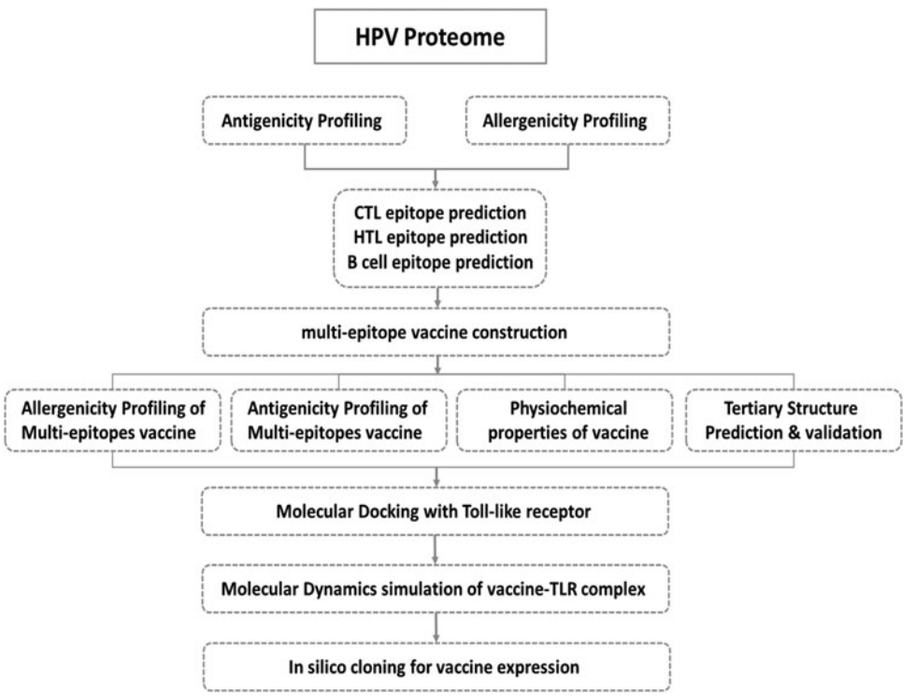

FIG. 1. Systematic flow diagram of the vaccine design for HPV.

Table 1. List of Candidate Proteins with Antigenicity Score Predicted Using VexiJen Server

Gene	*Protein name*	*Antigenicity*	*Score*
E1	Replication protein E1	ANTIGENIC	0.4470
E6	Protein E6	ANTIGENIC	0.6199
L2	Minor capsid protein L2	ANTIGENIC	0.6097
E2	Regulatory protein E2	ANTIGENIC	0.4273
L1	Major capsid protein L1	ANTIGENIC	0.5164

percentile rank higher binding affinity and vice versa. The same server was used for immunogenicity prediction as well (40).

Toxicity prediction

The predicted epitopes were submitted to ToxinPred (http://crdd.osdd.net/raghava/toxinpred/) for toxicity prediction. The server classifies the epitopes as toxic and nontoxic based on their physicochemical properties (14).

Construction of multiepitope vaccine sequence

The final predicted CTL and HTL epitopes, filtered through the discussed immunoinformatics strategy was used to construct a sequence. Beta-defensin was used as an adjuvant (16), and the epitopes were joined using AAY linkers for CTL epitopes and GPGPG linkers for HTL epitopes, respectively (34).

Allergenicity prediction of the vaccine

AlgPred (www.imtech.res.in/raghava/algpred/) webserver was utilized for allergenicity prediction. The server uses six different approaches for the allergenicity prediction with 85% accuracy at 0.4 thresholds (35).

Antigenicity prediction of the vaccine

VexiJen (www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) server was used for antigenicity prediction of the constructed vaccine with virus selected as a model organism and threshold set to 0.4. The server carries out the prediction of antigenic and nonantigenic amino acid sequences based on physicochemical properties only (46).

Physiochemical properties prediction

The physiochemical properties for the constructed sequence were analyzed with the help of the ProtParam web tool (http://web.expasy.org/protparam/). Using this tool, the vaccine's amino acid composition, *in vivo*/*in vitro* half-life, theoretical isoelectric point (PI), grand average of hydropathy (GRAVY) instability index, and molecular weight were evaluated for the constructed amino acid sequence (10).

Secondary structure prediction for the final vaccine sequence

PSIPRED tool (http://bioinf.cs.ucl.ac.uk/psipred/) was used for the secondary structure prediction of the constructed vaccine sequence. The server is available online and performs the assigned task with great accuracy (29).

Tertiary structure prediction

Robetta (http://robetta.bakerlab.org/) was used for tertiary structure prediction of the constructed sequence for the vaccine. Robetta uses Continuous Automated Model Evaluation (CAMEO), and every week it calculates up to 20 prereleased PDB targets. Robetta is believed to be the most accurate and consistent server since 2014. The server averages a Local Distance Difference Test score of around 69, a superposition independent score, and estimates interatomic distances with values ranging from 0 considered as bad to 100 as good (29,32). PyMOL was used for visualizing the three-dimensional (3D) structure of the construct vaccine sequence. ◀AU7

Tertiary structure refinement

Galaxy Refine tools (http://galaxy.seoklab.org/) were utilized for refining the final vaccine 3D structure. For the side chain reconstruction and repacking of the amino acid sequence, the server uses CASP10 refinement method. In contrast, the same method is employed for molecular dynamic simulation to relax the 3D structure of the query protein (17).

Tertiary structure validation

ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (43) ERRAT (http://services.mbi.ucla.edu/ERRAT/) and RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) were utilized for validation of the 3D structure of constructed vaccine. ProSA-web predicts a score to the input sequence as the sequence lies outside the range of native proteins, it is more likely that the structures have errors that are shown in the provided 3D structure in the query for more convenient analysis. ERRAT is another tool for 3D structure validation and is used to detect the nonbonded interaction in the structure (7). RAMPAGE utilizes the PROCHECK principle and provides a Ramachandran plot for the query structure (42).

Table 2. List of Final Nontoxic and Nonallergen Cytotoxic T Lymphocyte Epitopes with Their Combined and Immunogenicity Score, Respectively

Protein name	*Epitopes*	*Sequence*	*Combine score*	*Immunogenicity score*
Replication protein E1	CTL	GTGCNGWFY	2.9993	0.23769
Protein E6	CTL	ISEYRHYCY	2.7337	0.07501
Minor capsid protein L2	CTL	PTSINNGLY	3.2988	0.08995
Regulatory protein E2	CTL	QVDYYGLYY	3.4720	0.02026
Major capsid protein L1	CTL	YVARTNIYY	2.3850	0.20665

CTL, cytotoxic T lymphocyte.

AU17 ▶

Table 3. List of Selected Helper T-Lymphocyte Epitopes with Its Respective Percentile Rank Predicted Using IEDB Tool

Proteins	*Allele*	*Position*	*Peptide sequence*	*Method*	*Percentile rank*
Replication protein E1	HLA-DRB5*01:01	450–464	GVEFMSFLTALKRFL	Consensus (smm/nn/sturniolo)	0.02
	HLA-DRB5*01:01	448–462	YQGVEFMSFLTALKR	Consensus (smm/nn/sturniolo)	0.05
Protein E6	HLA-DRB5*01:01	128–142	KKQRFHNIRGRWTGR	Consensus (smm/nn/sturniolo)	7
	HLA-DRB5*01:01	126–140	LDKKQRFHNIRGRWT	Consensus (smm/nn/sturniolo)	9.4
Minor capsid protein L2	HLA-DRB5*01:01	281–295	PDFLDIVALHRPALT	Consensus (smm/nn/sturniolo)	0.4
	HLA-DRB3*02:02	459–473	RKRLPYFFSDVSLAA	NetMHCIIpan	2.1
Regulatory protein E2	HLA-DRB1*03:01	255–269	KLLHRDSVDSAPILT	Consensus (smm/nn/sturniolo)	3.9
	HLA-DRB3*02:02	175–189	AEKYSKNKVWEVHAG	NetMHCIIpan	4.1
Major capsid protein L1	HLA-DRB3*01:01	112–126	ICIEKQSRAAKRRLFQS	Consensus (smm/nn/sturniolo)	0.94

B cell epitope prediction

BCPred (http://ailab.ist.psu.edu/bcpred/) and ElliPro (http://tools.iedb.org/ellipro/) were utilized for the prediction of linear B cell epitopes and conformational B cell epitopes, respectively. BCPred uses the support vector machine method of the kernel method. It uses amino acid pair antigenicity for the prediction of linear B cell epitopes (9). In contrast, Ellipro assigns protrusion index value to the predicted epitopes described based on the center of mass of each residue located outside of the ellipsoid. Residues with large values are believed to have better solvent accessibility (31).

Molecular docking of vaccine with the receptor

HDOCK (http://hdock.phys.hust.edu.cn/) server was used for the docking of vaccine with human TLR-3 accessed through PDB ID: 1ziw to show the improvement of the immune response (8). The server is different from other docking servers for its specific ability to support amino acid sequences as input and a hybrid docking strategy (38,44). The server is robust and is known for the experimental information about the protein/protein binding site and small-angle X-ray drip, which can be incorporated throughout the docking and postdocking processes (45).

In silico *cloning and optimization of vaccine protein*

In the JCat tool (Java codon adaptation tool) (13) was used to express the B cell and T cell epitope-based vaccine in a suitable vector for codon optimization, expression, and reverse translation. The codon optimization was significant for the expression of vaccine structure in a host *Escherichia coli* (strain K12) as the usage of a codon is comparatively different in *E. coli* than the native host. Rho-independent transcription, restriction cleavage sites, and prokaryotic ribosome-binding sites were eluded through considering three extra options. Java codon adaptation tools provide the output in terms of CAI (codon adaptation index) and %GC content to confirm the high level of protein expression. For cloning, the final vaccine in *E. coli* pET-28a (+) vector modification N and C terminal with *Xho*I and *Nde*I restriction sites was performed, respectively. Finally, for expression, the prepared optimized sequence, along with the restriction sites, was installed to the pET-28a (+) vector utilizing the SnapGene tool.

Immune simulation

To understand the dynamics of the human immune system in response to foreign particles, a server that uses agent-based modeling, C-ImmSim, was used to predict the relationships between the human immune system and the foreign particle (33). Production of cytokines as well as other substances like interferon and antibodies were estimated by applying the PSSM method. Moreover, response for T helper ◀AU8 cell 1 and T helper cell 2 (Th1 and Th2) was also predicted with the defaulted parameters measure of diversity or Simps Index by the server.

Vaccine population coverage

To crosscheck the impact of amino acids substitution on vaccine population coverage. BLAST was performed, and the sequences were retrieved for alignment. ClustalW was used to perform sequence alignment. The predicted epitopes were then mapped and checked for any mutations. In our search, we selected global sequences to make sure that the vaccine is universal.

Results

Protein collection

In this study, the most suitable potential candidate proteins of HPV strain 16 (HPV16) were retrieved from UniProt under the UniProt ID: UP000009251 for designing B and T cell multiepitope vaccine using immunoinformatics

Table 4. B Cell Epitopes Predicted by BCPred with Defaulted Parameters

Rank	*Sequence*	*Start position*	*Score*
1	VWEVHAGGPGPGPDTSFYNP	261	1
2	RPALTGPGPGRKRLPYFFSD	203	1
3	VDSAPILTGPGPGAEKYSKN	240	1
4	FHNIRGRWTGPGPGPDFLDI	179	1
5	TALKRGPGPGKKQRFHNIRG	143	1
6	YVARTNIYYGPGPGGVEFMS	99	1
7	TALKRFLGPGPGYQGVEFMS	121	1
8	TQRLVWGPGPGSTILEDWNF	282	0.999
9	GRKCCRRKKEAAAKGTGCNG	37	0.983
10	YISEYRHYCYAAYPTSINNG	62	0.929

Table 5. B-Epitope Residues Predicted through Ellipro

No	*Residues*	*No. of residues*	*Score*
1	*A*:K39, A:C40, A:C41, A:R42, A:R43, A:K44, A:K45, A:E46, A:A47, A:A48, A:A49, A:K50, A:G51, A:T52, A:G53, A:C54, A:N55, A:G56, A:W57, A:F58	20	0.788
2	*A*:V264, A:H265, A:A266, A:G267, A:G268, A:P269, A:G270, A:P273, A:D274, A:T275, A:S276, A:F277, A:Y278, A:N279, A:P280, A:D281, A:T282, A:Q283, A:R284, A:L285, A:V286, A:W287, A:G288, A:P289, A:G290, A:P291, A:G292, A:S293, A:T294, A:I295, A:L296, A:E297, A:D298, A:W299, A:N300, A:F301, A:G302, A:L303, A:Q304, A:P305, A:P306, A:P307	42	0.776
3	*A*:Y71, A:A72, A:Y74, A:P75, A:T76, A:S77, A:I78, A:N79, A:N80, A:G81, A:A84, A:A85, A:Y86, A:Q87, A:V88, A:I182, A:R183, A:G184, A:R185, A:W186, A:T187, A:G188, A:P189, A:G190, A:P191, A:G192, A:P193, A:D194, A:F195, A:L196, A:D197, A:I198, A:V199, A:A200, A:L201, A:H202, A:R203, A:P204, A:A205, A:L206, A:T207, A:G208, A:P209, A:G210, A:P211, A:G212, A:R213, A:K214, A:R215, A:L216, A:P217, A:Y218, A:V240, A:D241, A:S242, A:A243, A:P244	57	0.67
4	*A*:G1, A:I2, A:L21, A:S22, A:C23, A:L24, A:P25, A:K26, A:E27, A:E28, A:Q29, A:I30, A:G31, A:K32, A:C33, A:S34, A:T35, A:R36, A:G37, A:R38, A:G130, A:P131, A:G132, A:A226, A:A227, A:G228, A:P229, A:G230, A:P231, A:G232	30	0.627
5	*A*:K255, A:S257, A:N259, A:K260, A:V261, A:W262, A:E263	7	0.584
6	*A*:Y90, A:Y91, A:K124, A:R125, A:F126, A:L127, A:G128, A:P129, A:Q134, A:G135, A:V136, A:M139, A:R237	13	0.542
7	*A*:R14, A:G15, A:G16, A:R17, A:C18, A:A19, A:K146	7	0.527

approach. HPV16 proteome comprises a total of nine proteins out of which five proteins, Replication protein E1, Protein E6, Minor capsid protein L2, Regulatory protein E2, and Major capsid protein L1, were used after antigenicity analysis as shown in Table 1. T1

Prediction of CTL epitopes

For all the 5 proteins, a total of 73 CTL epitopes were predicted. Out of which 18 CTL epitopes for Replication protein E1, 4 for Protein E6, 19 for Minor capsid protein L2, 13 for Regulatory protein E2, and 19 CTL epitopes for Major capsid protein L1 were predicted. However, based on toxicity, allergenicity, immunogenicity, and MHCI binding affinity, only three epitopes were selected (Table 2). T2

Prediction of HTL epitopes

HTL epitopes were finalized the same way as CTL epitopes, as shown in Table 3, and used to construct the multiepitope vaccine. For HTL epitope prediction, seven reference set of Human alleles was used. The final HTL epitopes for the five proteins positioned as follows:450–464 (HLA-DRB5*01:01), 448–462 (HLA-DRB5*01:01), 128–142 (HLA-DRB5*01:01), 126–140 (HLA-DRB5*01:01), 281–295 (HLA-DRB5*01:01), 459–473 (HLA-DRB3*02:02), 255–269 (HLA-DRB1*03:01) 175–189 (HLA-DRB3*02:02) 112–126 (HLA-DRB3*01:01), and 422–436 (HLA-DRB3*01:01). T3

B cell epitope prediction

A total of 10 linear B cell 20mer epitopes were predicted for the constructed vaccine using the BCpred server with the defaulted parameters (Table 4). Conformational B cell epitopes with $a > 0.5$ score were considered discontinuous (Table 5). The 3D model of the vaccine showing linear and discontinuous B cell epitopes is shown in Figure 2A. The ElliPro score represents the average value of PI for the epitope residues. The acquired probability score also approves and confirms the T4 T5 F2

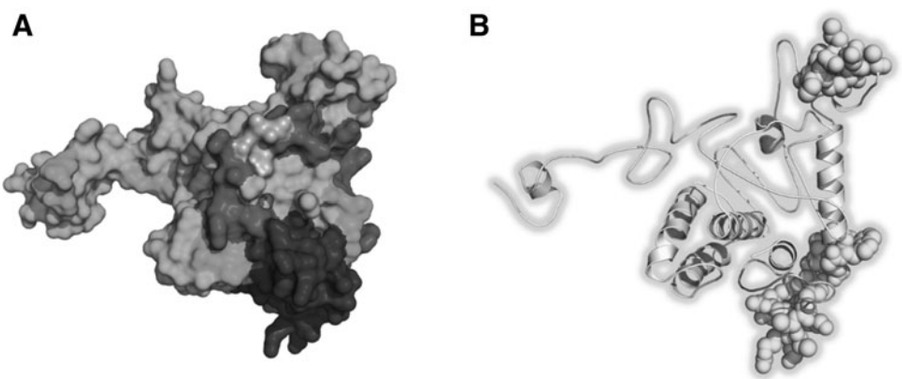

FIG. 2. **(A)** Vaccine 3D model showing linear, conformational B cell, CTL, and HTL epitopes. **(B)** ElliPro-predicted epitopes (*yellow* ball) displayed on the 3D vaccine structure (*grey* cartoon). 3D, three-dimensional; CTL, cytotoxic T lymphocyte; HTL, helper T-lymphocyte.

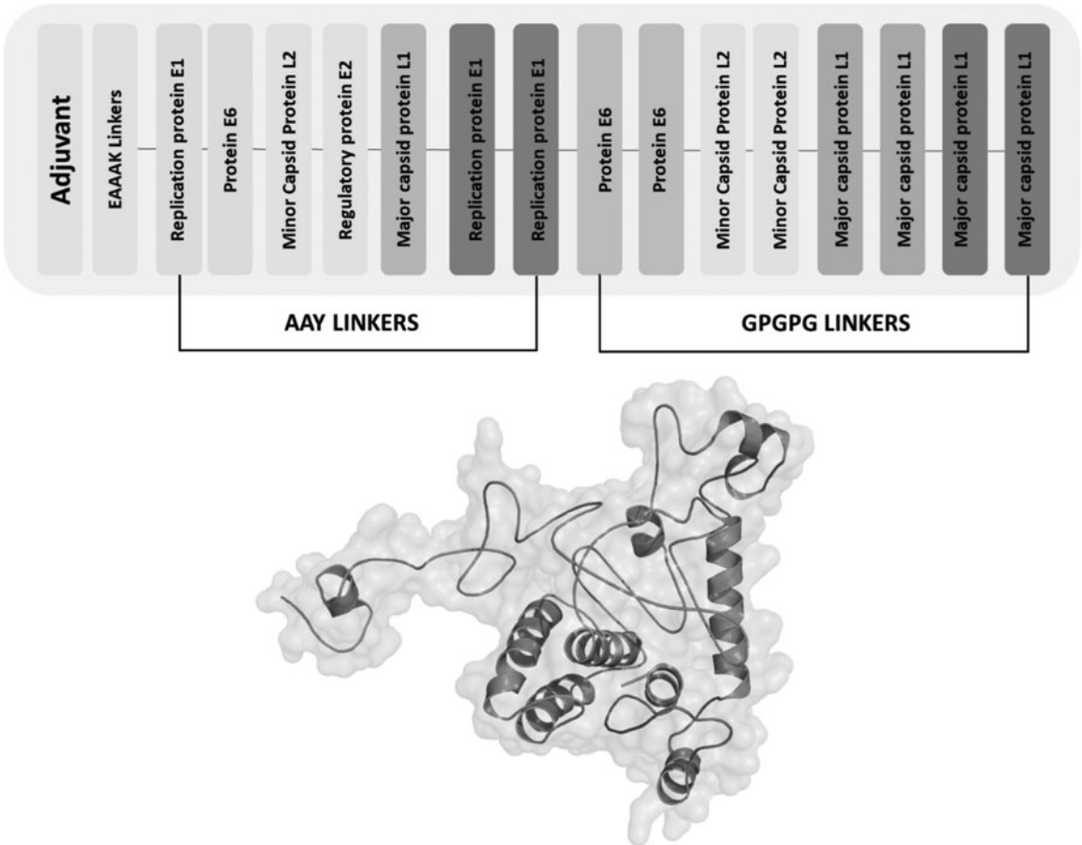

FIG. 3. Schematic presentation of the complete vaccine unit comprising a multiepitope peptide and linked adjuvant. Epitopes of a single protein are shown by the same color, whereas CTL and HTL epitopes are linked by separate linkers. The vaccine 3D structure (*brown cartoon ribbon*) given is also illustrated.

immunological behavior of the designed multiepitope subunit vaccine. ElliPro-predicted epitopes as shown on the vaccine 3D model is given as Figure 2B.

Designing a final vaccine construct

Five CTL and 10 HTL epitopes were used to construct a subunit vaccine sequence linked with a ratio of 1:2. Adjuvant was linked to the first CTL epitope N-terminal with EAAAK linker, and the rest of the CTL epitopes were joined using AAY linkers. The last CTL epitopes were fused with the HTL epitope using a GPGPG linker. At last, all the HTL epitopes were fused using GPGPG linkers. Schematic presentation of the complete vaccine unit comprising a multiepitope peptide and linked adjuvant is shown in Figure 3. F3 ▶

Three-dimensional structure prediction, validation, and refinement

The 3D structure for the constructed sequence was developed using the Robetta server. The server-initiated five models, out of which the best was selected was based on poststructure evaluation parameters. The five refined models provided by Galaxy Refinement server, the best model (model 1) was selected based on various factors like calculated RMSD, Mol Probity, High GDT-HA score Poor rotamers, and Rama favored. Out of five refined models provided by the server, model 1 was selected as a refined structure and used further in docking. Model 1 3D structure is provided in Figure 3. It was decided based on numerous factors, such as a high GDT-HA score of 0.9829, RMSD calculated was 0.317. The Mol Probity was measured by 1.571. The clash score measured atoms' overlapping, which was low for model 1 and was calculated as 11.0.

Similarly, Poor rotamers were calculated as 0.0, and the Rama favored was calculated as 99.3, which were far better than other models, as shown in Table 6. Therefore, the refined model 1 was selected for further analysis. The selected 3D model of the vaccine was then subjected to a structure assessment phase. ProCheck for protein model showed that 95.2% of amino acids are plotted in the most favorable areas, 4.8% in allowed regions, and only 0.0% in disallowed regions. Similarly, ERRAT and ProSA-web calculated and verified the overall quality of the basic 3D model. The resulting overall quality factor reported by ERRAT was 79.7%, and the ProSA-web showed a score of −4.58 for the 3D model of vaccine construct, which proved the accuracy of the model. Results from ProSA-web and rampage are given in Figure 4. ◀T6 ◀F4

Antigenicity and allergenicity of vaccine

Antigenicity and Allergenicity of the vaccine sequence were predicted using the VexiJen server and AllerTop

Table 6. Galaxy Refined Results for Vaccine Construct

Model	*GDT-HA*	*RMSD*	*MolProbity*	*Clash score*	*Poor rotamers*	*Rama favored*
MODEL 1	0.9829	0.317	1.571	11.4	0.0	99.3
MODEL 2	0.9805	0.318	1.585	11.8	0.0	99.0
MODEL 3	0.9805	0.332	1.624	13.0	0.0	99.0
MODEL 4	0.9780	0.318	1.592	12.0	0.0	99.0
MODEL 5	0.9796	0.322	1.534	10.3	0.0	99.0

The model was selected for its low MolProbity and high Rama-favored score.

server. VexiJen classified the sequence as Antigenic with a score of 0.4871 and AllerTop as nonallergen with the nearest protein in UniProtKB accession number Q86UU0 defined itself as nonallergen.

Physiochemical properties of vaccines

ProtParam predicted physicochemical properties of the proposed vaccine sequence as molecular weight 33.72 kDa, theoretical PI 9.58, which classify it as basic in nature. *In vitro* half-life in mammalian reticulocytes was estimated as 30 h: *in vivo* in yeast as >20 h and *in vitro* in *E. coli* as >10 h. The predicted instability index was 39.50 2, which represents the stability of the vaccine. The aliphatic index of the vaccine sequence was computed to be 59.80, indicating thermostability, and a negative GRAVY –0.521 represents the hydrophilic nature of the constructed sequence.

Prediction of secondary structure

The secondary structure of the vaccine construct can be categorized as; alpha-helix (Hh) 34.4%, extended strand (Ee) 21.06%, and random coil (Cc) 55%. Graphically, secondary structure elements are provided in Figure 5.

Molecular docking of constructed vaccine with human toll-like receptor 3

From a previous literature review showing the involvement of Human TLR3 in HPV-related diseases and infections, the docking of the vaccine was performed with TLR3 (PDB ID: 1ziw). Docking of both the molecules was performed through ClusPro 2.0. From the 30 models created, only 1 model on the basis of low energy and large size was selected as it showed better interaction between the receptor and ligand. Docked complex intermolecular conformation and chemical interactions are presented in Figure 6.

Codon adaptation and in silico *cloning*

The amino acid sequence of the vaccine was uploaded to JCAT to adapt vaccine sequence codon usage according to *E. coli* K12 strain. The calculated CAI was 0.96, with the GC contents 56.24%. Such results indicated a better

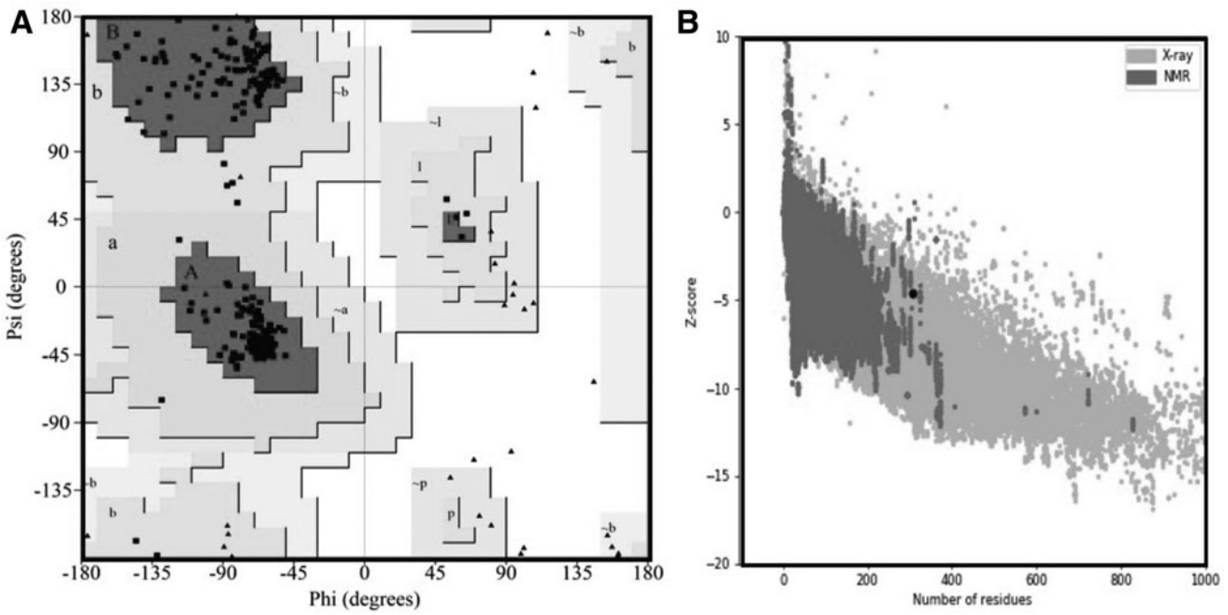

FIG. 4. Quality assessment of the vaccine model using rampage Ramachandran plot **(A)** and ProSA-web **(B)**. In the Ramachandran plot, the *black square* indicates torsion angles distribution relative to the core (*red*) and allowed (*yellow*) regions. Residues are plotted in generously allowed (*pale yellow*), and disallowed (*white*) regions. The ProSA plot score of the vaccine can be spotted as a *black dot* in the plot of all PDB structures quality scores of the same size. PDB.

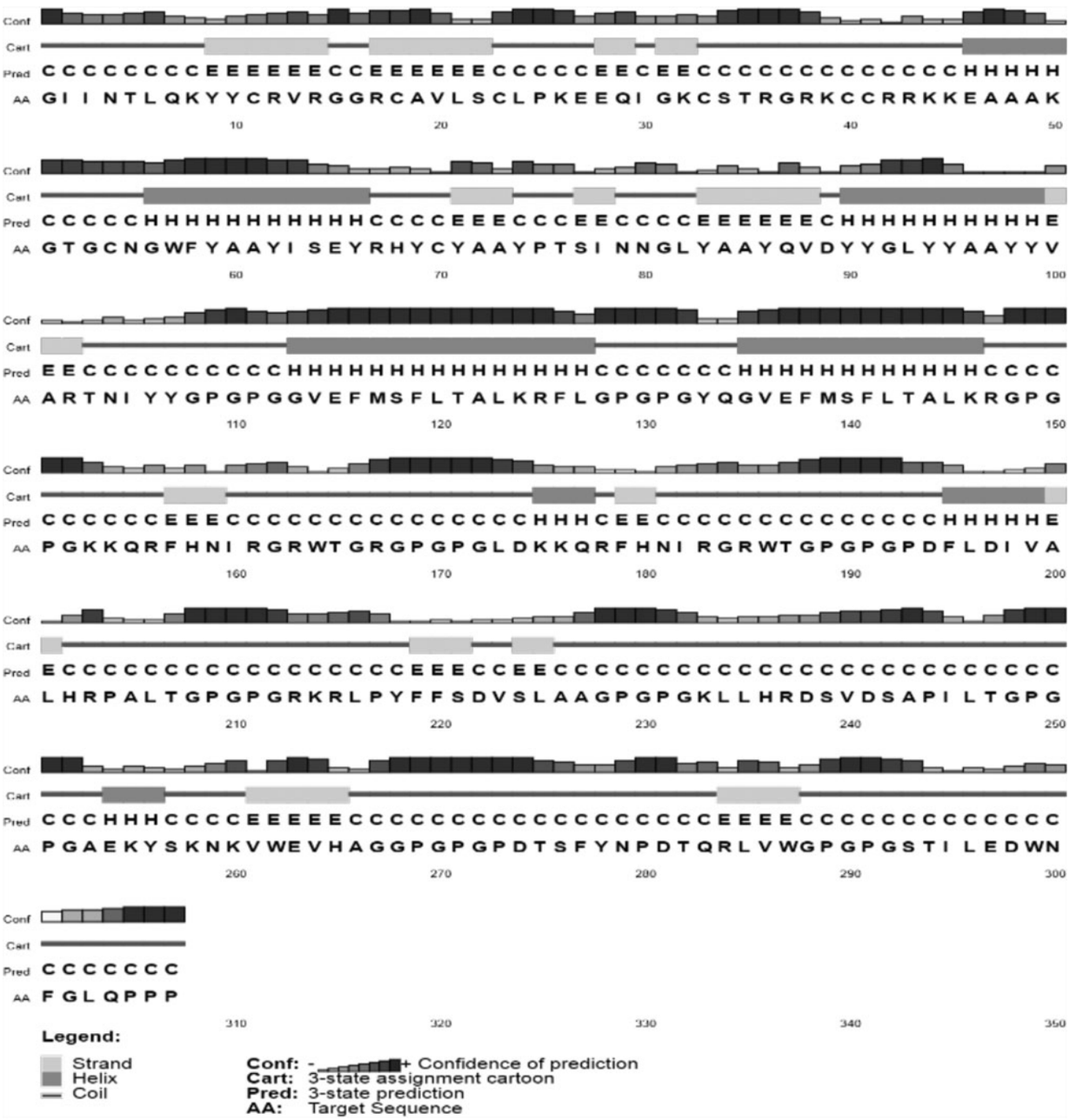

FIG. 5. Graphical representation of the secondary structure of the vaccine construct.

expression level of vaccine construct in *E. coli* (K12 strain). GC contents between 35% and 70% have been reported for better expression of the gene. Finally, the optimized vaccine was cloned into pET28a (+) plasmid using the restriction cloning tool of SnapGene. For this purpose, suitable restriction enzymes *Xh0*I and *Nde*I were selected, and their restriction sites were inserted at both ends of the adapted nucleotide sequence. Restriction sites for *Xh0* and *Nde*I were added at the N-terminal and C-terminal of the optimized vaccine construct and cloned in pET-28a (+) vector (Fig. 7). The total size of the cloned vector is 1,137 bps.

Immune simulation of the proposed vaccine and immune response triggering

The human immune system response was monitored through *in silico* immune simulation upon the antigen's injection in several doses. The immune response was significantly triggered, and the antibody titer was very high after the injection (Fig. 8A). It can be clearly observed that the combined IgM+IgG titer remained 700,000xx/mL. The IgM alone titer was reported to be around 55000xx/mL, whereas Ig1 was reported as 15,000xx/mL. We also checked the interleukin (IL) and cytokine responses (Fig. 8B). The IFN-g

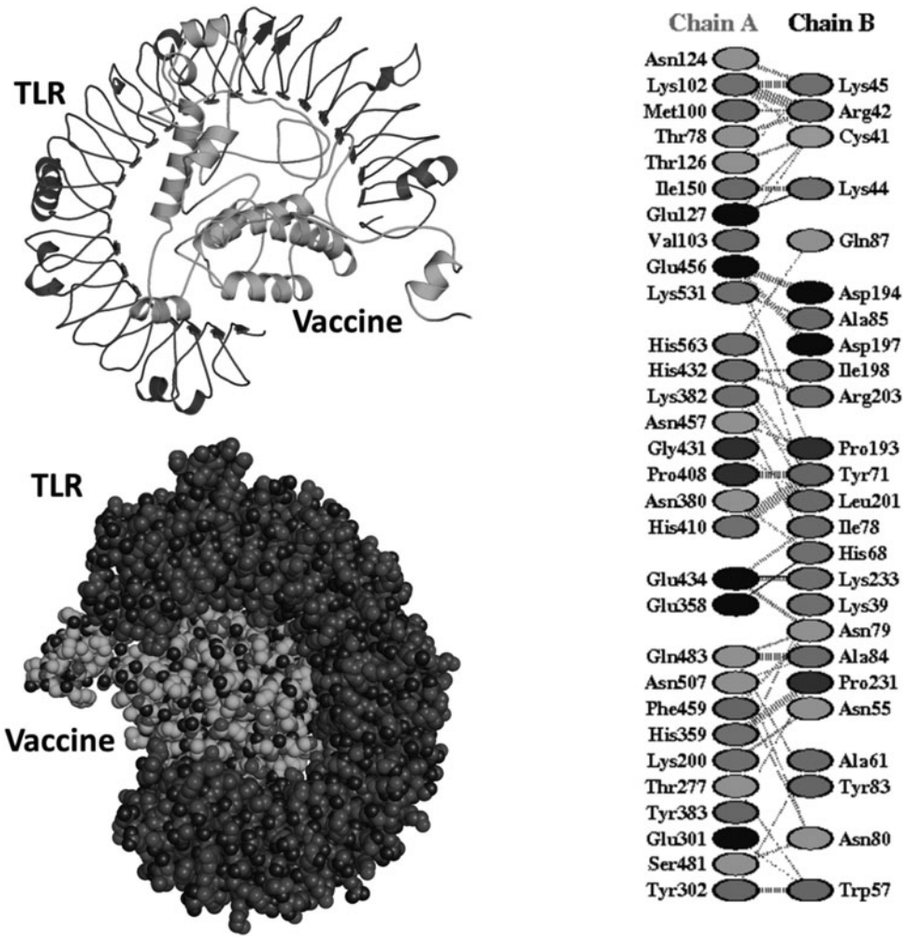

FIG. 6. Top-ranked model of docking assay. The different types of chemical interactions involving residues from TLR3 receptor and vaccine are also provided.

and IL-2 titers were significantly high. This shows the consistent and robust immune triggering response upon the injection. The cellular immune system response to pathogen identification at re-encounter was also robust, including memory cells' development. The maximum concentration of 375 cells/mm^3 for the phagocytic natural killer cell population was reported (Fig. 8C). The population of the T cell was reported to be >1,150 cells/mm^3 (Fig. 8D). Dendritic cells and phagocytic macrophage populations were 200 cell/mm^3 each (Fig. 8E, F). Hence these results confirm that our vaccine candidate effectively triggers the immune response.

Epitopes' conservation analysis

AU11 ► Using BLAST ~200 different strains from different countries, including United States, India, South Africa, Thailand, United Kingdom, and other countries were analyzed. To check the impact of any mutations reported in different strains, ClustalW-based alignment was performed for each protein. From the alignment it can be seen that all the epitopes in different proteins lie in conserved regions and would vaccine worldwide immunity. All the alignments are shown in Supplementary Figures S1–S5. These analyses suggest SF1 – SF5 ► that the predicted vaccine has wide population coverage. The sequential tutorial of this methodology is given in the Supplementary File.

Discussion

HPV-related cancers have emerged to be endemic worldwide. They play various types of cancers like head and neck squamous cell carcinomas (HNSCC) penile, anal, and cervical head and neck squamous cell carcinomas. HNSCC have increased the number of incidences in comparison to cervical cancers and is predicted to continue until 2060 (16).

Vaccination plays a substantial role in the effective control of infectious diseases. Epitope-based vaccine designing is beneficial, profitable, highly stable, nontoxic, and easy in engineering. Hence, it has the upper hand over conventional vaccines, which may have many apprehensions in immunocompromised individuals (20). These multiepitope-based vaccine peptides are retrieved from the pathogen's proteome and are designed from highly specific and immunogenic B cell and T cell epitopes (2,19,20). Immunization is one of the most effective methods to control infectious diseases. Immunoinformatics approaches are speedy and efficient tools

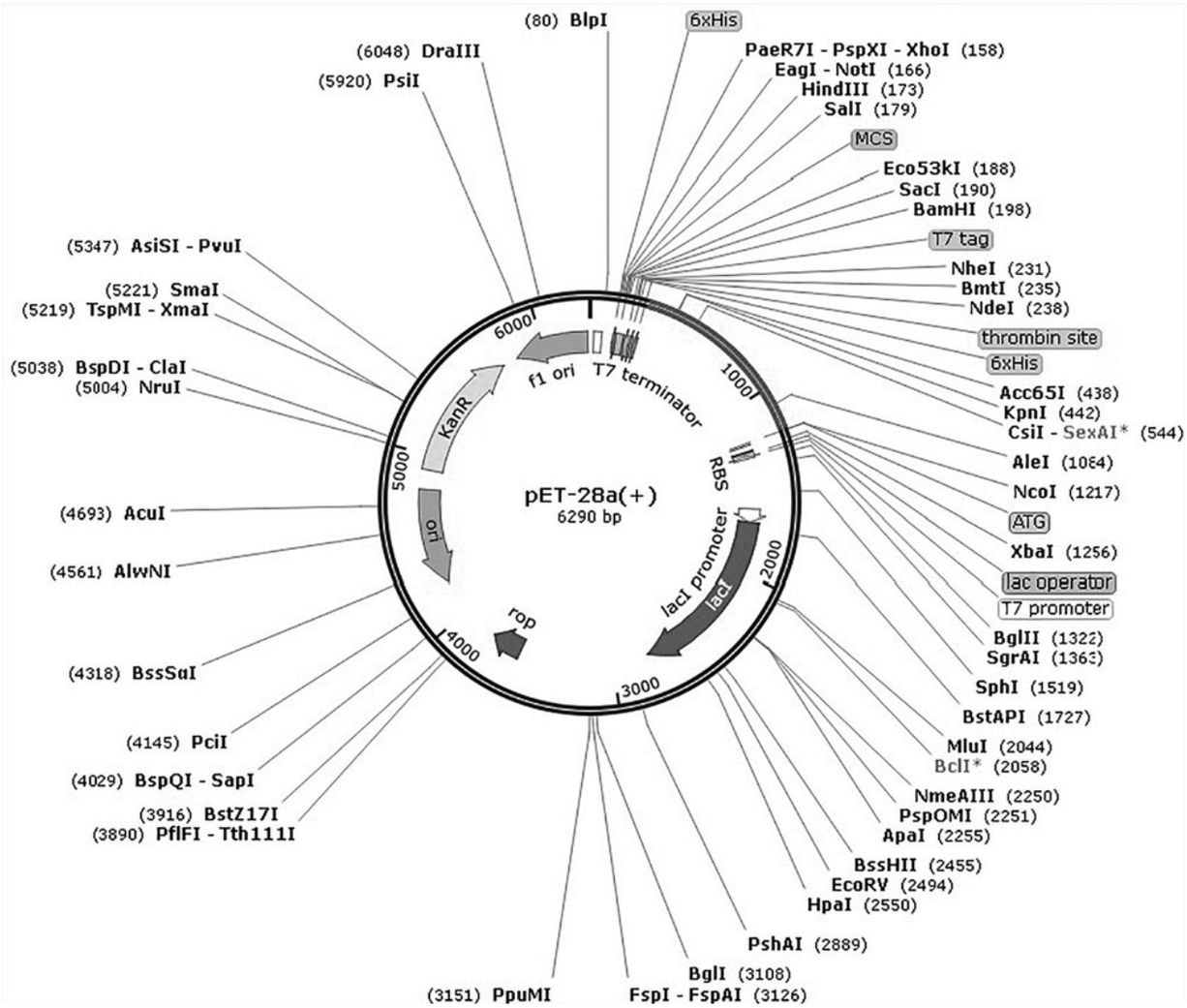

FIG. 7. Cloned vaccine sequence (*red*) into peT-28a vector.

for the identification of potential candidates for vaccine designing from a pathogen's proteome (2,18–26).

The proteome of HPV strain 16 consists of 9 proteins. This study used L1, L2, E1, E2, and E6 proteins to construct the multiepitope vaccine. Prediction of the best CTL, HTL, as well as B cell epitopes was carried out utilizing different bioinformatics tools. VaxiJen server predicted the antigenicity of vaccine construct to be 0.4871 while its allergenicity was predicted using AllerTop server classifying it as nonallergen with nearest protein UniProt ID Q86UU0 defined as nonallergen. The constructed vaccine's molecular weight was calculated as 33.72 kDa with a theoretical PI of 9.58, showing the basic nature of the vaccine construct. The instability index value was computed to be 39.50 classifying the developed vaccine as stable. The aliphatic index of the subunit vaccine is 59.8, and GRAVY is –0.5, which represents its hydrophilic nature.

Moreover, homology modeling approach was used to generate the tertiary structure of the developed subunit vaccine sequence. Meanwhile, the validation and accuracy check of the tertiary structure was performed with tools like ProSAWeb, ERRAT, and Procheck. Molecular docking of the vaccine and TLR-3 also revealed its strong interaction.

Codon optimization was performed for better expression and efficient transcription and translation of foreign protein in the *E. coli* K-12 strain. The efficiency of codon optimization was determined by calculating the total GC content and DNA sequence's CAI value. Finally, the vaccine was cloned into pET-28a (+) vector, where *Xho*I and *Nde*I restriction sites were inserted at the N- and C-terminals of the adapted codon, respectively. Later on, the optimized codons with restriction sites were introduced into the vector by restriction cloning. To close, a cloned construct was generated, having a sequence a total length of 1,137 bps. Furthermore, immune simulation also provoked the immune response.

Conclusions

The role of HPVs in cancers and other diseases is pronounced. An ultimate protocol is required for coping with the situation. In this study, we manipulated the proteome of HPV16 using a different easily available free online server

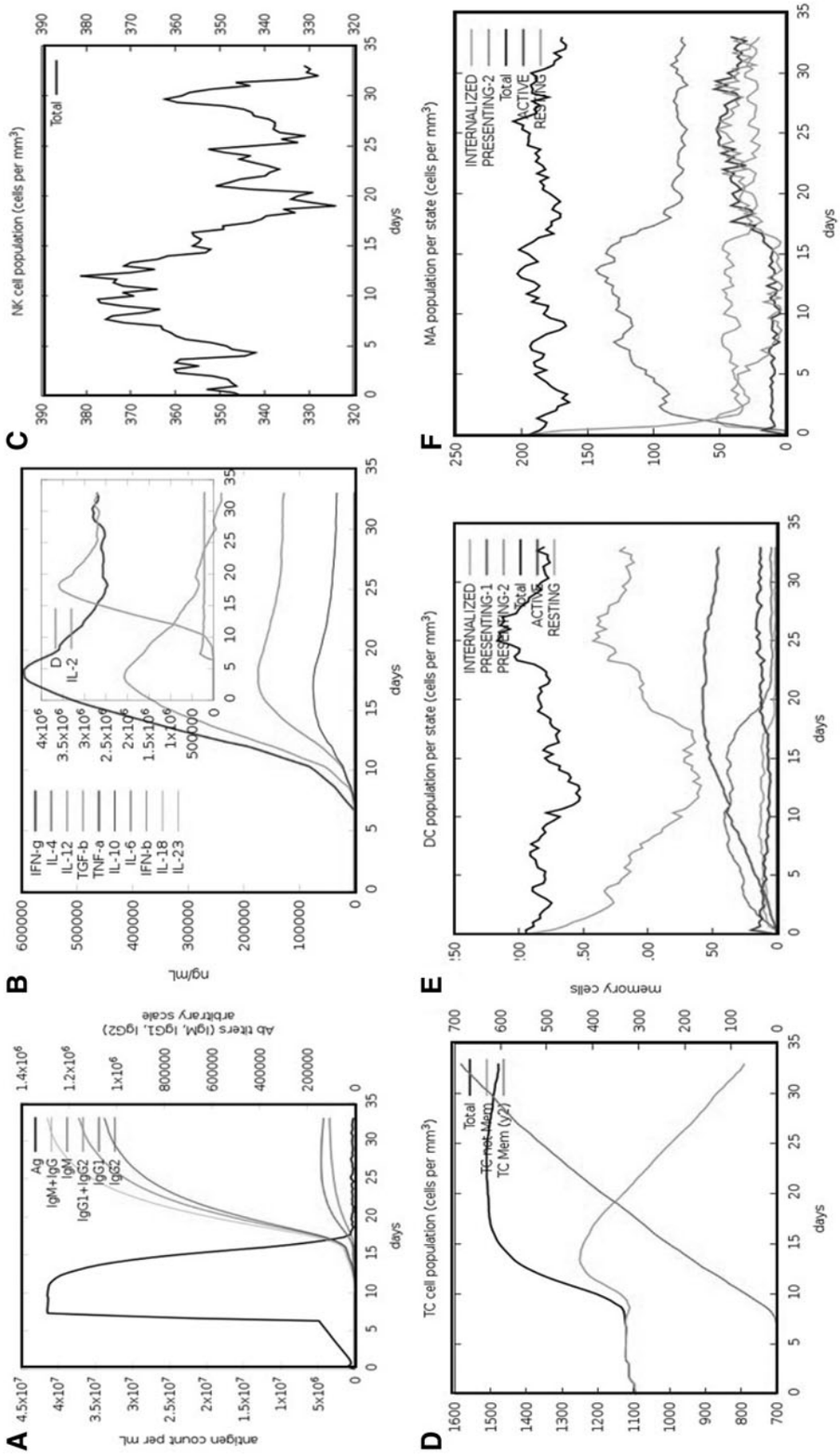

FIG. 8. *In silico* immune simulation by the vaccine followed by injection. The immune response to the vaccine is demonstrated by antibodies titer (**A**), interferon and interleukins (**B**), natural killer cells (**C**), cytotoxic T cell (**D**), dendritic cells (**E**), and macrophages (**F**).

for the prediction of immunogenic CTL, HTL, and B cell epitopes. A vaccine sequence was constructed, modeled, and evaluated for physicochemical properties. Additionally, immune simulation was performed for an insight into the human immune response to the constructed vaccine. Also, the interaction of the proposed construct vaccine with Human-TLR3 was performed, and at last, the sequence was cloned in an appropriate vector for expression. The developed vaccine construct might be able to prevent or lower HPV-related cases.

Authors' Contributions

All the authors were involved in design, analysis, writing and finalizing the article.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

D.-Q.W. is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant Nos. 32070662, 61832019, 32030063), the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), the Natural Science Foundation of Henan Province (162300410060), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Pengcheng Laboratory. and the Center for High-Performance Computing, Shanghai Jiao Tong University.

Supplementary Material

Supplementary File
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5

References

AU12 ▶
1. Accardi R, and Gheit T. Cutaneous HPV and skin cancer. La Presse Médicale 43:e435–e443. (3CLpro). J Biomol Struct Dynam 2014;1–12.
2. Ali A, Khan A, Kaushik AC, *et al.* Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV). Sci Rep 2019; 9:1–12.
3. Benson DA, Karsch-Mizrachi I, Lipman DJ, *et al.* GenBank. Nucleic Acids Res 2009;37:D26–D31.
4. Bouvard V, Baan R, Straif K, *et al.* A review of human carcinogens—Part B: biological agents. Lancet Oncol 2009; 10:321.
5. Brancaccio RN, Robitaille A, Dutta S, *et al.* Generation of a novel next-generation sequencing-based method for the isolation of new human papillomavirus types. Virology 2018; 520:1–10.
6. Bravo IG, and Félez-Sánchez M. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. Evol Med Public Health 2015;2015:32–51.
7. Colovos C, and Yeates T. ERRAT: an empirical atom-based method for validating protein structures. Protein Sci 1993;2:1511–1519.
8. Daud II, Scott ME, Ma Y, *et al.* Association between toll-like receptor expression and human papillomavirus type 16 persistence. Int J Cancer 2011;128:879–886.
9. EL-Manzalawy Y, Dobbs D, and Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recogn 2008;21:243–255.
10. Gasteiger E, Hoogland C, Gattiker A, *et al.* Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook. Springer 2005;571–607. ◀AU13
11. Gheit T. Mucosal and cutaneous human papillomavirus infections and cancer biology. Front Oncol 2019;9:355.
12. Gottschling M, Göker M, Stamatakis A, *et al.* Quantifying the phylodynamic forces driving papillomavirus evolution. Mol Biol Evol 2011;28:2101–2113.
13. Grote A, Hiller K, Scheer M, *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:W526–W531.
14. Gupta S, Kapoor P, Chaudhary K, *et al.* In silico approach for predicting toxicity of peptides and proteins. PLoS One 2013;8:e73957.
15. Halec G, Schmitt M, Dondog B, *et al.* Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. Int J Cancer 2013;132:63–71.
16. Hathaway JK. HPV: diagnosis, prevention, and treatment. Clin Obstet Gynecol 2012;55:671–680.
17. Heo L, Park H, and Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res 2013;41:W384–W388.
18. Hussain I, Pervaiz N, Khan A, *et al.* Evolutionary and structural analysis of SARS-CoV-2 specific evasion of host immunity. Genes Immun 2020;21:409–419.
19. Khan A, Ali SS, Khan MT, *et al.* Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease. J Biomol Struct Dyn 2020:1–12. [Epub ahead of print]; DOI: 10.1080/07391102.2020.1779128.
20. Khan A, Junaid M, Kaushik AC, *et al.* Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. PLoS One 2018;13:e0196484.
21. Khan A, Khan M, Saleem S, *et al.* Decoding the structure of RNA-dependent RNA-polymerase (RdRp), understanding the ancestral relationship and dispersion pattern of 2019 Wuhan Coronavirus. 2020. ◀AU14
22. Khan A, Khan M, Saleem S, *et al.* Phylogenetic analysis and structural perspectives of RNA-dependent RNA-polymerase inhibition from SARs-CoV-2 with natural products. Interdiscip Sci 2020;12:335–348.
23. Khan A, Khan MT, Saleem S, *et al.* Structural Insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. Comput Struct Biotechnol J 2020;18:2174–2184.
24. Khan A, Junaid M, Li C-D, *et al.* Dynamics insights into the gain of flexibility by Helix-12 in ESR1 as a mechanism of resistance to drugs in breast cancer cell lines. Front Mol Biosci 2020;6:159.
25. Khan MT, Khan A, Rehman AU, *et al.* Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. Sci Rep 2019;9:1–12.

26. Khan A, Rehman Z, Hashmi HF, *et al.* An integrated systems biology and network-based approaches to identify novel biomarkers in breast cancer cell lines using gene expression data. Interdiscip Sci 2020;12:155–168.
AU15 ► 27. Khan S, Khan A, Rehman AU, *et al.* Immunoinformatics and structural vaccinology driven prediction of multi-epitope vaccine against Mayaro virus and validation through in-silico expression. Infect Genet Evol 2019;73:390–400.
28. Leung TW, Liu SS, Leung RC, *et al.* HPV 16 E2 binding sites 1 and 2 become more methylated than E2 binding site 4 during cervical carcinogenesis. J Med Virol 2015;87: 1022–1033.
29. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–405.
30. Mühr LSA, Eklund C, and Dillner J. Towards quality and order in human papillomavirus research. Virology 2018; 519:74–76.
31. Ponomarenko J, Bui H-H, Li W, *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9:514.Virology 2008;519:74–76.
32. Raman S, Vernon R, Thompson J, *et al.* Prediction report. Proteins 2009;77:89–99.
33. Rapin N, Lund O, and Castiglione F. Immune system simulation online. Bioinformatics 2011;27:2013–2014.
34. Ravichandran L, Venkatesan A, and Febin Prabhu Dass J. Epitope-based immunoinformatics approach on RNA-dependent RNA polymerase (RdRp) protein complex of Nipah virus (NiV). J Cell Biochem 2019;120:7082–7095.
35. Saha S, and Raghava G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 2006;34:W202–W209.
36. Song Y, DiMaio F, Wang RY-R, *et al.* High-resolution comparative modeling with RosettaCM. Structure 2013;21: 1735–1742.
37. Stranzl T, Larsen MV, Lundegaard C, and Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 2010;62:357–368.
38. Sussman JL, Lin D, Jiang J, *et al.* Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallogr D 1998;54:1078–1084.
39. Van Doorslaer K, and McBride AA. Molecular archeological evidence in support of the repeated loss of a papillomavirus gene. Sci Rep 2016;6:33028.
40. Vita R, Overton JA, Greenbaum JA, *et al.* The immune epitope database (IEDB) 3.0. Nucleic Acids Res 2015;43: D405–D412.
41. Wang R, Pan W, Jin L, *et al.* Human papillomavirus vaccine against cervical cancer: opportunity and challenge. Cancer Lett 2020;471:88–102.
42. Wang W, Xia M, Chen J, *et al.* Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in Gossypium raimondii and G. arboreum. Data Brief 2016; 9:345–348.
43. Wiederstein M, and Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–W410.
44. Yan Y, Tao H, He J, and Huang S-Y. The HDOCK server for integrated protein–protein docking. Nat Protoc 2020;15: 1829–1852.
45. Yan Y, Zhang D, Zhou P, *et al.* HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res 2017;45:W365–W373.
46. Zaharieva N, Dimitrov I, Flower D, and Doytchinova I. Immunogenicity prediction by VaxiJen: a ten year overview. J Proteom Bioinform 2017;10:298–310.

Address correspondence to:
Abbas Khan ◄ AU16
Department of Bioinformatics and Biological Statistics
School of Life Sciences and Biotechnology
Shanghai Jiao Tong University
Shanghai 200240
P.R. China

E-mail: abbaskhan@sjtu.edu.cn

Dr. Dong-Qing Wei
Department of Bioinformatics and Biological Statistics
School of Life Sciences and Biotechnology
Shanghai Jiao Tong University
Shanghai 200240
P.R. China

E-mail: dqwei@sjtu.edu.cn