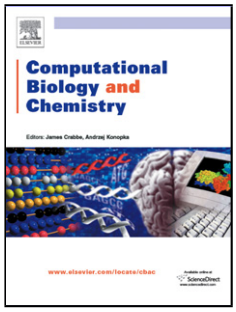


Designing a potent L1 protein-based HPV peptide vaccine; a bioinformatics approach

Zahra Yazdani, Alireza Rafiei, Reza Valadan, Hossein Ashrafi, MarziehSharifi Pasandi, Mostafa Kardan

PII: S1476-9271(19)30275-0

DOI: https://doi.org/10.1016/j.compbiolchem.2020.107209

Reference: CBAC 107209

To appear in: *Computational Biology and Chemistry*

Received Date: 1 April 2019

Revised Date: 8 January 2020

Accepted Date: 16 January 2020

Please cite this article as: Yazdani Z, Rafiei A, Valadan R, Ashrafi H, Pasandi M, Kardan M, Designing a potent L1 protein-based HPV peptide vaccine; a bioinformatics approach, *Computational Biology and Chemistry* (2020), doi: https://doi.org/10.1016/j.compbiolchem.2020.107209

Designing a potent L1 protein-based HPV peptide vaccine; a bioinformatics approach

Zahra Yazdani[1], AlirezaRafiei[1], Reza Valadan[2], Hossein Ashrafi[3], MarziehSharifi Pasandi[2],Mostafa Kardan[1]

1. Department of Immunology, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

2. Molecular and Cell Biology Research Center, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

3. School of Life Science, Pharmacy & Chemistry, SEC Faculty, Cancer Theme, Kingston University London, Kingston upon Thames, London KT1 2EE, UK

Corresponding author: Alireza Rafiei

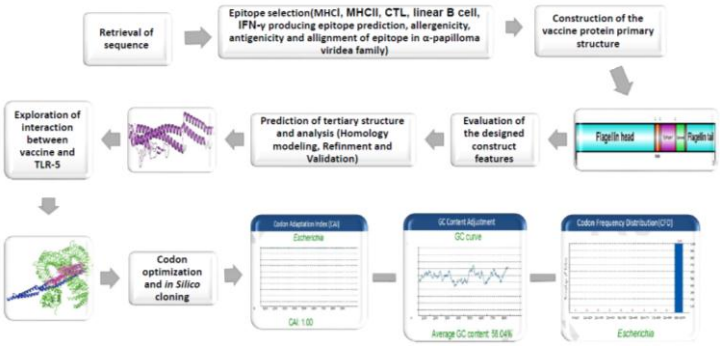
Department of Immunology, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran, 18 Km Khazar Blvd, Khazar Sq. Sari, Iran

Phone: +98-11-33543088

Fax: +98-11-3354-3087

Email: rafiei1710@gmail.com

Graphical abstract



Number of figures: 5

Number of tables: 4

Highlights

- **The vaccine consists of four major sections: toll-like receptor 4 adjuvant, toll-like receptor 5 adjuvant and two epitopes.**
- **Each section was joined together by appropriate linkers.**
- **Different strategies were applied to enhance immunogenicity of peptide vaccine.**
- **The vaccine is able to protection of population on all members of α-papillomaviridea family.**
- **The vaccine has a high quality structure, appropriate physicochemical properties and a high potential to be expressed in E. coli as host.**

Abstract

Background: Oncogenic human papilloma viruses (HPV) are the cause of various types of cancer, specifically cervical cancer. L1 protein is the main protein of HPV capsid which targeted in many vaccine-producing attempts. However, they have not enough

coverage on the various high risk HPV types. Therefore, having a low cost potent HPV vaccine to protect against all members of the α-papillomaviridea family will be promising. In this study, L1 protein-based peptide vaccine was designed using immunoinformatics methods which provides physicochemical properties such as stability in room temperature, potential of antigenicity, non-allergic properties and no requirement with eukaryotic host system.

Results: The designed vaccine has two HPV conserved epitopes with lengths 18 and 27 amino acids in all members of α-papillomaviridea. These peptides promote humoral and cellular immunity and INF-γ responses. In order to ensure strong induction of immune responses, Flagellin, a Toll like receptor 5(TLR-5) agonist, and a short synthetic toll like receptor 4 (TLR-4) agonist were also joined to the epitopes. Structure of the designed-vaccine was validated using Rampage and ERRAT and a high quality 3D structure of the vaccine protein was provided. Docking studies demonstrated an appropriate and stable interaction between the vaccine and TLR-5.

Conclusions: The vaccine is expected to have a high quality structure and suitable properties including high stability, solubility and a high potential to be expressed in *E.coli*. High potentiality of the vaccine in inducing humoral and cellular immune responses, may be considered as an anti-tumor vaccine.

Keywords: Conserved epitopes; Vaccine, HPV, Adjuvant; Immunoinformatics;Tertiary structure analysis; in silico cloning

1. Introduction

Cervical cancer (CC) is the fourth common cancer and the fourth leading cause of cancer death in women worldwide (Bray et al., 2018). Human papillomaviruses (HPV) are the main causes of CC in whichmore than 99% of CC tissues were infectedwith the DNA of the HPV virus(Monie, Hung, Roden, & Wu, 2008; Yuan et al., 2001). HPV also has a key role in other cancers such as vulva, anus, penis, vagina, mouth and throat cancer(Tommasino, 2014). HPV is a member of Papillomaviridea which infect squamous epithelium in different regions. The cellular structure of this virus consists of circular double-stranded DNA with approximately 8,000 bp that contains early regions that encodeearly viral proteins including (E6, E7, E8, E1, E2, E4, E5), late regions which encode capsid component proteins (L1 and L2), and anon coding region as long control region (LCR) which has a critical role in replication and transcription(Clifford, Smith, Plummer, Munoz, & Franceschi, 2003).L1 is the major protein in the capsid structure of HPV which weighs 55 kilo Dalton. Morphologically it is similar to virions and is made up of 72 pentamers called capsomers. This protein has a spontaneously self-assembling ability as virus like particles (VLPs). Assembled VLPs are considered to be strong immunogens that may be rapidly identified by B cells (Buck, Day, & Trus, 2013; Chen, Garcea, Goldberg, Casini, & Harrison, 2000; Kirnbauer, Booy, Cheng, Lowy, & Schiller, 1992). It has been shown that injection of HPV16 L1-VLPs without adjuvant generates high levels of anti-HPV antibody responses (Harro et al., 2001; Tumban, Peabody, Peabody, & Chackerian, 2013). After the virus particles bind to the basal membrane, mature virus binds to the outer surface of the host cells using L1 protein. Then L1 becomes pliable resulting in releasing the virus genome in the host cell and its replication(Buck et al., 2013).HPVs are classified into both mucosal and cutaneous groups. These groups are divided into two groups of low-risk (LR-HPV) and high-risk

(HR-HPV), depending on the lesion's power in developing progressive malignancy. Low-risk mucosal HPVs produce genital warts; while high-risk HPVs produce squamous epithelial lesions and can cause invasive squamous carcinoma, such as CC(Clifford, Smith, Aguado, & Franceschi, 2003; Clifford, Smith, Plummer, et al., 2003; Kurman, Henson, Herbst, Noller, & Schiffman, 1994; Yuan et al., 2001; Zur Hausen, 1994). Studies have demonstrated that HPV16 and HPV18contribute toabout 70% of CC cases in which HPV16 and HPV18 are responsible for 51% and 16% of CC cases, respectively(Burd, 2003; Fernández-San Millán et al., 2008; Muñoz et al., 2003; Tumban et al., 2013; Yuan et al., 2001). However, epidemiological and experimental studies have shown that other members of the α-paplumaviridea family also play a role in carcinogenesis(Fakhraei et al., 2016; Haghshenas et al., 2013). For instance, according to the international agency for research in cancer in 2012, 12 types of HPV including HPV16, HPV 18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58 and HPV59 are responsible for CC and typesHPV26, HPV53, HPV66, HPV67, HPV68, HPV70 and HPV73, in rare cases, cause CC(Arbyn, Tommasino, Depuydt, & Dillner, 2014). Epidemiological studies suggest that CC could be prevented following the immunization against HPV infection through vaccination and that it could potentially decline CCfrequency(Roden & Wu, 2006). However, more knowledge about properties of this virus is needed for production and development of an effective vaccine(Monie et al., 2008). On other hand, production of an effective HPV vaccine can improve the prevention of cervix obstruction progress and reduce the treatment costs(Schiller & Lowy, 2012). Three VLP-based vaccines, Cervarix® bivalent vaccine (GlaxoSmithKline) and Gardasil® 4-valent and 9-valent vaccine have been introduced based on L1 antigen against HPV infection. There are some close morphological similarities between these

vaccines and native virions that lead to increase in antibody secretion, T cell cytotoxic activity and subsequently prevention of HPV infection. It is well-known that Cervarix® has a preventive maintenance against of HPV types 16 and 18. Also, it has been documented that Gardasil® 4-valent vaccine is remarked as a key protective against HPV16, HPV 18, HPV-6 and HPV-11. Gardasil® 9-valent causes protection againsta numbers of HPV types including31, 33, 45, 52, 16, 18, 6, 11, and 58(Negahdaripour et al., 2017; Panatto et al., 2015; Schiller & Lowy, 2014). But such vaccines should be purified from eukaryotic cells and they need more restricted cold chain for storage than those produce in prokaryotic system (Negahdaripour et al., 2017; Pouyanfard & Müller, 2017; Zhai & Tumban, 2016).These facilities are not well-available in alldeveloping countries where 80% ofCC occurs(Ferlay et al., 2015). Therefore, manufacturing a safer and easier vaccine for wider use in the world is one of the most important health priorities.A HPV 16/18 bivalent vaccine, ceolin, has been well-characterized that it is produced in *E.coli*{Gu, 2017 #95} and entered phase III clinical trial{Hu, 2014 #96}{Wu, 2015 #97}. Production of peptide vaccines is considered to be new targeted method that eliminates problems associated to current vaccines. Peptide vaccines are more commonly used than VLP-based vaccines. They are expressed in *E.coli* and are able to trigger high amount of humoral and cellular responses with high stability in room temperature(Naz & Dabir, 2007; Patronov & Doytchinova, 2013). Successfully, Immunoinformatics fields allow scientists to study epitopes, which in addition to reducing the numbers and facilitating experiments, which systematically leads to probable candidate epitopes identification besides facilitating experiments. As a result of new epitopes identification, the use of inactivated viral proteins could be eliminated(Naz & Dabir, 2007; Negahdaripour et al., 2017). Based on this, we designed a recombinant bioinformatics

vaccine based on L1 antigen of HPV type 16. This novel vaccine has two HPV epitopes that are able to stimulate the immune responses of CD4+ T, CD8+ T cells, humoral immunity, and interferon-γ (IFN-γ) production with higher antigenicity and no allergenic ability. In recent years, Toll-like receptor (TLR)-agonists were utilized as adjuvants. TLRs are a group of Pattern Recognition Receptors (PRRs) that can activate the innate immune system through Pathogen Associated Molecular Patterns (PAMPs) and enhance presenting of antigens though antigen presenting cells (APC)(Jiménez-Dalmaroni, Gerswhin, & Adamopoulos, 2016; Reed, Hsu, Carter, & Orr, 2016). Given that TLR-4 plays a natural role in controlling plenty of infectious diseases, a large number of adjuvants have been developed for TLR-4, some of them are available in commercial vaccines against HPV, such as Cervarix.

These adjuvants by optimal activation of APCs increase the responsiveness of the vaccine (Fox, Friede, Reed, & Ireton, 2010; Giannini et al., 2006; Johnson, 2013). Flagellin, a TLR-5 agonist, causes activation of both innate and adaptive immune system(Fox et al., 2010; Giannini et al., 2006; Johnson, 2013; Moyle, 2017; Moyle & Toth, 2013)and has been used as a vaccine adjuvant in several studies(Braga et al., 2010; Mori et al., 2012; Negahdaripour et al., 2017; Tarahomjoo, 2014). Therefore, in recent work, we focused on these two TLR-agonists as a potent adjuvant to enhance immunization in vaccine structure. Therefore, we evaluated structural stability, structural conformation, and its interaction with TLR-5 as well as solubility and successful expression of this structure in *E.coli.* Finally, in order to apply this vaccine to other HPVs, we used alignment of the amino acid sequences of the HPV type 16 virus with other members of α-papillomaviridea family.

2. Materials and Methods

2.1 Sequence retrieval

The amino acid sequence of HPV16 L1 was retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/) database with accession number of AAD33259.1 and saved in FASTA format for subsequent analysis.

2.2 Epitope prediction

2.2.1 Major histocompatibility complex class I (MHC-I) binding epitope prediction was performed by IEDB (www.iedb.org) database and Net-MHC 4.0 online server (http://www.cbs.dtu.dk/services/NetMHC/) and MHC-I humans allele with 9 mer length was selected. IEDB (Instructor/Evaluator Database) as resource database includes a list of B and T cell epitopes and information about MHC, binding MHC ligand of human and animal species established by the National Institute of Allergy and Infectious Diseases. It uses artificial neural network (ANN), stabilized matrix method (SMM), Combinatorial Peptide Libraries (CombLib), and NetMHCpan, based on availability of predictors and previously observed predictive performances. These methods predict peptides binding to MHC class I molecules (Kim et al., 2012). NetMHC 4.0 software is one of the tools that quantitatively predict the interaction between MHC-I and peptides, because of owning artificial neural networks which allows insertions and deletions in the alignment does has a better performance than prediction methods based on peptides of single lengths (Andreatta & Nielsen, 2015). In the software, the thresholds for strong and weak binders were set as rank 0.5 % and 2%, respectively.

2.2.2 CTL epitopes were determined using Combined Approach prediction in CTL Pread online server (http://crdd.osdd.net/raghava/ctlpred/index.html). CTL epitopes are the useful candidate for vaccine design which directly predicts CTL epitopes based on quantitative matrix (QM) and machine learning techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN) and facilitates MHC restriction in Tcell epitopes(Bhasin & Raghava, 2004).

2.2.3 Prediction of MHC-II binding epitopes was done by RANKPEP online server (http://imed.med.ucm.es/Tools/rankpep.html) that predicts MHC-II binding epitope using position specific scoring matrices (PSSMs)(Wang et al., 2008).

2.2.4 Prediction of Linear B Cell epitopes was performed using LBtope (http://crdd.osdd.net/raghava/lbtope) and BepiPred2.0 (http://www.cbs.dtu.dk/services/BepiPred). LBtope software is a method for prediction of B cell epitopes through SVM using dipeptide composition generated from the query amino acid sequence with an overall accuracy around 80%(Singh, Ansari, & Raghava, 2013). Then setting function was run onLBtope_Variable_non_redundant.

BepiPred software 2.0 combines hidden Markov model and a propensity scale method and predicts the location of linear Bcell epitopes. When working with the BepiPred software on the results page of this software, the Epitope Threshold can be changed from zero to one, which indicates specificity and sensitivity of the interaction with the epitope(Larsen, Lund, & Nielsen, 2006).

2.2.5 Selection of appropriate epitope: The results of all the above predictions were pooled and compared to identify the regions with high overlaps.

2.3 Evaluation of the selected epitopes; Based on IFN-γ inducing epitopes, antigenicity and allergenicity epitopes prediction

2.3.1 Selected epitopes were assessed using IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope) to determine the ability of epitopes to induce IFN-γ production. This web server classifies MHC binder epitopes into IFN-γ inducing (positive numbers) and non-inducing IFN-γ (negative numbers). Several methods apply in this software including; machine learning technique, motifs-based search, and hybrid approach were used for classification. Best prediction based on hybrid approach has 82.10% accuracy whereas the hybrid approach has maximum accuracy of 81.39% (Dhanda, Vir, & Raghava, 2013). To evaluateepitope prediction in this software, FASTA format of sequence will enter in respective box and in setting was evaluated Motif and SVM hybrid and IFN-γ versus Non IFN-γ.

2.3.2 Antigencity of epitopes was assessed via ANTIGENpro server (http://scratch.proteomics.ics.uci.edu/) and VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html).Web server of ANTIGENpro is the first server for predicting antigenicity based on protein sequence, which is often done by protein microarray data analysis (Magnan et al., 2010). To predict antigenic epitopes, the sequence of selected epitopes was entered into the corresponding box and then run the ANTIGENpro method.VaxiJen is the first server for prediction of antigens, which is available free online. Prediction bases of this server is based on physicochemical properties protein amino acid sequences instead of a sequence alignment approach(Doytchinova & Flower, 2007). Antigenicity of sequences was performed using VaxiJen. Therefore,

FASTA sequence of the protein was given to the software and virus word was selected in the box of the target and set the threshold on 0.4, and then submission was performed .

2.3.3 Allergenicity of selected epitopes was assessed using AllergenFP v.1.0 (http://ddg-pharmfac.net/AllergenFP/). This server identifies allergens based on their physicochemical properties. Accuracy of this server is about 88% (Dimitrov, Naneva, Doytchinova, & Bangov, 2013). After mentioned Steps, appropriate epitopes were chosen based on above stages.

2.4 Alignment of selected epitopes in family members of α-papillomaviridea

To understand that selected epitopes are conserved in family members of α-papillomaviridea, we aligned the selected epitopes in HPV16 and other α-papillomaviridea family members. First, the amino acid sequence of L1 of α-papillomaviridea family members was retrieved from the NCBI (https://www.ncbi.nlm.nih.gov/) database then blast of epitopes was done using CLC sequence viewer 8.0. CLC sequence viewer is one of the software for viewing and analyzing alignment of the protein, RNA and DNA(Sarwar, Rehman, & Ferzund, 2016).

2.5 Vaccine engineering, evaluation of construct physicochemical properties, antigenicity and allergenicity

The epitopes were selected based on the results of the previous sections, and the whole construction was designed by joining these epitopes to two TLR agonist adjuvants sequences (AS09 as TLR-4 adjuvant and N and C-terminals of flagellin

of *Salmonella enterica subsp.* as TLR-5). After designing the vaccine, the physicochemical properties of the vaccine structure including the number of amino acids, molecular weight, number and composition of atoms, molecular formula, number of positive and negative amino acid compounds, total number of positive and negative residues, isoelectric point (pI), instability index, aliphatic index and grand average hydropath city (GRAVY) were evaluated by Expasy'sProtParam online server (http://web.expasy.org/protparam/)(Gasteiger et al., 2005) and the solubility of vaccine protein in *E. coli* was evaluated by two severs: ccSOL omics server (http://service.tartaglialab.com/grant_submission/ccsol_omics) and Solpro (http://scratch.proteomics.ics.uci.edu). ccSOL omics is a web server that predicts solubility of proteins in *E. coli* based on coil/disorder, hydrophobicity B-sheet and α-helix properties. Accuracy of this server is 74% (Agostini, Cirillo, Livi, Delli Ponti, & Tartaglia, 2014). Solpro is another web server that predicts solubility of proteins upon overexpression in *E. coli*. This server has an accuracy of higher than 74% using multiple runs of 10-fold cross validation (Magnan, Randall, & Baldi, 2009). Allergenicity of vaccine was assessed using AllergenFP v.1.0 (http://ddg-pharmfac.net/AllergenFP/) and its antigenicity was assessed by ANTIGENpro server (http://scratch.proteomics.ics.uci.edu) and VaxiJen 2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) at a threshold value 0.4.

2.6 Tertiary structure analysis and prediction of discontinuous Bcell epitopes

The prediction of Tertiary structure (3D) model was determined using I-TASSER online server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). I-TASSER server is a modeling server of protein tertiary structure based on secondary-structure enhanced Profile-Profile threading Alignment (PPA) and iterative implementation of the Threading Assembly Refinement (TASSER) program. In this server unlike many MQAP programs that evaluate models solely based on the structure of the final models, the performance scoring includes simulation information and modeling parameters. This software is free for university users and allows them to automatically design high-quality models of 3-D protein structure based on their amino acid sequences(Yang & Zhang, 2015). This server presents five models as a result of the prediction. Therefore, the model with highest confidence score (c-score) was accepted and observed with Pymol software v2.1.1. Pymol is one of the most software tools to visualize molecular structure(Shin, Lee, Heo, Lee, & Seok, 2014). After that, Galaxy Refine was used to determine the refinement of3D model. Galaxy Refine is an open source online software in which was successfully tested in CASP10 (Critical assessment of techniques for protein structure prediction) as a refinement method. This method first rebuilt the side chains of protein using molecular dynamic simulationside-chain repacking and also overall structure relaxation. According to the assessment report of CASP10, this method indicates the best performance in quality improvement in local structure, as well as it can improve quality of global and local structure, also it can improve quality of both global and local structure(Shin et al., 2014).

Finally, to validate the refined 3D structures and compare the models, The RAMPAGE (Ramachandran Plot Assessment) (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), and ERRAT (http://services.mbi.ucla.edu/ERRAT/) servers were used. Rampage analysis backbone conformation through calculation of phi-psi torsion angles for amino acids in the structure. Then classified the amino acids in 3 region of favored, allowed and outlier(Lovell et al., 2003). ERRAT is a web server that evaluates the validation of protein structure based on characteristic atomic interaction. This server compares the statistics of non-bonded atom–atom interactions in the query sequence with a database of 96 reliable high-resolution crystallography structures. This server is sensitive to errors in position of backbone atoms on the order of 1.5 A° and method used for atomic refinement so that unrefined structures generally do not score well(Colovos & Yeates, 1993). Discontinuous Bcell epitopes were predicted from the 3D structure of the vaccine by Disco Tope 2.0 server (http://www.cbs.dtu.dk/services/DiscoTope/). This server calculates the final scores by two approaches: contact numbers derived from surface accessibility and a novel epitope propensity amino acid score(Kringelum, Lundegaard, Lund, & Nielsen, 2012). The default threshold value was set on-3.7 in which the sensitivity and specificity are 0.47 and 0.75, respectively.

2.7 Exploration of interaction between vaccine and human Toll like receptor 5

First, the three-dimensional structure of the TLR-5 was obtained from PDB data base (www.rcsb.org) with code of 3J0A. Then protein-protein docking of vaccine structure and TLR-5was separately done using CLUSPRO 2.0 online server

(cluspro.bu.edu/login.php). CLUSPRO 2.0 is a web server for protein-protein docking that assays direct docking of two interacting proteins in 3 stages:

1. Running of PIPER (a rigid body docking program based on the Fast Fourier Transform (FFT) correlation approach).
2. RMSD which based on clustering of the structures generated was done to find the largest clusters that will represent the most likely models of the complex.
3. Analyzing stability of the clusters and refinement of structure were carried out by short Monte Carlo simulations and the medium-range optimization method SDU, respectively (Kozakov et al., 2013; Kozakov et al., 2010). To do molecular docking study by this software, the candidate vaccine was considered as a ligand and TLR-5 considered as a receptor.

2.8 Codon optimization and in silico cloning

In order to prepare a suitable vaccine sequence to be clone and express in an appropriate expression vector, the reverse translation and codon optimization of the designed vaccine sequencewas performed by the Codon Usage Wrangler server (https://www.mrc-lmb.cam.ac.uk/ms/methods/codon.html). Properties of the optimized DNA sequence such as Codon Adaptation Index (CAI), GC content, Codon Frequency and Distribution (CFD) have a key role in achieving a high-level of protein expression in the host. Therefore, the mentioned properties were assessed by GenScript online server (https://www.genscript.com/tools/rare-codon-analysis)(Negahdaripour et al., 2017)and finally, the restriction sites of *ECORI* and *BamHI*were added to the N- and C-terminals of the DNA sequence of vaccine and it was prepared for cloning in *E. coli* as a host.

3. Results

3.1 Primary structure

The protein sequence of HPV16 L1 with accession number of AAD33259.1 was driven in NCBI .

3.2 Epitope prediction

3.2.1 MHC-I binding epitope prediction

MHC-I binding epitope with a length of 9-mer was predicted by database of IEDB and NetMHC 4.0. MHC-I binding epitopes that have an overlap in IEDB and NetMHC 4.0 were selectedfor future studies.

3.2.2 Prediction of CTL epitopes

Prediction of CTL epitopes was done by free online server CTL Pred. The prediction result of this server is given in several formats, including overlap display, color display, and a tabular display. In the color display format, predicted epitopes of CTL are indicated by staining. The beginning amino acid of each epitope is indicated by a red color and other amino acids are blue. The results are displayed in 100 amino acid columns per line. This format is very useful in detecting accumulation regions of CTL epitopes in a sequence. In overlap display, all CTL epitopes are displayed in separate lines. This display is presented on a scale that indicates the position of the epitope in the antigenic sequence and epitopes are displayed by staining of amino acids and final output format is tabular that is the most commonly displayed format. This format is widely used by most predictive methods. In this format, peptides are displayed in a table according to low scoring. Sequences that overlap in the results of all three formats are selected for subsequent studies.

3.2.3 MHC-II binding epitope prediction

MHC-II binding epitope of L1 was done using RANKPEP.

3.2.4 Linear B cell epitope prediction

This prediction was done by two servers separately and their results were compared.

1) LBtope: On the result page of this software, a colorful block is shown for each sequence. The color of each residue displays the percentage of the probability that the amino acid is in the middle of the motif. So the possibility of immunodominantepitpe arranged based on colors in which red for 81-100%, green for 61-80%, blue for 41-60%, yellow for 21-40%, and black for 0-20%. Albeit there was no yellow and block colors in our sequence. Motifs that have a possibility above 80% were considered as epitopes.
2) BepiPred: The prediction results are shown with several bars. The first bar contains the amino acid sequence of the protein in which characterized by orange color and more colorful bar has higher percentage of residue in epitope. Amino acids that have a threshold 0.5 or higher were indicated with the symbol E above the amino acids. Second bars indicate the protein structure. In this bar, pink, blue, and orange colors to indicate helix, sheet and coil, respectively. Exposed epitopes are indicated with a capital letter E and buried epitopes are indicated with a capital letter B. Overlapping epitopes were selected in two servers for future studies.

3.3 Selection of appropriate epitopes

The obtained results from all above predictions were compared to each other in order to obtain overlapping regions that probably have a stronger immune response (Table 1).

3.4 Evaluation of the predicted-epitopes based on IFN-γ inducing, Allergenicity and Antigenicity

Table 1 shows the results of prediction of allergenicity and antigenicity in selected epitpes. The selected epitopes were evaluated by IFNepitope server software to determine which of them had a better ability to induce IFN-γ. Antigenic capacity of selected epitopes were assessed by ANTIGENpro and vaccijen2.0 servers. In addition, allergenicity properties were also predicted by AllergenFP v.1.0.

3.5 Confirmation of conservativity of selected epitopes in all members of the α-papilloma viridea family

Conservation of selected epitopes in members of the α-papillomaviridea family (Types HPV 26, 51, 69, 82, 30, 53, 56, 66, 18, 39, 45, 59, 68, 70, 85, 97, 16, 31, 33, 35, 52, 58, 67, 34and 73) were assessed by retrieve L1 sequence in members of the α-papillomaviridea family in NCBI. Alignment was carried out in CLC Sequence viewer 8.0. As it has shown in Fig 1, alignment of L1 sequences indicated that screening peptides, especially in anchoring amino acids 2, 8 and 9, are found to share high sequence similarity among all members of α-papillomaviridea family.

3.6 Vaccine engineering and evaluation of physicochemical properties of the candidate vaccine structure, antigenicity and allergencity

Peptide sequences of YGDSLFFYLRREQMFVRH and YVARTNIYYHAGTSRLLAV GHPYFPIK were selected as immunodominant epitopes in the vaccine structure. These epitopes are joined to two TLR agonist adjuvants: RS09 that is a short peptide TLR-4 agonist with two parts, the N-terminal head and C-terminal tail of flagellin which is as an

agonist of TLR-5. A short linker sequence of GGS was used to link peptide sequences to each other and with two TLR agonist adjuvants. Finally, the vaccine structure that is a peptide with 278 amino acids was designed by linking the sequences. Fig. 2 shows the schematic diagram of the vaccine was designed by Illustrator for Biological Sequences (IBS)Ver 1.0 server (61).

Physicochemical properties and molecular structure of the designed vaccine were analyzed by ProtParam server (Table 2). The results showed that the vaccine is stable (instability index: 39.94) with a molecular weight (29.96 KDa) and isoelectric point (pI: 7.97). Since protein insolubility is a major problem for many experimental studies, therefore, we should examine the status of protein solubility upon overexpression. ccSOL omics and Solpro servers predicted the probability of the vaccine solubility upon expression in E. coli host is 25% and 0.544158, respectively. These findings emphasize that our designed peptide vaccine is soluble (Table 2). In addition, prediction of consensus antigenicity of candidate vaccine using predictions using ANTIGENpro and vaccijen 2.0 tools revealed that the final whole protein construct is as an antigen. Meanwhile, the result of AllergenFP tool showed that our construct has no allergenic properties.

3.7 Tertiary structure analysis and prediction of discontinuous B-cell epitopes

Tertiary structure analysis was examined using I-TASSER, Pymol, Galexi Refine, ERRAT and RAMPAGE (Fig. 3 and Fig. 4 and Table 3). Discontinuous B-cell epitopes were predicted from the 3D structure of the vaccine by Disco Tope 2.0 server (Table4).

3.8 Molecular docking of the vaccine with TLR-5

To evaluate the interaction of TLR-5 with the candidate vaccine structure, protein-protein docking of 3D models was performed using the Cluspro 2.0 online software (Fig.5). The lower Gibs free energy for an interaction of TLR5-vaccine model was -1274.6J.

3.9 Codon optimization and in silico cloning

CFD means rare codons that decrease efficiency of translation and even hind the translational machinery. These results indicate that the optimized DNA sequence of the vaccine construction is suitable for cloning and expression in *E.coli* (Fig.6).

Finally, restriction sites of EcoRI and BamHI were inserted at the start and end of the DNA sequence, respectively. Finally, the genetic vaccine was prepared for cloning and expression in *E. coli.*

4. Discussion

Oncogenic human papillomaviruses are the main causes of CC (Monie et al., 2008; Yuan et al., 2001). These viruses are also involved in the initiation and development of other cancers, such as Vulva, Anus, Penis, Vagina, Mouth, Throat and gastric cancer (Fakhraei et al., 2016; Tarahomjoo, 2014). L1 is the main protein in the capsid structure of these viruses that has a role key in infection (Buck et al., 2013; Chen et al., 2000; Kirnbauer et al., 1992). Producing of a vaccine on L1 protein can lead to immunization against HPV infection to potentially and prevent related cancers (Roden& Wu, 2006). Available HPV vaccines have not protective coverage against all of HPV types (Arbyn et al., 2014), also these vaccines should be purified in eukaryotic cells in which meet some obstacles in massive production and more restricted to cold chain storage. These limitations might

highly increase cost of vaccination program in middle and low income developing countries, where about 80% of CC exist (Ferlay et al., 2015; Pouyanfard & Müller, 2017; Zhai &Tumban, 2016). Therefore, thorough efforts should be made to replace existing vaccines with cheaper vaccines and more stable at room temperature. Moreover, the efficacy of candidate vaccine should be large enough to produce strong humoral and cellular immune responses against the most important HPV types that are known to cause CC and other HPV associated cancers. In order to develop a vaccine using bioinformatics, two sequences were identified as a capable epitopes, which activates humoral response, Th1 and CTL cells. Unlike having high antigenicity, the vaccine should not be allergic. The sequence of the immunodominant epitopes must be selected in such a way as to maximize structural similarity among the members of α-papillomaviridea family.

Using an appropriate adjuvant for vaccine production may enhance the efficacy of target antigens, potency and stability of vaccine in individuals who do not respond well to the vaccine. Using the adjuvants in the vaccine structure produces stronger and longer immune responses. In recent years, TLR agonists have been used as adjuvants in the vaccine structure and their use has been increasingly developed (Jiménez-Dalmaroni et al., 2016). TLR-4 has been reported to express in various regions of the cervix. On the other hand, Giannini *et al* revealed that the use of monophosphoryl lipid A (MPL) and aluminum salt as TLR-4 agonist in Cervarix vaccine this drug as a saline agonist in comparison to aluminum salt alone has a greater effect on the induction and elongation of immune response (Giannini et al., 2006). We used RS09 as a successful synthetic TLR-4 agonist peptide that resembles LPS by interacting with TLR-4 (Shanmugam et al., 2012). It has been reported that the combination of two or three TLR agonists may

significantly increase cellular and humoral immune responses and so that it is more beneficial (Kasturi et al., 2011; Orr et al., 2014; Zhu et al., 2010).

Flagellin improves innate and acquired immunity by increasing cytokine production using a variety of innate cell types, attracting T and B cells into the secondary lymphoid sites, and activating TLR5 , $CD11c^{+}$, and T lymphocyte cells in a separate method of cognate-Ag recognition, Therefore it has been used in vaccine structure as an adjuvant. Furthermore, it has also been shown that TLR5 is present in the cervix (Fazeli, Bruce, &Anumba, 2005; Hayashi et al., 2001; Lu & Sun, 2012; Nguyen et al., 2013; Tarahomjoo, 2014; Yoon et al., 2012).That's why we used the flagellin of Salmonella enterica subsp. enterica serovar Dublin (TLR5 agonist) as the second adjuvant in vaccine structure. It has been demonstrated that flagellin has 4 domains including D0, D1, D2, and D3. Only the D0 and D1 domains have N- termini and C-termini of the molecule interact with the TLR5 and its D2 and D3 domains are responsible for antigenicity (Lu & Sun, 2012; Song, Jeon, Namgung, Hong, & Yoon, 2017; Yonekura, Maki-Yonekura, &Namba, 2003; Yoon et al., 2012). In line to other studies, D2 and D3 domains of flagellin were deleted and only the N- and C- termini of the flagellin were used in head and tail sections of the vaccine structure (Hajighahramani et al., 2017). Studies have proven that the use of linkers in multi-epitope vaccines would be useful for facilitating antigen processing and avoiding the formation of junctional epitopes (Livingston et al., 2002; Nezafat et al., 2016). Therefore, GGS linkers were used to join different vaccine sections (Negahdaripour et al., 2017). The small, non-polar glycine and polar serine amino acids cause flexibility and solubility of the vaccine structure (Argos, 1990; Bhattacharya, Nowotny, Cao, & Cheng, 2016). Since N- and C- termini of flagellin can interact freely with TLR5, no linker was added to the two termini of the designed vaccine. Finally, the protein vaccine was

designed with 278 amino acids. In addition to activating effective immune responses, an efficient vaccine must have good physicochemical properties during production, formulation, storage and administration. Based on the results of bioinformatics predictions, our designed vaccine was soluble, stable, non-allergen, immunogen with pI 7.97. Furthermore, according to the results of Ramachandran plot and ERRAT evaluations the vaccine structure was refined, which led to a higher quality. The capacity of the final structure of candidate vaccine in inducing humoral response was assessed in Discotope 2.0 which predicts conformational B cell epitope, and their results reveal the vaccine is a strong immunogenic peptide. After the design of a high-quality vaccine, its interaction with TLR5 was evaluated by CLUSPRO2.0 docking online software. It found that our construct interacted efficiently with TLR5 with a negative Gibs free energy.

To have a high throughput recombinant protein expression in prokaryotic system, GenScript tool was shown higher expression efficacy of the full length peptide vaccine in *E. coli* host. Also the high solubility protein which was predicted by the Solpro and ccsolomics software is another positive feature of the vaccine that prevents the accumulation of protein in the inclusion bodies. The results showed our designed vaccine is suitable to be used against cervical cancer. We added appropriate sequence of restriction enzymes to two N- and C-terminus of the DNA vaccine construct. The final vaccine construct might be expressed in *E. coli* and used to prevent the HPV virus infections to reduce the burden of cervical cancer and other HPV related cancers. Although the outputs of bioinformatics softwares confirm high immunogenicity of the modeled multiepitopes HPV vaccine, the presence of low accessibility B cell linear epitope might be a limitation in our study.

5. Conclusions

In this study, using several available bioinformatics tools, we tried to design an efficient HPV peptide vaccine by using two specific immunodominant epitopes. The vaccine is able to induce humoral and cell-mediated immunities, which are required for protection against all types of HPV viruses. It is comprise of several adjuvants to increase immunity. The physicochemical properties of the structure were shown to be appropriate by using computational tools. Thus, this vaccine might be expected to prevent the HPV infections and other HPV-related cancers.

6. List of abbreviations

abbreviation	Full name
Ag	Antigen
A°	Angstrom
bp	base pair
CD	Cluster of Differentiation
CTL	Cytotoxic T-Lymphocyte
D	Domain
HPV	human papilloma viruses
hTLR-5	Human toll like receptor 5
IFN-γ	Interferon gamma
J	Joule
LPS	lipopolysaccharide
MHC-I	Major histocompatibility complex class I
MHC-II	Major histocompatibility complex class II
PI	Isoelectric point
TLR	Toll like receptor

3D	Three Dimensional

Declarations

Statement

Zahra Yazdani collecting and interpretation of data, performing the analysis and writing the draft.

Alireza Rafiei supervision the project, analyzing and interpreting the results and editing of manuscript.

Reza Valadan having technical help revised the manuscript critically.

Hossein Ashrafi having technical help revised the manuscript critically.

MarziehSharifi Pasandi screeching and editing of manuscript.

Mostafa Kardan screeching and editing of manuscript

All authors read and approved the final manuscript.

Funding

This study received a small grant (Grant no: 5269) from Research and Technology deputy of Mazandaran University of Medical Sciences.

Availability of data and materials

In this study, protein sequence of HPV16 L1 was obtained from NCBI (https://www.ncbi.nlm.nih.gov/) and analyzed by different software.

Authors' contributions

Z.Y contributed to collecting and interpretation of data, performing the analysis and writing the draft. A.R supervised the project and analyzed and interpreted the results and editing of manuscript. R.V and H.A had technical help revised the manuscript

critically. M.S and M.K contributed in screeching and editing of manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The protocol of this study approved by Research Ethics committee of Mazandaran University of Medical Sciences

Consent for publication

Not applicable.

Declarations of interest

The authors declared no conflict of interest.

Acknowledgements

The authors would like to cordially thank Seyed Ahmad Mohammadi for his helpful advices.

References

Agostini, F., Cirillo, D., Livi, C. M., Delli Ponti, R., & Tartaglia, G. G. (2014). cc SOL omics: A webserver for solubility prediction of endogenous and heterologous expression in Escherichia coli. *Bioinformatics, 30*(20), 2975-2977.

Andreatta, M., & Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics, 32*(4), 511-517.

Arbyn, M., Tommasino, M., Depuydt, C., & Dillner, J. (2014). Are 20 human papillomavirus types causing cervical cancer? *The Journal of pathology, 234*(4), 431-435.

Bhasin, M., & Raghava, G. (2004). Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine, 22*(23-24), 3195-3204.

Braga, C. J., Massis, L. M., Sbrogio-Almeida, M. E., Alencar, B. C., Bargieri, D. Y., Boscardin, S. B., . . . Ferreira, L. C. (2010). CD8+ T cell adjuvant effects of Salmonella FliCd flagellin in live vaccine vectors or as purified protein. *Vaccine, 28*(5), 1373-1382.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*.

Buck, C. B., Day, P. M., & Trus, B. L. (2013). The papillomavirus major capsid protein L1. *Virology, 445*(1-2), 169-174.

Burd, E. M. (2003). Human papillomavirus and cervical cancer. *Clinical microbiology reviews, 16*(1), 1-17.

Chen, X. S., Garcea, R. L., Goldberg, I., Casini, G., & Harrison, S. C. (2000). Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Molecular cell, 5*(3), 557-567.

Clifford, G., Smith, J., Aguado, T., & Franceschi, S. (2003). Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis. *British journal of cancer, 89*(1), 101.

Clifford, G., Smith, J., Plummer, M., Munoz, N., & Franceschi, S. (2003). Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *British journal of cancer, 88*(1), 63.

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science, 2*(9), 1511-1519.

Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology direct, 8*(1), 30.

Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2013). AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics, 30*(6), 846-851.

Doytchinova, I. A., & Flower, D. R. (2007). Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine, 25*(5), 856-866.

Fakhraei, F., Haghshenas, M. R., Hosseini, V., Rafiei, A., Naghshvar, F., & Alizadeh‐Navaei, R. (2016). Detection of human papillomavirus DNA in gastric carcinoma specimens in a high-risk region of Iran. *Biomedical reports, 5*(3), 371-375.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer, 136*(5), E359-E386.

Fernández-San Millán, A., Ortigosa, S. M., Hervás-Stubbs, S., Corral-Martínez, P., Seguí-Simarro, J. M., Gaétan, J., . . . Veramendi, J. (2008). Human papillomavirus L1 protein expressed in tobacco chloroplasts self-assembles into virus-like particles that are highly immunogenic. *Plant biotechnology journal, 6*(5), 427-441.

Fox, C. B., Friede, M., Reed, S. G., & Ireton, G. C. (2010). Synthetic and natural TLR4 agonists as safe and effective vaccine adjuvants. In *Endotoxins: structure, function and recognition* (pp. 303-321): Springer.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607): Springer.

Giannini, S. L., Hanon, E., Moris, P., Van Mechelen, M., Morel, S., Dessy, F., . . . Losonksy, G. (2006). Enhanced humoral and memory B cellular immunity using HPV16/18 L1 VLP vaccine formulated with the MPL/aluminium salt combination (AS04) compared to aluminium salt only. *Vaccine, 24*(33-34), 5937-5949.

Haghshenas, M., Golini-Moghaddam, T., Rafiei, A., Emadeian, O., Shykhpour, A., & Ashrafi, G. H. (2013). Prevalence and type distribution of high-risk human papillomavirus in patients with cervical cancer: a population-based study. *Infectious agents and cancer, 8*(1), 20.

Harro, C. D., Pang, Y.-Y. S., Roden, R. B., Hildesheim, A., Wang, Z., Reynolds, M. J., . . . Karron, R. A. (2001). Safety and immunogenicity trial in adult volunteers of a human papillomavirus 16 L1 virus-like particle vaccine. *Journal of the National Cancer Institute, 93*(4), 284-292.

Jiménez-Dalmaroni, M. J., Gerswhin, M. E., & Adamopoulos, I. E. (2016). The critical role of toll-like receptors—from microbial recognition to autoimmunity: a comprehensive review. *Autoimmunity reviews, 15*(1), 1-8.

Johnson, D. A. (2013). TLR4 agonists as vaccine adjuvants: a chemist's perspective. *Expert review of vaccines, 12*(7), 711-713.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., . . . Bourne, P. E. (2012). Immune epitope database analysis resource. *Nucleic acids research, 40*(W1), W525-W530.

Kirnbauer, R., Booy, F., Cheng, N., Lowy, D., & Schiller, J. (1992). Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic. *Proceedings of the National Academy of Sciences, 89*(24), 12180-12184.

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics, 81*(12), 2159-2166.

Kozakov, D., Hall, D. R., Beglov, D., Brenke, R., Comeau, S. R., Shen, Y., . . . Paschalidis, I. C. (2010). Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. *Proteins: Structure, Function, and Bioinformatics, 78*(15), 3124-3130.

Kringelum, J. V., Lundegaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS computational biology, 8*(12), e1002829.

Kurman, R. J., Henson, D. E., Herbst, A. L., Noller, K. L., & Schiffman, M. H. (1994). Interim guidelines for management of abnormal cervical cytology. *JOURNAL-AMERICAN MEDICAL ASSOCIATION, 271*, 1866-1866.

Larsen, J. E. P., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome research, 2*(1), 2.

Lovell, S. C., Davis, I. W., Arendall III, W. B., De Bakker, P. I., Word, J. M., Prisant, M. G., . . . Richardson, D. C. (2003). Structure validation by Cα geometry: φ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics, 50*(3), 437-450.

Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics, 25*(17), 2200-2207.

Magnan, C. N., Zeller, M., Kayala, M. A., Vigil, A., Randall, A., Felgner, P. L., & Baldi, P. (2010). High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics, 26*(23), 2936-2943.

Monie, A., Hung, C.-F., Roden, R., & Wu, T. C. (2008). Cervarix™: a vaccine for the prevention of HPV 16, 18-associated cervical cancer. *Biologics: targets & therapy, 2*(1), 107.

Mori, J., Vranac, T., Smrekar, B., Černilec, M., Šerbec, V. Č., Horvat, S., . . . Jerala, R. (2012). Chimeric flagellin as the self-adjuvanting antigen for the activation of immune response against Helicobacter pylori. *Vaccine, 30*(40), 5856-5863.

Moyle, P. M. (2017). Biotechnology approaches to produce potent, self-adjuvanting antigen-adjuvant fusion protein subunit vaccines. *Biotechnology advances, 35*(3), 375-389.

Moyle, P. M., & Toth, I. (2013). Modern subunit vaccines: development, components, and research opportunities. *ChemMedChem, 8*(3), 360-376.

Muñoz, N., Bosch, F. X., De Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K. V., . . . Meijer, C. J. (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. *New England Journal of Medicine, 348*(6), 518-527.

Naz, R. K., & Dabir, P. (2007). Peptide vaccines against cancer, infectious diseases, and conception. *Frontiers in bioscience: a journal and virtual library, 12*, 1833-1844.

Negahdaripour, M., Eslami, M., Nezafat, N., Hajighahramani, N., Ghoshoon, M. B., Shoolian, E., . . . Ghasemi, Y. (2017). A novel HPV prophylactic peptide vaccine, designed by

immunoinformatics and structural vaccinology approaches. *Infection, Genetics and Evolution, 54*, 402-416.

Panatto, D., Amicizia, D., Bragazzi, N. L., Rizzitelli, E., Tramalloni, D., Valle, I., & Gasparini, R. (2015). Human papillomavirus vaccine: State of the art and future perspectives. In *Advances in protein chemistry and structural biology* (Vol. 101, pp. 231-322): Elsevier.

Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open biology, 3*(1), 120139.

Pouyanfard, S., & Müller, M. (2017). Human papillomavirus first and second generation vaccines–current status and future directions. *Biological chemistry, 398*(8), 871-889.

Reed, S. G., Hsu, F.-C., Carter, D., & Orr, M. T. (2016). The science of vaccine adjuvants: advances in TLR4 ligand adjuvants. *Current opinion in immunology, 41*, 85-90.

Roden, R., & Wu, T.-C. (2006). How will HPV vaccines affect cervical cancer? *Nature Reviews Cancer, 6*(10), 753.

Sarwar, M. A., Rehman, A., & Ferzund, J. (2016). Database Search, Alignment Viewer and Genomics Analysis Tools: Big Data for Bioinformatics. *International Journal of Computer Science and Information Security, 14*(12), 317.

Schiller, J. T., & Lowy, D. R. (2012). Understanding and learning from the success of prophylactic human papillomavirus vaccines. *Nature Reviews Microbiology, 10*(10), 681.

Schiller, J. T., & Lowy, D. R. (2014). Raising expectations for subunit vaccine. *The Journal of infectious diseases, 211*(9), 1373-1375.

Shin, W.-H., Lee, G. R., Heo, L., Lee, H., & Seok, C. (2014). Prediction of protein structure and interaction by GALAXY protein modeling programs. *Bio Design, 2*(1), 1-11.

Singh, H., Ansari, H. R., & Raghava, G. P. (2013). Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PloS one, 8*(5), e62216.

Tarahomjoo, S. (2014). Utilizing bacterial flagellins against infectious diseases and cancers. *Antonie Van Leeuwenhoek, 105*(2), 275-288.

Tommasino, M. (2014). *The human papillomavirus family and its role in carcinogenesis.* Paper presented at the Seminars in cancer biology.

Tumban, E., Peabody, J., Peabody, D. S., & Chackerian, B. (2013). A universal virus-like particle-based vaccine for human papillomavirus: longevity of protection and role of endogenous and exogenous adjuvants. *Vaccine, 31*(41), 4647-4654.

Wang, P., Sidney, J., Dow, C., Mothe, B., Sette, A., & Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS computational biology, 4*(4), e1000048.

Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research, 43*(W1), W174-W181.

Yuan, H., Estes, P. A., Chen, Y., Newsome, J., Olcese, V. A., Garcea, R. L., & Schlegel, R. (2001). Immunization with a pentameric L1 fusion protein protects against papillomavirus infection. *Journal of virology, 75*(17), 7848-7853.

Zhai, L., & Tumban, E. (2016). Gardasil-9: A global survey of projected efficacy. *Antiviral research, 130*, 101-109.

Zur Hausen, H. (1994). Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. In *Human pathogenic papillomaviruses* (pp. 131-156): Springer.

Figure Legends

Fig. 1 Alignment of selected epitopes in α-papillomaviridea: (a) Alignment of selected epitopes in HPV-16 and α-5 species, b) Alignment of selected epitopes in HPV-16 and α-6 species, c) Alignment of selected epitopes in HPV-16 and α-7 species, d) Alignment of selected epitopes in HPV-16 and α-9 species, e) Alignment of selected epitopes in HPV-16 and α-11 species. The locations of MHC (amino acids with numbers 2, 8 and 9), are conserved in α-papillomaviridea family.

Fig. 2

a) Graphical picture of vaccine structure, the vaccine is consisted of five segments: Two epitopes and adjuvants: flagellin (TLR5 agonist), which has two segments: N terminal (head) and C-terminal (tail) and RS09 (TLR-4 agonist) and these segments are linked together by short linkers.

b)The refined 3D structure of the designed vaccine. The 3D structure of the designed vaccine was suggested through homology modeling by I-Tasser, then the best proposed model was refined by GalaxyRefine and display using Pymol software.

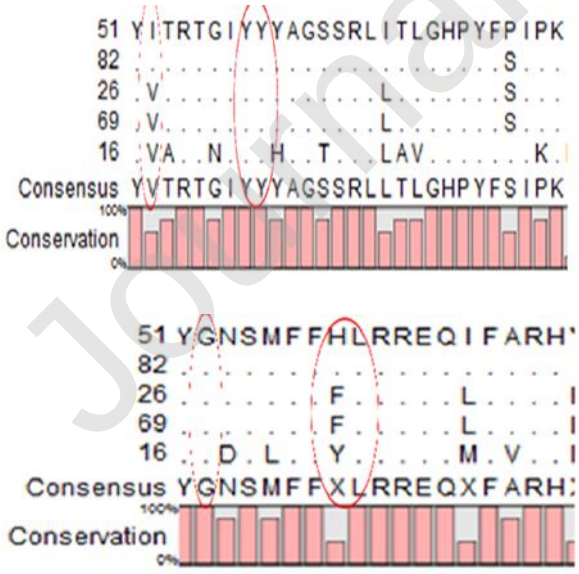
Fig. 3 The result of the structural quality of the vaccine in ERRAT. ERRAT plot showed the overall quality factor of the refined structure as 96.617. Good high resolution structures usually produce values around 95% or higher.

Fig.4 Docking model (cartoon representation) of human TLR5 protein in complex with the vaccine molecule obtained by Cluspro. TLR5 protein is shown in white. N-terminus and C-terminus of flagellin are shown in blue and red, respectively.TLR4 adjuvant is shown in green and epitopes are shown in yellow color. Some of the interacting residues

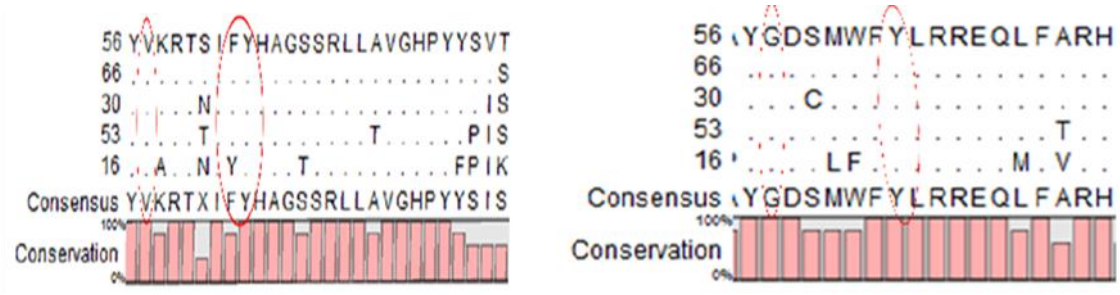
of flagellin N and C- terminals and TLR4 are shown in shown in magnify. Docked model was visualized via Pymol software.

Fig. 5 Evaluation of the three important parameters of the codon-optimized gene for high-level protein expression in *E. coli* as host. a) CAI of gene sequence is 1. A CAI of > 0.8 is considered as good for expression in selected host. b) The average GC content of the sequence is 59.04%. c) Codon frequency distribution (CFD) value of gene sequence is100. The CFD = 100 supports maximum protein expression in the desired host

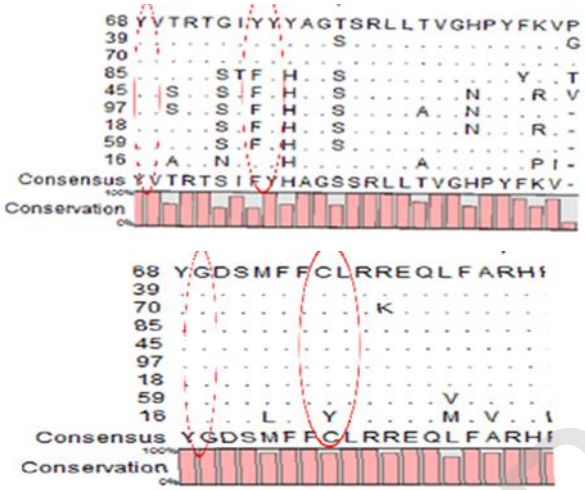
a)



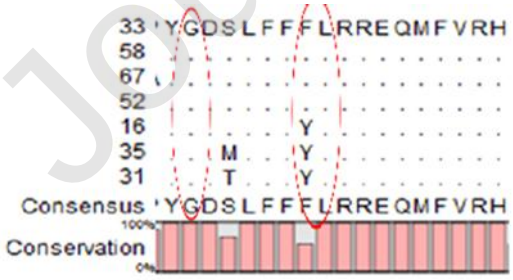
b)



c)



d)



e)

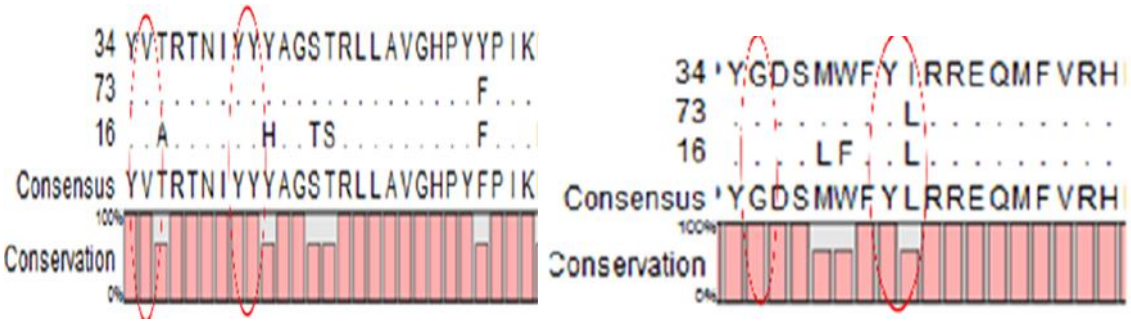
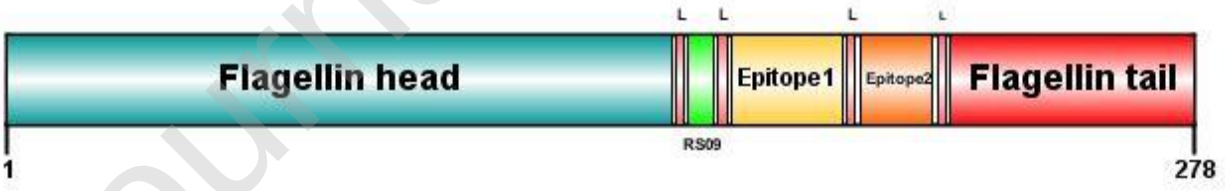


Figure1

a)



b)

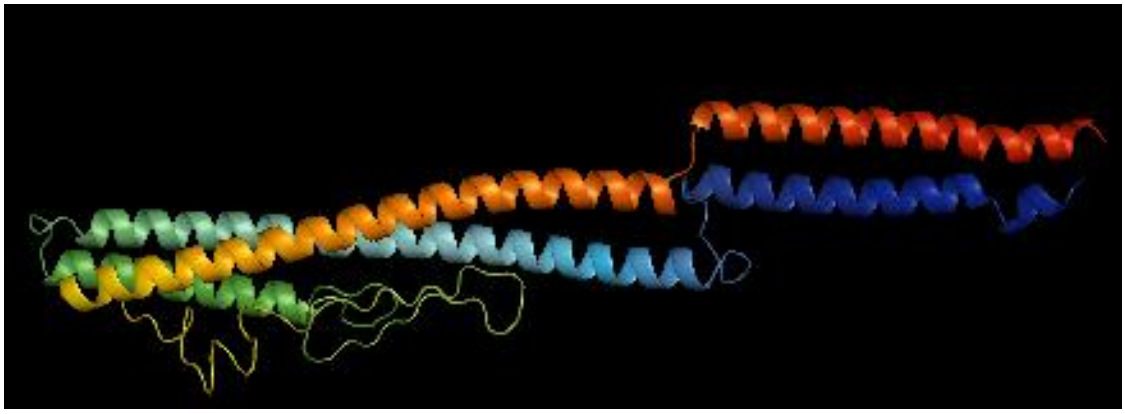


Figure2

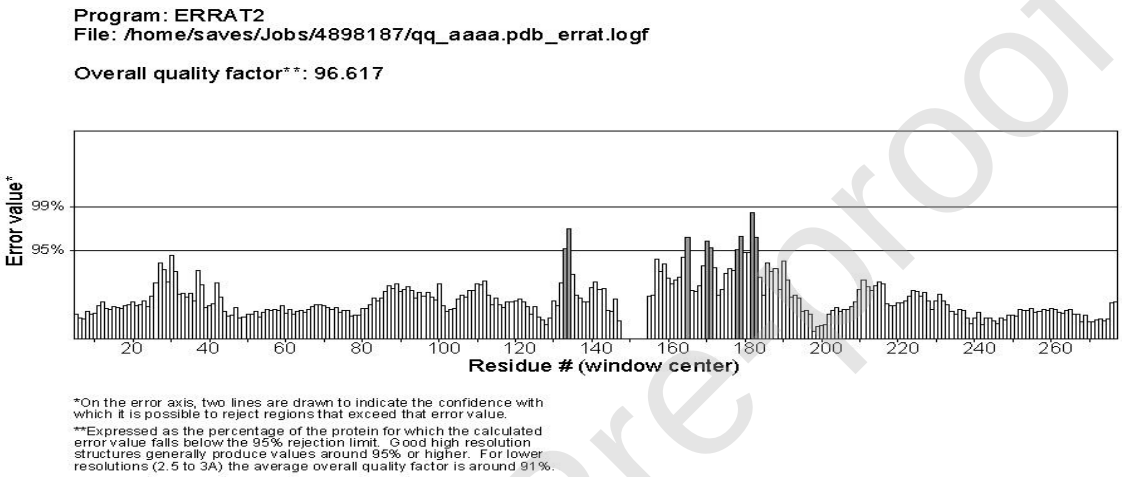


Figure3

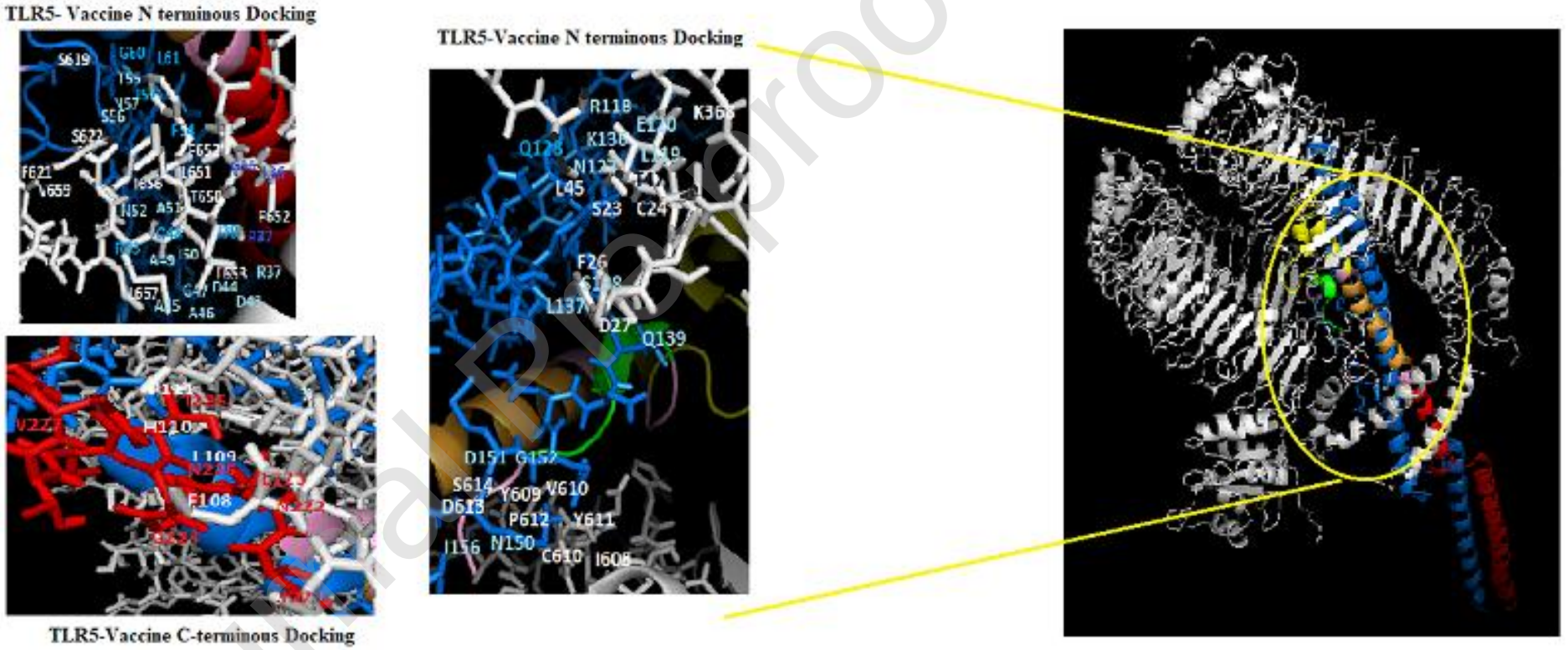
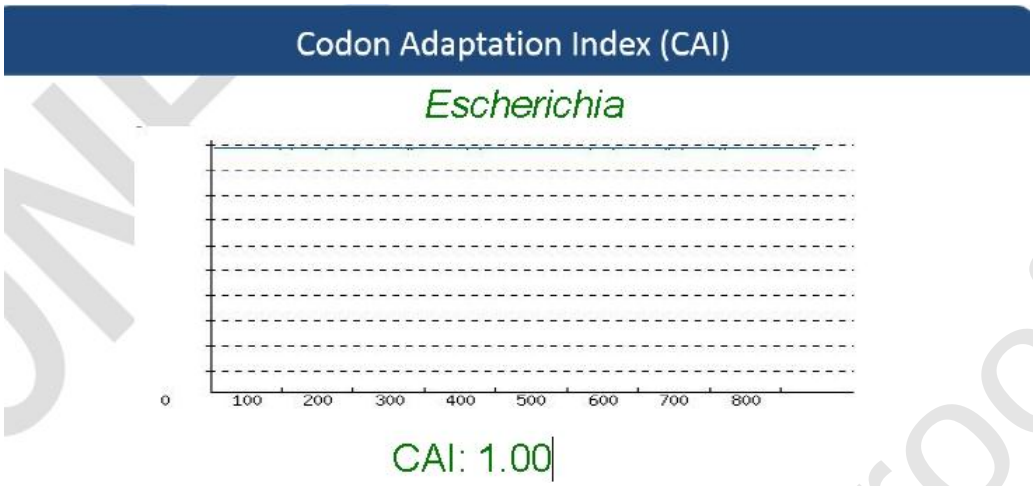
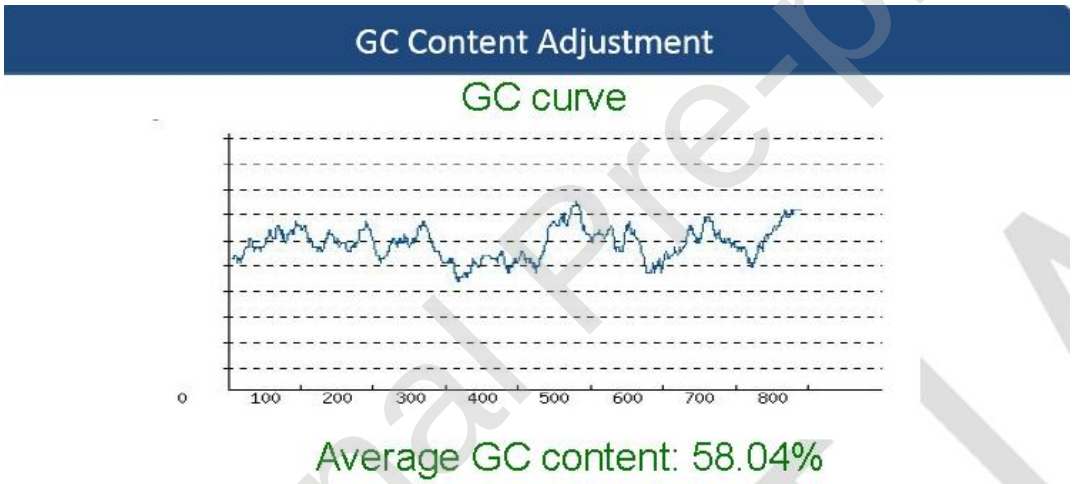


Figure4

a)



b)



c)

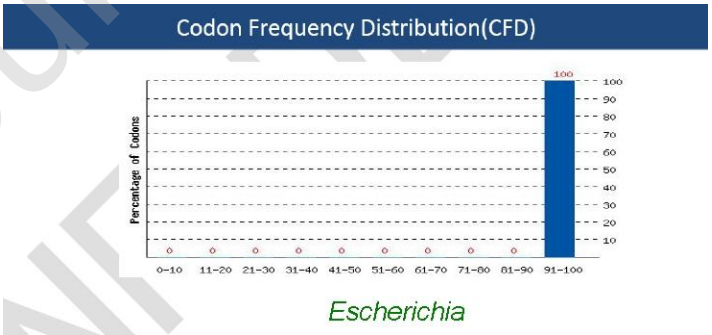


Figure 5

Tables:

Sequence	POS.	MHCI (NET MHC)	MHCI (IEDB)	MHC2	CTL	Linear B Cell	IFN-γ (potentiality/valu)	Allergen FP	Vaxijen
YVARTNIYYHAGTSRLLAVGHPYFPIK	53-79	A0101, A2601,B3901	A2902, B3501, A0101,B3901	DR15 (DRB1*1501), DR4, DR11 (DRB1*1101), DR1 (DRB1*0101), DR4(DRB1*0401), DR51(DRB5*0101)	*	ARTNIYYHAGTSRLL	3.7444874	Non-allergen	Antigen
YGDSLFFYLRREQMFVRH	268-285	B0702	C1402	DR51(DRB5*0101)	*		1	Non-Allergen	Antigen
PLKKYTFWEVNLKEK	470-478	A0301	A1101, A6801	DR8(DRB1*0801)	*	PLKKYTFWEVNLKEK	-0.54339563	Allergen	Antigen
FYLRREQMF	274-282	A2402	A2301, C1402	DR1, DR4(DRB1*0402)			-0.59288568	Allergen	Antigen
YYHAGTSRL	60-68	A2402, B3901	C1402, G0101, G0102, G0103, G0104, G0106	DR1, DR1 (DRB1*0101), DR-2, DR4(DRB1*0401), DR7(DRB1*0701)			0.24105226	Allergen	Non-antigen
YIKMVSEPY	260-268	A2601, B3901	B1402, B1501	DR1 (DRB1*0101, DR4, DR4(DRB1*0401)	*		-1.1145414	Non allergen	Antigen
SADLDQFPLGRKFLL	488-496	B0702, B3901	B5501, B5601	DR1 (DRB1*0101)	*	SADLDQFPLGRKFLL	-0.023525888	Allergen	Non-antigen

FPTPSGSMV	318-326	B0702	B5501, B5601		*		-0.9576615	Non-allergen	Non-antigen
YHIFFQMSL	21-29	B3901	B3901	DR15 (DRB1*1501), DR4(DRB1*0401), DR4(DRB1*0404)	*		-0.50439068	Allergen	Antigen

Table 1. Selected epitopes from Epitope prediction servers and IFNepitope, ANTIGENpro, Vaccijen 2.0 and AllergenFP v.1.0 server. The bolded sequences were finally chosen for the vaccine design. The IFN-γ results in positive values show an IFN-γ inducing epitope and the higher value means the more potent epitope.

Physicochemical properties	Value
Molecular Weight (Da)	29960.11
Instability index	39.94 Stable
Gravy	-0.470
Aliphatic index	84.96
Theoretical pI	7.97
No. amino acids	278
Total no. of negatively charged residues (Asp+Glu)	23
Total no. of positively charged residues (Arg+Lys)	24
No. of atoms	4178
Solubility / cclomics	25%
Solubility / solpro	0.544158
Antigenicity	
Antigenicity / ANTIGENpro	0.938806 (Antigen)
Antigenicity /vaxijen	0.4283 (Probable ANTIGEN)
Allergenicity / AllergenFP v.1.0	Probable Non-Allergen

Table 2. Evaluation of physicochemical properties structure, antigenicity and allergencity of designed vaccine

Favored region	Allowed region	Outlier region
272 (97.5%)	6 (2.2%)	1 (0.4%)

Table 3. Quality evaluation vaccine structure based on Rampage Ramachandran

	No. residue	Amino acid	Contact no.	Propensity score	Disco Tope score
1	33	SER	8	-2.844	-3.437
2	41	ALA	0	-2.682	-2.374
3	45, 46	ALA, ALA	0.2	-3.347, -3.444	-2.962, -3.278
4	49	ALA	3	-3.491	-3.435
5	53	ARG	8	-2.592	-3.214
6	56	SER	8	-2.265	-2.924
7	58- 60	ILE, LYS, GLY	15, 5, 17	-2.075, -1.487, -1.181	-3.561, -1.891, -3.000
8	62, 63	THR, GLN	7, 14	-0.863, -0.854	-1.569, -2.366
9	65- 67	SER, ARG, ASN	14, 6, 23	-0.424, 1.399, -0.059	-1.985, 0.548, -2.697
10	69, 70	ASN, ASP	8, 22	-0.581, -0.553	-1.434, -3.019
11	72, 73	ILE, SER	16, 12	-1.888, -1.044	-3.511, -2.304
12	76, 77	GLN, THR	8, 24	-1.216, -0.608	-1.996, -3.298
13	79, 80	GLU, GLY	16, 11	-0.612, 0.034	-2.382, -1.234
14	83, 84	ASN, GLU	6, 20	-0.737, 0.487	-1.342, -1.869
15	68- 88	ASN, ASN, ASN	15, 11, 28	-1.392, -0.579, -0.234	-2.957, -1.778, -3.427
16	90, 91	GLN, ARG	7, 17	-1.677, -0.697	-2.289, -2.571
17	94	GLU	9	-0.841	-1.779
18	97, 98	ASN, GLY, THR	6, 15	-1.009, -1.238	-1.583, -2.821
19	101-103	ASN, GLY, THR	12, 3, 4	-1.820, -2.347, -1.836	-2.991, -2.422, -2.085
20	105-119	SER, ASP, SER, ASP, LEU, LYS, SER, ILE, GLN, ASP, GLU, ILE, GLN, GLN, ARG	3, 1, 0, 12, 21, 3, 7, 31, 20, 6, 19, 31, 9, 8, 27	-0.865, -1.739, -0.700, -1.009, -1.284, 0.308, 1.029, 0.180, -0.221, 1.466, 1.510, -0.075,	-1.111, -1.654, -0.619, -2.273, -3.551, -0.073, 0.106, -3.405, -2.495, 0.608, -0.849, -0.245, 1.315, -2.649

				0.893, 2.525, 0.515	
21	121, 122	GLU, GLU	5, 12	1.407, 0.696	0.670, -1.454
22	124, 136	ASP, ARG, VAL, SER, ASN, GLN, THR, GLN, PHE, ASN, GLY,VAL, LYS	13, 6, 24, 26, 8, 9, 30, 5, 24, 8, 0, 21, 14	1.108, 2.607, 1.533, 1.292, 1.620, 1.610, 1.309, 0.655, 0.705, 1.638, 1.279, 0.794, 0.956	-0.515, 1.618, -1.403, -1.846, 0.513, 0.390, -2.291, 0.005, -2.136, 0.530, 1.132, -1.713, -0.764
23	139-143	SER, ASP, GLN, ASP, ASN, GLN	11, 0, 8, 18, 6	0.277, 0.156, -1.154, -0.114, 0.203	-1.020, 0.138, -1.942, -2.171, -0.510
24	145	LYS	6	-0.182	-0.851
25	147-155	GLN,VAL, GLY, ALA, ASN, ASP, GLY, GLU, THR	11, 27, 8, 21, 6, 0, 0, 19, 10	0.666, 0.060, -0.651, -1.417, -0.811, 0.628, 0.977, -0.169, 0.048	-0.675, -3.052, -1.496, -3.669, -1.408, 0.555, 0.865, -2.335, -1.107
26	157	THR	6	-1.058	-1.626
27	159	ASP	12	-1.943	-3.100
28	162, 163	SER, ALA	10, 22	-0.841, 0.181	-1.894, -2.369
29	165, 166	PRO, HIS	29, 13	-0.948, -0.627	-3.484, -2.050
30	172	SER	0	-4.018	-3.556
31	175-178	ALA, ARG, THR, ASN	0, 9, 0, 15	-2.670, -1.018, 0.409, -1.246	-2.363, -1.936, 0.362, -2.828
32	181, 182	TYR, HIS	19, 10	-1.309, -0.939	-3.344, -1.981
33	184	GLY	11	-1.672	-2.745
34	186, 187	SER, ARG	7, 5	-2.216, -2.971	-2.766, -3.204
35	192	GLY	6	-3.027	-3.369

36	206	SER	10	-2.181	-3.080
37	210	TYR	6	-2.898	-3.255
38	213, 214	ARG, GLU	8, 9	-2.540, -3.008,	-3.168, -3.697
39	217	PHE	6	-2.941	-3.293
40	220, 221	HIS, GLY	9, 7	-1.516, -0.724	-2.377, -1.446
41	223-225	SER, THR, ASN	18, 6, 12	-0.623, -0.567, -0.679	-2.621, -1.192, -1.980
42	227, 228	GLY, ASN	9, 6	-0.799, -0.704	-1.742, -1.313
43	231, 232	THR, ASN	6, 8	-1.849, -2.139	-2.326, -2.813
44	234	ASN	11	-2.644	-3.605
45	246	ALA	0	-3.536	-3.129
46	249, 250	VAL, SER	2, 5	-3.791, -3.272	-3.585, -3.471
47	281	ARG	1	-2.879	-2.663

Table 4. Conformational B-cell epitopes identified in the refined 3D structure of the designed vaccine using Disco Tope