International Journal of Peptide Research and Therapeutics
https://doi.org/10.1007/s10989-018-9735-4

Designing a Fusion Protein Vaccine Against HCV: An In Silico Approach

Amir Atapour[1] · Pooneh Mokarram[1,2] · Zohreh MostafaviPour[1] · Seyed Younes Hosseini[3] · Younes Ghasemi[1,4,5] · Shiva Mohammadi[1] · Navid Nezafat[4,5]

Accepted: 9 July 2018

Abstract
Hepatitis C virus (HCV) infection is a major global issue that leads to serious liver disease such as chronic liver inflammation and hepatocellular carcinoma. At present, no approved vaccine is available for control or treatment of HCV infection. Therefore, the development of an efficient vaccine against HCV is an urgent need. Today, designing an effective vaccine against hepatitis C is one of the outmost propriety for researchers. Fusion protein vaccines containing the immunogen proteins and adjuvant molecules are able to stimulate both humoral and cellular responses that are crucial for eradicating HCV infection. Herein, in silico design of fusion forms of vaccine candidates against HCV, including flagellin (fliC) from *Pseudomonas aeruginosa* and NS5B antigen (NT300) from HCV was performed. First, two forms of fusion protein (NT300-fliC and fliC-NT300) were designed and analyzed using different bioinformatics tools. For this aim, the Iterative threading assembly refinement (I-TASSER) server was used for modeling the fusion forms of protein; namely, NT300-fliC and fliC-NT300, then the high-rank 3D model of fusion protein was selected, subsequently various physico-chemical, and structural parameters were examined bioinformatically. After the selection of the best construct (fliC-NT300), the interaction of flagellin part of vaccine with toll-like receptor 5 (TLR5) was evaluated via docking studies. Our results represented that based on data obtained from various servers, and the docking analyses of two constructs, fliC-NT300 fusion form showed better results than NT300-fliC. For this reason, the fliC-NT300 form was selected for further evaluations. In sum, structural and immunological computational studies showed that the fliC-NT300 can be introduced as a prophylactic or therapeutic candidate vaccine against the HCV, after the efficacy of that was confirmed via in vitro and in vivo assays.

Keywords Hepatitis C virus (HCV) · Toll-like receptor 5 (TLR5) · Flagellin · Docking · Immunoinformatics · Vaccine

Introduction

Hepatitis C virus (HCV) is a major global health issue. According to the World Health Organization (WHO) reports, a worldwide prevalence of HCV is estimated to be over 180 million people. HCV is the most important cause of liver diseases that eventually results in chronic liver inflammation, cirrhosis, liver failure, and hepatocellular carcinoma (El-Serag 2012; Mohamed et al. 2015; Petruzziello et al. 2016). Despite progress in standard drug therapy, several challenges such as high cost and side effects of drugs are still present in HCV therapy (Ghasemi et al. 2015). Recently, several prophylactic and therapeutic vaccine candidates, based on peptide, recombinant protein, DNA, vector-based, and virus-like particles, have developed against HCV infection. However, some vaccine candidates have been tested in clinical trials, but they have not been efficient in treatment or prevention of HCV infection. Due to their ineffectiveness

✉ Pooneh Mokarram
mokaram2@gmail.com

✉ Navid Nezafat
n.nezafat@srbiau.ac.ir

1 Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran

2 Department of Biochemistry and Colorectal Research Center, Namazi Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

3 Department of Bacteriology and Virology, Shiraz University of Medical Sciences, Shiraz, Iran

4 Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

5 Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

Published online: 14 July 2018

in stimulating the proper humoral and cellular responses as well as the lack of inciting long-lasting immunity, designing an efficient HCV vaccine is an urgent need (Ghasemi et al. 2015; Li et al. 2015). Accordingly, new vaccine strategies have focused on inducing immune responses against structural and non-structural (NS) HCV proteins of different HCV genotypes (Atapour et al. 2017; Feinstone et al. 2012; Hajizadeh and Mokarram 2013). Because hepatitis C patients have the deficiency in immune responses, using certain molecules with efficient immunostimulatory properties, including an efficient vaccine can be applied to restore proper immune responses (Mosley et al. 2012; Torresi et al. 2011; Xue et al. 2014; Yu and Chiang 2010). In the context of the HCV virulence factors, the virus genome encodes 11 viral proteins (structural and nonstructural); E1, E2, Core, NS3, NS4A are among the most important antigens. The non-structural 5B (NS5B) protein (65 kDa) is a viral protein that has a key role in replication of HCV genome (Barnes et al. 2012; Halliday et al. 2011). Previous studies illustrated that HCV non-structural proteins are able to stimulate an immune system response; moreover, it was demonstrated that NS5b protein is an appropriate target for designing the vaccine (Barnes et al. 2012; Brass et al. 2010; Habersetzer et al. 2011). Today, vaccination strategies rely on the use of safe and effective adjuvants, including bacterial toxins, endogenous immunomodulator like cytokines, bacterial cell wall components, which can stimulate the robust immune responses (Maisonneuve et al. 2014; Sivakumar et al. 2011). Toll-like receptor (TLR) ligands are the main class of natural adjuvant. In brief, TLRs are generally expressed on the surface of immune cells such as macrophages, dendritic cells (DCs) and specific epithelial cells, which can recognize diverse pathogen-associated molecular patterns (PAMPs) (Medzhitov 2001; Suresh and Mosser 2013). PAMPs are evolutionarily conserved structures on pathogens that do not express on host cells and contains lipids, lipoproteins, proteins and nucleic acids. Recognition of PAMPs by TLRs leads to the secretion of immune cytokines, chemokines, interferons (IFNs), and up regulation of co-stimulatory molecules (Olive 2012; Suresh and Mosser 2013). Currently, TLR agonists are promising molecular adjuvant in designing the subunit vaccines to overcome the low immunogenicity of such vaccines. Flagellin, as the natural agonist of TLR5, is highly conserved bacterial protein; it has been confirmed as an effective adjuvant for vaccines and immunotherapy. Binding of flagellin to TLR5, which is present on target of antigen presenting cells (APCs), initiates signaling cascade leading to activation of the host inflammatory responses (Lockner et al. 2015; Pandey 2011; Pérez et al. 2013). It has been reported that different forms of flagellin (native, recombinant and truncated) can be used as an immunological adjuvant or as a carrier in vaccine design (Guo et al. 2017; Lockner et al. 2015; Vassilieva et al. 2011). Because designing a novel vaccine is time-consuming process and requires extensive experimental studies, applying bioinformatics tools helps scientists to confer aforesaid obstacles (Karkhah et al. 2017). In this study, we aim to apply bioinformatics tools for designing a novel recombinant fusion protein as subunit vaccine against hepatitis c infection, which contains the N-terminal region of NS5B antigen (NT300) that is considered as antigenic segment and *Pseudomonas aeruginosa* flagellin (fliC) as an adjuvant part. Therefore, we designed two fusion forms of full sequence of B-type flagellin (fliC) and NT300 by bioinformatics tools. Structural and immunological computational studies represented that the fliC-NT300 is the best vaccine candidate. Subsequently, the interaction between different forms of fusion protein with TLR-5 was investigated by docking method. Our designed construct can be applied as a potential candidate vaccine against HCV, after in vitro and in vivo immunological assays.

Materials and Methods

Sequence Collection and Design of Construct

At the present study, the amino acid sequence of B-type flagellin (fliC) (P72151), NS5B (O39930_9HEPC) was retrieved from Uniprot Knowledgebase data at http://www.Uniprot.org in FASTA format. Additionally, the first 300 amino acid of HCV NS5B protein and the full length of the flagellin sequence were used and then the two types of construct arrangements namely NT300-fliC and fliC-NT300 was designed. The following analyses were performed on aforesaid constructs.

Sequence Analysis

Various physico-chemical parameters such as instability index, in vitro and in vivo half-life, theoretical isoelectric point (pI), amino acid composition, grand average of hydropathicity (GRAVY) and molecular weight (Mw) were investigated by ProtParam online server at http://web.expasy.org/protparam/ (Gasteiger et al. 2005; Jahangiri et al. 2018a).

Structural Analyses of Vaccine Construct

Building 3D Models

Three different servers, including Phyre2, SWISS-MODEL, Galaxy web and I-TASSER were used for 3D structure modeling (Kelley and Sternberg 2009; Roy et al. 2010; Schwede et al. 2003; Shin et al. 2014). According to validation results of models obtained from above-mentioned servers, the I-TASSER 3D models possess the highest scores, for this

reason, I-TASSER server was applied as a final server for modeling our protein construct.

I-TASSER software at http://zhanglab.ccmb.med.umich.edu/I-TASSER/ utilizes multiple-threading alignments and iterative template fragment assembly simulations methods for modeling protein structure (Roy et al. 2010). In order to analyze the accuracy of predicted structures, a confidence score named C-score is assigned based on the quality of the threading alignments and the convergence of the I-TASSER's structural assembly refinement simulations.

Refinement of 3D Structures

The best model that has been created by I-TASSER server was submitted to GalaxyRefine software at http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE in order to refine whole protein (Heo et al. 2013). The GalaxyRefine procedure first reconstructs all side-chains and performs side-chain repacking and following overall structure relaxation (mild and aggressive relaxation) by molecular dynamics simulation.

Validation of 3D Structure Model

In order to identify potential errors in the 3D structures and comparing the model before and after the refinement processes, validation and analysis of the models were performed by ProSA-web at https://prosa.services.came.sbg.ac.at/prosa.php (Wiederstein and Sippl 2007), ERRAT server at http://nihserver.mbi.ucla.edu/ERRATv2/ (Colovos and Yeates 1993), and Ramachandran plot analysis at http://mordred.bioc.cam.ac.uk/rapper/rampage.php (Lovell et al. 2003). ProSA-web calculates an overall quality score for a specific input structure. The ProSA-web z-score indicates overall model quality. Its value is displayed in a plot that contains the Z-scores of experimentally determined structures deposited in PDB. Ramachandran plot (RAMPAGE server) was verified the residue-by-residue stereochemical qualities of models and shows the number of residues in favored, allowed and outer areas. The ERRAT server analyzed the statistics of non-bonded atom–atom interactions in the reported structure (compared to a database of reliable high-resolution structures) (Colovos and Yeates 1993; Lovell et al. 2003; Wiederstein and Sippl 2007).

Analyses of Fusion Protein–TLR5 Docking

For protein–protein interaction of the flagellin segment of our fusion protein as a ligand and the TLR5 as the receptor, SwarmDock server at https://bmm.crick.ac.uk/~svc-bmm-swarmdock/ was exploited (Torchala and Bates 2014). SwarmDock applies the flexible modelling of protein–protein complexes using the SwarmDock algorithm, which includes a normal modes approach. The pipeline includes three consecutive stages: (1) pre-processing and minimization of input structures, (2) docking, using a hybrid particle swarm optimization/local search, (3) minimizing, re-ranking, and clustering of the docked poses.

Refinement of Docked Model

The FireDock web server at http://bioinfo3d.cs.tau.ac.il/FireDock/is the first web server for flexible refinement and scoring of protein–protein docking solutions (Mashiach et al. 2008). The server was applied to refine the best selected ligand/receptor complex structure produced by docking analysis. It includes optimization of side-chain conformations and rigid-body orientation and permits a high-throughput refinement. The method refines each candidate and ranks all the candidates according to the global energy (Mashiach et al. 2008).

Immunoinformatics Analyses of Vaccine Construct

Allergenicity and Antigenicity Evaluation

The allergenicity of the fusion protein was analyzed by the AllerTOP 1.0 and AlgPred servers. The AllerTOP 1.0 server at http://www.pharmfac.net/allertop/ is based on auto cross covariance transformation of protein sequences into uniform equal-length vectors. The proteins are categorized by k-nearest neighbor algorithm (kNN, k = 3) based on a training set including 2210 known allergens from different species and 2210 non-allergens from the same species (Dimitrov et al. 2013). Moreover, the AlgPred web server at http://www.imtech.res.in/raghava/algpred/ was used for allergenicity prediction in six different approaches (Saha and Raghava 2006). Due to high accuracy of hybrid approach (SVMc + IgE epitope + ARPs BLAST + MAST), it was chosen in our study.

The VaxiJen v2.0 and ANTIGENpro servers were used for antigenicity evaluation of final fusion protein. The VaxiJen v2.0 server at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html is based on alignment-independent prediction of protective antigens and classifies antigens solely based on the physicochemical properties of proteins without recourse to sequence alignment (Doytchinova and Flower 2007). According to the target organisms (bacterial, viral, and tumor protein datasets), server accuracy varies between 70 and 89% at the default threshold. ANTIGENpro at http://scratch.proteomics.ics.uci.edu/ is a sequence-based, alignment-free, and pathogen independent, applying protein antigenicity microarray data for predicting of protein antigenicity (Magnan et al. 2010).

B-Cell and CTL Epitopes Prediction

ElliPro and BCPRED servers were used to identify conformational and linear B-cell epitopes from the final construct, respectively.

ElliPro is a structure-based server for the prediction of antibody epitopes, which is based on three algorithms: (1) approximation of the protein shape as an ellipsoid, (2) calculation of the residue protrusion index (PI), and (3) clustering of neighboring residues based on their PI values (Ponomarenko et al. 2008). The BCPRED at http://ailab.ist.psu.edu/bcpred/predict.html which predicts linear B-cell epitopes applying a novel method of a subsequence kernel with 74.57% accuracy was used in this study. Moreover, the 20-mer B-cell epitopes with the default specificity threshold (75%) was chosen for prediction (EL-Manzalawy et al. 2008). CTLpred at http://crdd.osdd.net/raghava/ctlpred/ is a method for prediction of CTL epitopes vital in epitope vaccine design. The method is based on elegant machine learning techniques, such as neural network (ANN) and support vector machine (SVM). Five different approaches (quantitative matrices based, only ANN based, only SVM based, consensus approach, combined approach) can be chosen (Bhasin and Raghava 2004). In this study, combined approach was used that has 75.8% accuracy.

Interferon-Gamma Inciting Epitopes Prediction

IFN-gamma is a crucial factor for stimulating cellular immunity. The potential IFN-gamma inciting epitopes were predicted by IFNepitope server at http://crdd.osdd.net/raghava/ifnepitope/ (Dhanda et al. 2013). The server is based on a dataset comprising of IFN-gamma inducing and non-inducing MHC class II binders through various approaches, such as machine learning technique, motifs-based search, and hybrid approach with the 81.39% accuracy.

MHC-I and MHC-II Binding Epitope Prediction

NetMHC 4.0 server at (http://www.cbs.dtu.dk/services/NetMHC/) was applied for the prediction of MHC-I binding epitopes. The NetMHC server predicts the peptide-MHC class I binding peptides using an artificial neural networks (ANNs) method (Andreatta and Nielsen 2015).

Additionally, NetMHC II 2.3 server at (http://www.cbs.dtu.dk/services/NetMHCII/) was used for prediction of MHC-II binding epitopes (Jensen et al. 2018).

In Silico Cloning

In order to clone and express the fusion protein in appropriate expression vector, the reverse translation and codon optimization of protein sequence were simultaneously done by the Codon Usage Wrangler server (http://www.mrc-lmb.cam.ac.uk/ms/methods/codon.html) (Negahdaripour et al. 2017a). The optimized DNA sequence was introduced to the GenScript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis) for assess of various characteristics of sequences, including Codon Adaptation Index (CAI), GC content, and codon frequency distribution (CFD), which have a critical role to reach high-level of protein expression in the host (Mortazavi et al. 2016). The restriction sites were included to N and C-terminals of the sequence. In the current research, the *E. coli* was chosen as the appropriate host.

Results

Physico-Chemical Parameter Evaluation

The Mw and theoretical pI of protein were computed as 82.6 and 8.87 kDa, respectively. The predicted half-lives were calculated as 30 h (mammalian reticulocytes, in vitro), > 20 h (yeast, in vivo), and > 10 h (*E. coli*, in vivo). The instability index was 25.67, which showed the fusion protein is stable (a value above 40 predicts that the protein may be unstable). GRAVY and aliphatic index were defined − 0.181 and 84.53, respectively. The aliphatic index indicates the fusion protein is stable at a wide range of temperatures, and according to the very low GRAVY index (− 0.181), our protein is classified as a hydrophilic protein.

3D Structure Modeling, Refinement and Validation

Different servers were used for 3D structure modeling such as phyre2, SWISS-MODEL, Galaxy web and I-TASSER. Based on results obtained from different servers, we concluded that I-TASSER is proper software for 3D structure modeling of our protein (data not shown). Among the five primary models predicted by I-TASSER software, the model with the highest C-score = − 1.3 were chosen for later refinement process (the C-score range is typically within − 5 to 2, the higher scores shows the higher the confidence). As seen in the below, the validation results of the selected I-TASSER model, shows that the model needs refinement processes. The refinement step was performed by GalaxyRefine server to achieve high-quality 3D model. The starting and refined models were compared in Fig. 1.

ProSA z-score of the primary input model was − 3.31 (Fig. 2a), and after refinement was − 4.43 (Fig. 2b). In the ERRAT server, the overall quality factors of initial and final models were 74.75 and 95.52, respectively that are shown in Fig. 3a, b. Before and after refinement processes, in order to validate the 3D models of fusion protein, Ramachandran plot analysis was also carried out on models. In the initial model,

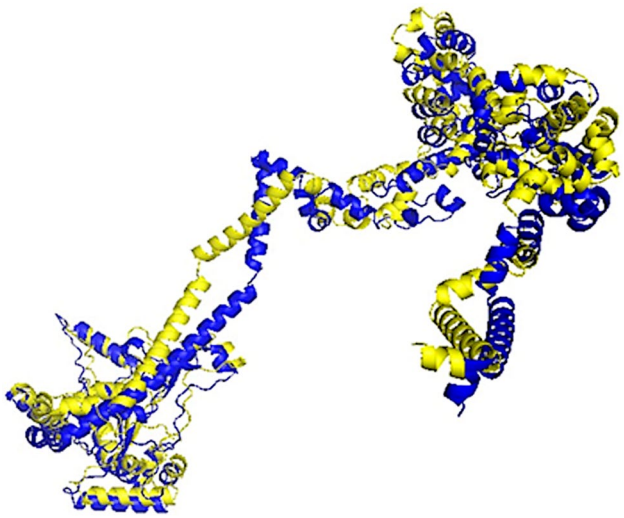

Fig. 1 Superimposition of initial and refined 3D proteins structure conformations. The initial and refined models are colored in yellow and blue, respectively. (Color figure online)

74.5, 14.3 and 11.2% of residues are located in favored, allowed and outlier regions, respectively (Fig. 4a), whereas, in the refined model, 98.6, 1.1, and 0.3% of residues are located in favored, allowed and outlier regions, respectively (Fig. 4b).

Defining Binding Sites and Docking Studies

Based on the biochemical characteristics and previous knowledge on the binding regions between TLR5 and fusion protein [D1 domain of *P. aeruginosa* flagellin (pafliC) and leucin-rich repeats (LLRs) of TLR5], the molecular docking process was performed (Fig. 5). Docking analysis of the 3D model (fliC-NT300) was done by SwarmDock server. The four docked models were chosen and then, in order to select the best model based on global energy; these models were evaluated by FireDock software.

The energy level of the four models were, −32.5, −95.6, −34.5, and −66.0 kcal/mol. The refined model with the global energy −95.6 kcal/mol was chosen as the best model.

Immunoinformatics Analyses

Antigenicity and Allergenicity Evaluation

The antigenicity of fusion protein was predicted 0.9446% by ANTIGENpro and 0.6906% by Vaxijen at 0.4% threshold for bacterial model. The obtained results demonstrated that with a high probability, our fusion protein construct is antigenic. Moreover, the allergenicity of fusion protein was evaluated using AlgPred and AllerTOP 1.0 servers. Based on AllerTope and AlgPred results, the protein is probably a non-allergen.

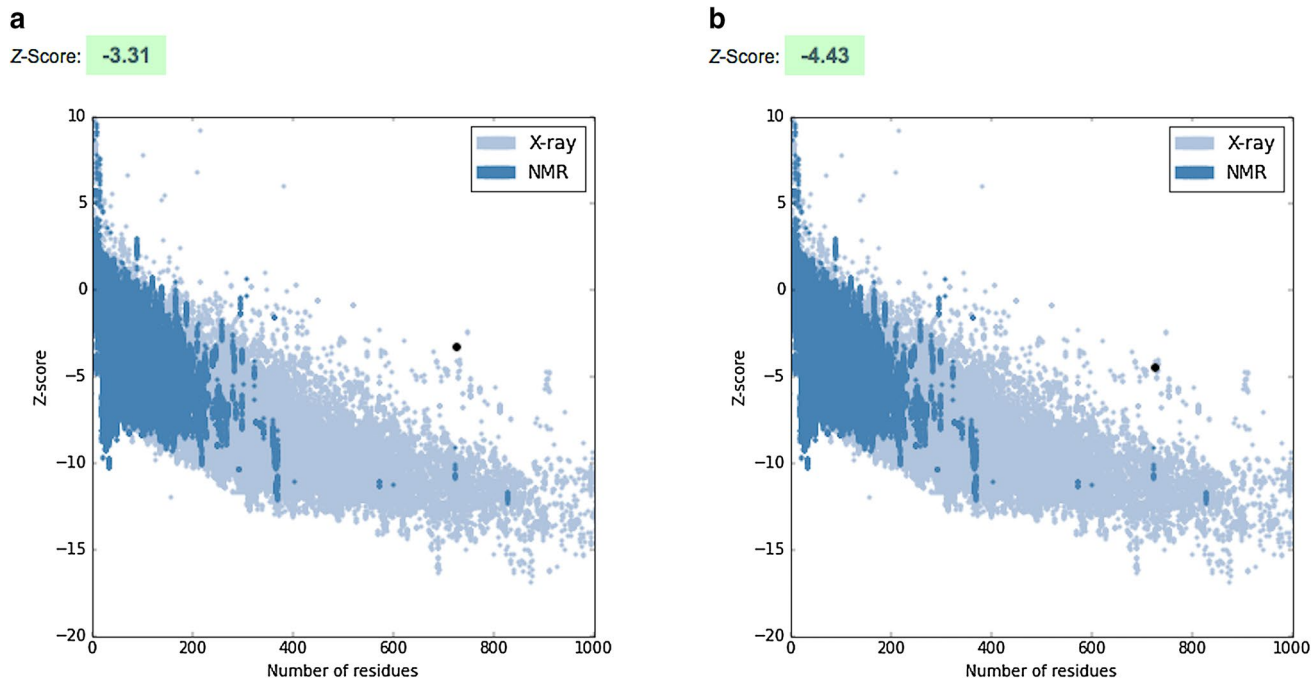

Fig. 2 ProSA-web Z-score plot. **a** The Z-score of the model before refinement is −3.31 (shown in a large black spot). **b** The Z-score of the model after refinement is −4.43

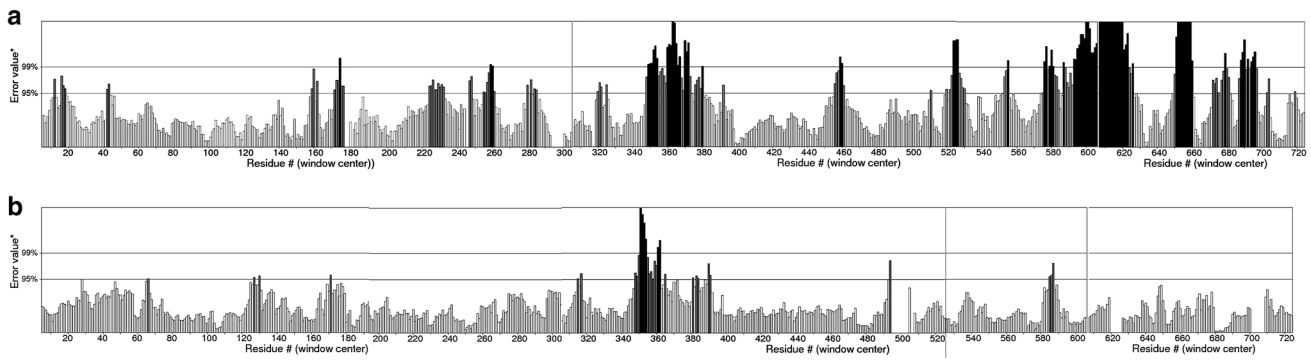

Fig. 3 ERRAT plots. **a** The overall quality factor of the initial structure is 74.750. **b** The overall quality factor of the refined structure is 95.527

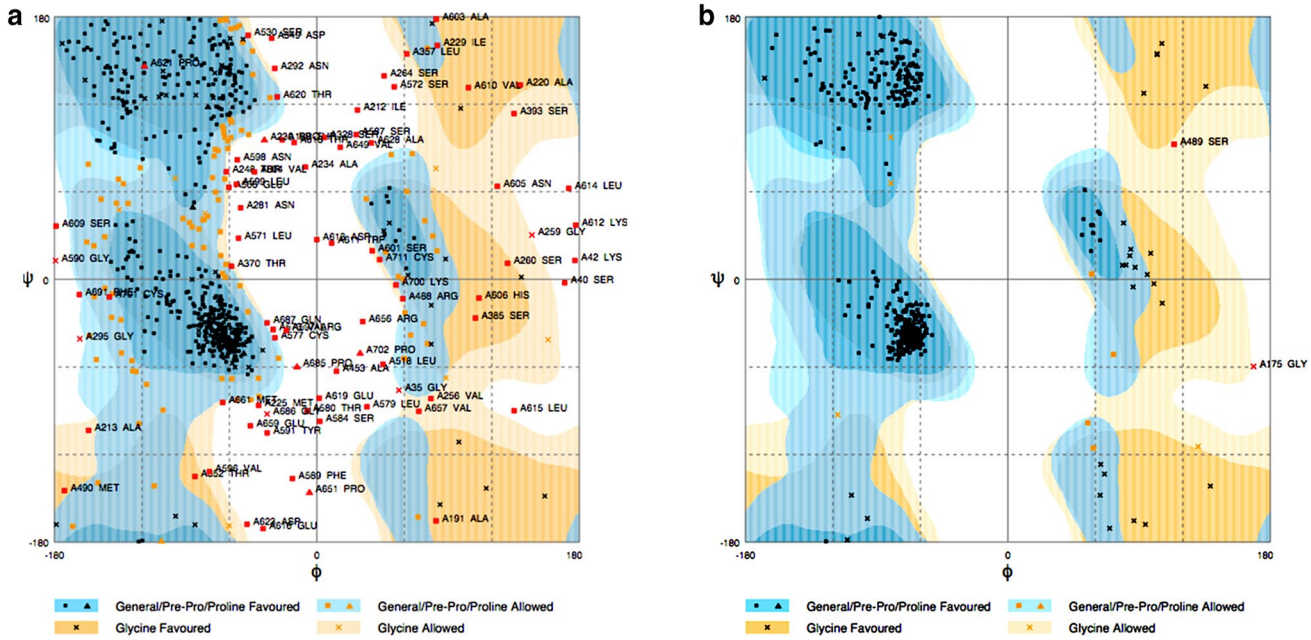

Fig. 4 Ramachandran plots. **a** The plot of the initial model. **b** The plot of the refined model

B-Cell and CTL Epitopes Prediction

Applying ElliPro, 366 conformational B-cell epitopes candidate were determined in the final 3D model of protein out of a total of 788 residues (Table 1). The ten high-ranked and common linear and conformational B-cell epitopes that had been predicted by both ElliPro and BCPRED were selected as final epitopes (Table 2). The ten high-scored CTL epitopes were identified in protein vaccine by CTLpred, as listed in Table 3.

IFN-Gamma Inducing Epitopes Prediction

According to IFNepitope results, ten regions were identified as IFNγ inducing epitopes in the vaccine construct (Table 4).

MHC-I and II Peptide Prediction

Since most HLA molecules have a strong preference for binding to 9-mer epitopes; peptides with 9-mer length and human HLAs were selected. Subsequently, high rank binding peptides were chosen (Table 5).

Additionally, the DRB1_1101, DRB1_0301 and DRB1_1001 alleles were chosen for predicting MHC-II

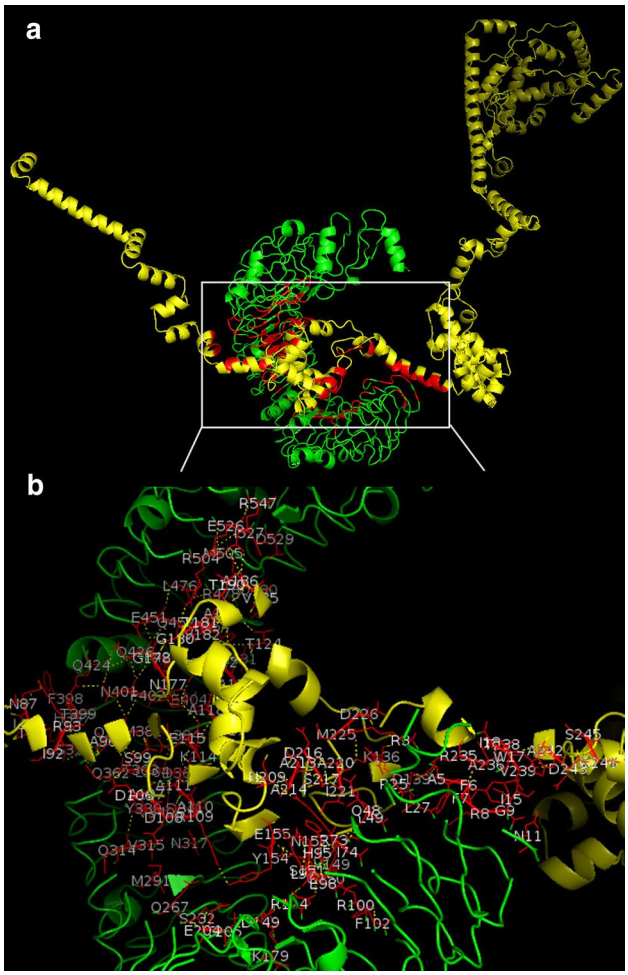

Fig. 5 Docking model of the mouse TLR5 molecule and fusion protein obtained by the Swarmdock server. **a** TLR5 molecule and fusion protein are shown in green and yellow, respectively. **b** The interacting residues, which indicate involvement of the D1 domain of flagellin in the interaction with TLR5, are shown in red. Pymol software was used for visualization of Docked model. (Color figure online)

binding epitopes. The obtained results are shown in the Table 6.

In Silico Cloning

Reverse translation and codon optimization of fusion protein were simultaneously performed by Codon Usage Wrangler server, in order to high-level protein expression in *E. coli*. A CAI of our optimized nucleotide sequence was 1; a CAI of > 0.8 is remarked as good for expression in the selected host (Fig. 6a). Likely codons with a frequency distribution (CFD) < 30% to reduce efficiency of transcriptional and translational. The CFD value of our gene sequence is 100% value, which indicated maximum protein expression in the host (Fig. 6b).

CG content between 30 and 70% was considered as the optimum range. Any regions outside of this range have a negative impact on the efficiency of gene expression. The mean CG content of our sequence is 60.88% (Fig. 7). Moreover, none of the negative repetitive elements in the optimized gene were found. Finally, in order to clone the gene in *E. coli* vectors, the restriction sites *Nco*I and *Xho*I were introduced to the N and C-terminals of the sequence, respectively.

Discussion

Currently, vaccination is a very effective way to combat many infection diseases and save millions of lives. Different vaccine candidates, such as recombinant protein, peptide, DNA, DC-based, and vector-based vaccines have developed against HCV infection; each of them has some advantages and disadvantages (Ghasemi et al. 2015; Naderi et al. 2014). In order to control HCV infection, it is essential to stimulate both humoral and particularly cellular immune responses. Until now, no approved HCV vaccine is available on the market, in this regard, designing an efficient vaccine is a high-priority need. Vaccine production is a time-consuming and costly process. According to recent advances in immunobiology and computer sciences, researchers are able to design the efficient vaccines, particularly in the field of infectious diseases; via computational methods (Farhadi et al. 2015; Hajighahramani et al. 2017; Jahangiri et al. 2018a, b; Shahbazi et al. 2016). In the present study, we intend to use bioinformatics and structural vaccinology approaches to design a fusion protein vaccine consists of *P. aeruginosa* fliC (pafliC) and NS5B protein. Different studies show that fliC has four major domains (D0–D3). The amino acid sequence alignment indicated that the N- and C-terminal regions in flagellin proteins are highly conserved, which is related D0 and D1 domains (Faezi et al. 2016). On the other hand, domains D2 and D3 are composed of hypervariable regions as a folded β-sheet structure. D1 domain organizes packed α-helical structures and includes highly conserved regions. Various studies confirmed that TLR-5 recognition site located within D1 domain that is essential for TLR5–fliC interaction, leading to flagellin signaling and eventually cytokines secretion (Faezi et al. 2016; Honko and Mizel 2005; Song and Yoon 2014). According to the in silico study, Smith and colleagues showed that a short sequence of ten residues (88–97; LQRIRDLALQ) in the D1 domain of fliC has a significant role in TLR-5–fliC interaction (Smith et al. 2003; Verma et al. 2005). TLR-5 as a type I receptor contain three domains, N-terminal extracellular, a single–pass transmembrane, and C-terminal intracellular. The extracellular domain includes LLRs that have a main role in interaction with flagellin

Table 1 Conformational B-cell epitope residues of protein vaccine identified using the ElliPro server

Peptide regions and residues number	Number of residues	ElliPro score
M1, A2, L3, T4, V5, N6, T7, N8, I9, A10, S11, L12, N13,T14, A:Q15, R16, N17, L18, N19, A20, S21, S22, N23, D24, L25, N26, T27, S28, L29, Q30, R31, L32, T33, T34, G35, Y36, I38, N39, S40, A41, K42, D43, D44, A45, A46, G47, L48, Q49, I50, S51, N52, R53, A:L54, S55, N56, Q57, I58, S59, G60, L61, N62, V63, A64, T65, R66, N67, A68, N69, D70, G71, I72, S73, L74, A75, Q76, T77, A78, E79, G80, A81, A:L82, Q83, Q84, S85, T86, N87, I88, L89, Q90, R91, I92, R93, D94, L95	94	0.881
T241, A242, D243, V244, S245, G246, V247, T248, G249, G250, S251, L252, N253, F254, D255, V256, T257, V258, G259, S260, N261, T262, V263, S264, L265, A266, G267, V268, T269, S270, T271, Q272, D273, L274, A275, D276, Q277, L278, N279, S280, N281, S282, S283, K284, L285, G286, I287, T288, A289, S290, I291, N292, D293, K294, G295, V296, L297, T298, I299, T300, S301, A302, T303, G304, E305, N306, V307, K308, F309, A318, G319, Q320, V321, A322, V323, K324, V325, Q326, G327, S328, D329, G330, K331, F332, E333, A334, A335	87	0.759
D617, T618, E619, T620, P621, I622, D623, T624	8	0.711
R544, A:L545, Q546, V547, L548, D549, D550, H551, Y552, R553, D554, V555, L556, K557, E558, M559, K560, A561, K562, A563, S564, T565, V566, K567, A568, K569, L570, L571, S572, V573, E574, E575, A576, C577, K578, L579, T580, P581, P582, H583, S584, A585, K586, S587, K588, F589, G590, Y591, G592, A593, K594, D595, V596, R597, N598, L599, S600, S601, K602, A603, V604, N605, H606, I607,R608, S609, V610, W611, K612, D613, L614,L615, E616, P651, D652, L653, G654, V655, R656, V657, C658, E659, K660,M661, A662, L663, Y664, D665, V666, V667, S668, T669, L670, P671, Q672, A673,V674, M675, G676, S677, S678, Y679, Q682, Y683, S684, Q687, E690, T715, V716, T717, E718, N719, D720,:I721, R722, V723, E724, E725, S726, I727, Y728	121	0.675
F691, L692, V693, N694, T695, W696, K697, S698, K699, K700, C701, P702, M703, G704, F705, S706	16	0.647
A471, I472, L473, A474, Q475, A476, N477, Q478, L479, P480, Q481, A482, V483, L484, S485, L486, L487, R488, S489, M490, S491, Y492, T493, W494, T541, F542, D543	27	0.647
T495, G496, A497, L498	4	0.52
V202, G203, G204, G205, Q206, K208, N209, I210, I212	9	0.516

Table 2 The common and high-ranked conformational and linear B-cell epitopes in the final vaccine that were predicted by ElliPro and BCPRED servers

Start–end	Peptides	Number of residues	Score
630–649	NEVFCVQPEKGGRKPARLIV	20	0.999
576–595	ACKLTPPHSAKSKFGYGAKD	20	0.95
692–108	LVNTWKSKKCPMGFSYD	17	0.919
715–728	TVTENDIRVEESIY	14	0.91
300–308	TSATGENVK	9	0.908
246–265	GVTGGSLNFDVTVGSNTVSL	20	0.879
202–205	VGGG	4	0.766

Table 3 CTL epitopes predicted by CTLpred server

Peptide rank	Start–end position	Sequence	Score (ANN/SVM)
1	656–664	RVCEKMALY	0.92/1.0005481
2	641–649	GRKPARLIV	0.89/0.94787475
3	81–89	ALQQSTNIL	0.94/0.72832316
4	558–566	EMKAKASTV	0.96/0.69006745
5	231–239	NLSARARTV	0.02/1.5824666
6	400–408	NAIAVVDNA	0.96/0.60916048
7	681–689	FQYSPGQRV	0.95/0.57056709
8	234–242	ARARTVFTA	0.00/1.5109405
9	371–379	QASQVFGNA	0.84/0.63708378
10	482–490	AVLSLLRSM	0.85/0.62336213

(Godfroy III et al. 2012; Matsushima et al. 2015; Song et al. 2017; Yoon et al. 2012). The comparative structural analysis by crystallography demonstrated that in fliC D1 domain, amino acids R90 and E114, are located in the center of the interaction site, as a common hot spot, that prepares shape and chemical complementarity to a cavity created by the LLR9 in TLR5 (Smith et al. 2003; Yoon et al. 2012). Different bioinformatics studies have been conducted along with in vitro and in vivo studies on the importance of flagellin as a component of the vaccine against various diseases. In several studies, some of the features of fusion molecules have been evaluated by bioinformatics methods, including interactions between various forms of flagellin and TLR5 (Campodónico et al. 2010; Savar and Bouzari 2014; Savar et al. 2013). In our study, to reach the proper arrangement form of flagellin and NS5B protein, after the initial design of fusion proteins, the amino acid sequences of two forms of proteins were introduced to I-TASSER server for generating 3D models of designed proteins namely, fliC-NT300 and NT300-fliC.

Table 4 IFN-gamma inducing epitopes that were predicted by IFNepitope sever

Start–end position	Sequence	Score
32–146	LTTGYRINSAKDDAAGLQISNRLSNQISGLNVATRNANDGISLAQTAEGALQQSTNILQRIRDLALQSANGSNSDADRAALQKEVAAQQAELTRISDTTTFGGRKLLDGSFGTTS	1
157–173	DISLQNASASAIGSYQ	0.73
181–197	TVASVAGTATASGIASG	0.7
327–342	GSDGKFEAAAKNVVAA	0.99
341–356	AAGTAATTTIVTGYVQ	4
441–485	TNARSRIKDTDFAAETAALSKNQVLQQAGTAILAQANQLPQAVLS	1
522–540	HNMVYATTSRSASLRQKKV	1
555–569	VLKEMKAKASTVKAK	0.29
631–645	EVFCVQPEKGGRKPA	0.33
779–788	NTLTCYLKAS	0.24

According to the results obtained from two fusions, finally, fliC-NT300 form was selected as the appropriate form, then the 3D model structure of this form was refined via GalaxyRefine software. After the refinement process, the refined model was submitted to the SwaramDock software for the molecular docking analysis (Negahdaripour et al. 2017a; Nezafat et al. 2016, 2017). Furthermore, the quality of primary and refined chimeric protein models were determined by ERRAT, Ramachandran plot and ProSA-Web tools (Negahdaripour et al. 2017a; Nezafat et al. 2017). The results showed that the quality of the final refined model has been improved considerably. Docking analysis between fliC regions of chimeric protein and TLR5 was performed via the Swarmdock server. The final docked model of the protein complex was submitted to Firedock server for refinement process (Jahangiri et al. 2018a; Negahdaripour et al. 2017b). The Fire Dock server classifies the models based on the global energy level, so that the model with lower energy levels was selected as the best docked model (−95.6 kcal/mol). Moreover, immunoinformatics analyses of fliC-NT300 indicated that the vaccine construct is a robust antigen, and non-allergen. 366 potential conformational B-cell epitopes, ten IFN-gamma inducing epitopes, ten high-ranked linear B-cell epitopes and CTL epitopes, presented our vaccine is able to incite humoral and cellular responses efficiently, for this reason it can potentially be used as a prophylactic and therapeutic vaccine against HCV. Based on previous studies, it has been shown that the following MHC-II alleles: DRB1_1101, DRB1_0301, DRB1_1001, and MHC-I allele (HLA-A0201) play a significant role in the clearance of HCV. In this regard, both NetMHC and NetMHC II servers were used to obtain immunodominant epitopes (Huang et al. 2016; Ocal et al. 2014).

Finally, in silico cloning process of the chimeric protein was performed for efficient protein expression in the *E. coli* host. In order to achieve a high-level protein

Table 5 MHC-I binding epitopes predicted by the NetMHC 4.0 server

Position	HLA	Peptide sequence	Binding level	Affinity (nM)	%Rank
661	HLA-A0201 (human)	ALYDVVSKL	SB[a]	13.55	0.17
223	HLA-A0201	KMDGAIPNL	SB	13.77	0.17
135	HLA-A0201	KLLDGSFGT	SB	14.11	0.17
613	HLA-A0201	LLEDSVTPI	SB	33.01	0.50

SB strong binder

Table 6 MHC-II binding epitopes predicted by the NetMHC II 2.3 server

Position	HLA	Peptide sequence	Binding level	Affinity (nM)	% Rank
1	DRB1_1101	QAVLSLLRSMSYSWT	SB	15.3	1.50
2	DRB1_1101	PQAVLSLLRSMSYSW	SB	16.8	1.70
3	DRB1_1101	AVLSLLRSMSYSWTG	SB	19.0	1.90
1	DRB1_0301	RLIVFPDLGVRVCEK	SB	15.6	0.50
2	DRB1_0301	ARLIVFPDLGVRVCE	SB	16.7	0.50
3	DRB1_0301	PARLIVFPDLGVRVC	SB	21.3	0.80
4	DRB1_0301	LIVFPDLGVRVCEKM	SB	26.8	1.10
1	DRB1_1001	QAVLSLLRSMSYSWT	SB	8.3	0.60
2	DRB1_1001	PQAVLSLLRSMSYSW	SB	8.6	0.60
3	DRB1_1001	AVLSLLRSMSYSWTG	SB	9.1	0.70

SB strong binder

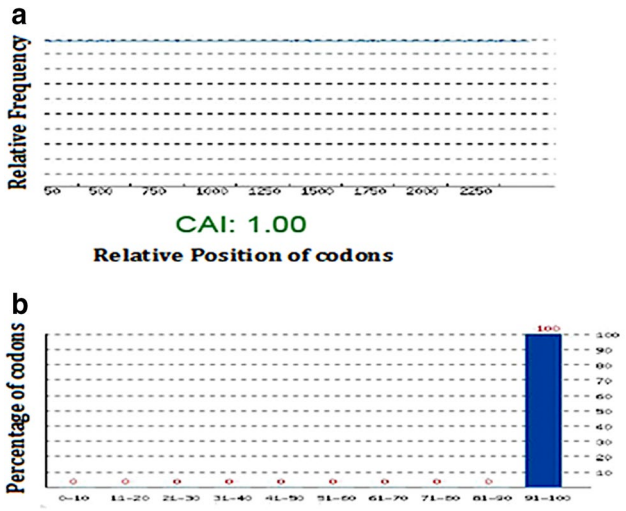

Fig. 6 Analysis the two important values of codon-optimized gene for maximum protein expression in *E. coli*. **a** The CAI value of gene sequence is 100; the CAI = 1 is believed to be perfect for expression in desired host and the CAI > 0.8 is good for high-level expression. **b** The codon frequency distribution (CDF) of DNA sequence; 100% codons have a CFD value more than 90% in gene sequence; codons with values lower than 30 seem to hamper expression efficiency

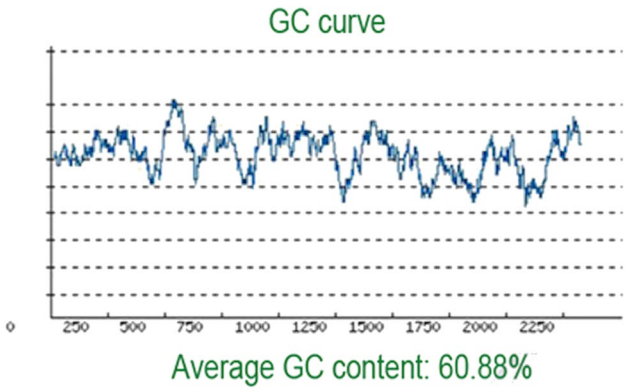

Fig. 7 GC content of the optimized fusion gene. Almost all the peaks are located in the allowed region (30–70%)

expression in *E. coli*, various parameters like CAI, CFD, and GC content of the gene should be optimized. All the above-mentioned indices of the optimized gene showed that our synthetic chimeric construct can be well expressed in the *E. coli* host.

Conclusion

In this study, two different fusion proteins that consist of flagellin (fliC) from *P. aeruginosa* and N-terminal of NS5B from HCV (NT300-fliC and fliC-NT300) were designed and evaluated using various bioinformatics methods. The results of the in silico structural and immunological evaluations showed that fliC-NT300 fusion was the best form and can be used as preventive and therapeutic vaccine for further experimental studies. If the results of experimental studies are acceptable, it can be applied as a candidate vaccine against hepatitis C, in the future.

Acknowledgements The present article was extracted from the PhD thesis written by Amir Atapour and was financially supported by Shiraz University of Medical Sciences Grant Number 93-7369.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

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

Atapour A, Mokarram P, Mostafavi-Pour Z, Ramezani A (2017) Molecular cloning, expression, and purification of a recombinant fusion protein (rNT-gp96-NT300). BioPharm Int 30:38–44

Barnes E et al (2012) Novel adenovirus-based vaccines induce broad and sustained T cell responses to HCV in man. Sci Transl Med 4:115ra111–115ra111

Bhasin M, Raghava G (2004) Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine 22:3195–3204

Brass V, Gouttenoire J, Wahl A, Pal Z, Blum HE, Penin F, Moradpour D (2010) Hepatitis C virus RNA replication requires a conserved structural motif within the transmembrane domain of the NS5B RNA-dependent RNA polymerase. J Virol 84:11580–11584

Campodónico VL, Llosa NJ, Grout M, Döring G, Maira-Litrán T, Pier GB (2010) Evaluation of flagella and flagellin of *Pseudomonas aeruginosa* as vaccines. Infect Immun 78:746–755

Colovos C, Yeates T (1993) ERRAT: an empirical atom-based method for validating protein structures. Protein Sci 2:1511–1519

Dhanda SK, Vir P, Raghava GP (2013) Designing of interferon-gamma inducing MHC class-II binders. Biol Direct 8:30

Dimitrov I, Flower DR, Doytchinova I (2013) AllerTOP-a server for in silico prediction of allergens. BMC Bioinform 14:S4

Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8:4

EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21:243–255

El-Serag HB (2012) Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology 142:1264–1273

Faezi S, Bahrmand AR, Mahdavi M, Siadat SD, Nikokar I, Sardari S, Holder IA (2016) High yield overexpression, refolding, purification and characterization of *Pseudomonas aeruginosa* type B-flagellin: an improved method without sonication. Int J Mol Cell Med 5:37

Farhadi T, Nezafat N, Ghasemi Y, Karimi Z, Hemmati S, Erfani N (2015) Designing of complex multi-epitope peptide vaccine based on omps of *Klebsiella pneumoniae*. Int J Pept Res Ther 21:325–341

Feinstone SM, Hu DJ, Major ME (2012) Prospects for prophylactic and therapeutic vaccines against hepatitis C virus. Clin Infect Dis 55:S25–S32

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. Springer, New York

Ghasemi F, Rostami S, Meshkat Z (2015) Progress in the development of vaccines for hepatitis C virus infection. World J Gastroenterol 21:11984

Godfroy JI III, Roostan M, Moroz YS, Korendovych IV, Yin H (2012) Isolated Toll-like receptor transmembrane domains are capable of oligomerization. PLoS ONE 7:e48875

Guo F et al (2017) Prompt and robust humoral immunity elicited by a conjugated chimeric malaria antigen with a truncated flagellin. Bioconjug Chem 29:761–770

Habersetzer F et al (2011) A poxvirus vaccine is safe, induces T-cell responses, and decreases viral load in patients with chronic hepatitis C. Gastroenterology 141:890–899

Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y (2017) Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. Infect Genet Evol 48:83–94

Hajizadeh MR, Mokarram P (2013) Recombinant nonstructural 3 protein, rNS3, of hepatitis C virus along with recombinant GP96 induce IL-12, TNFα and α5integrin expression in antigen presenting cells. Hepat Mon 13

Halliday J, Klenerman P, Barnes E (2011) Vaccination for hepatitis C virus: closing in on an evasive target. Expert Rev Vaccines 10:659–672

Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic Acids Res 41:W384–W388

Honko AN, Mizel SB (2005) Effects of flagellin on innate and adaptive immunity. Immunol Res 33:83–101

Huang J et al (2016) The associations of HLA-A* 02: 01 and DRB1* 11: 01 with hepatitis C virus spontaneous clearance are independent of *IL28B* in the chinese population. Sci Rep 6:31485

Jahangiri A, Amani J, Halabian R (2018a) In silico analyses of staphylococcal enterotoxin B as a DNA vaccine for cancer therapy. Int J Pept Res Ther 24:131–142

Jahangiri A, Rasooli I, Owlia P, Fooladi AAI, Salimian J (2018b) Highly conserved exposed immunogenic peptides of Omp34 against *Acinetobacter baumannii*: an innovative approach. J Microbiol Methods 144:79–85

Jensen KK et al (2018) Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 154:394–406

Karkhah A, Saadi M, Nouri HR (2017) In silico analyses of heat shock protein 60 and calreticulin to designing a novel vaccine shifting immune response toward T helper 2 in atherosclerosis. Comput Biol Chem 67:244–254

Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371

Li D, Huang Z, Zhong J (2015) Hepatitis C virus vaccine development: old challenges and new opportunities. Natl Sci Rev 2:285–295

Lockner JW et al (2015) Flagellin as carrier and adjuvant in cocaine vaccine development. Mol Pharm 12:653–662

Lovell SC et al (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins Struct Funct Bioinform 50:437–450

Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P (2010) High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 26:2936–2943

Maisonneuve C, Bertholet S, Philpott DJ, De Gregorio E (2014) Unleashing the potential of NOD-and Toll-like agonists as vaccine adjuvants. Proc Natl Acad Sci USA 111:12294–12299

Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 36:W229–W232

Matsushima N, Miyashita H, Enkhbayar P, Kretsinger RH (2015) Comparative geometrical analysis of leucine-rich repeat structures in the nod-like and toll-like receptors in vertebrate innate immunity. Biomolecules 5:1955–1978

Medzhitov R (2001) Toll-like receptors and innate immunity. Nat Rev Immunol 1:135–145

Mohamed AA, Elbedewy TA, El-Serafy M, El-Toukhy N, Ahmed W, El Din ZA (2015) Hepatitis C virus: a global view. World J Hepatol 7:2676

Mortazavi M, Nezafat N, Negahdaripour M, Gholami A, Torkzadeh-Mahani M, Lotfi S, Ghasemi Y (2016) In silico evaluation of rare codons and their positions in the structure of cytosine deaminase and substrate docking studies. Trends Pharm Sci 2

Mosley RT et al (2012) Structure of hepatitis C virus polymerase in complex with primer-template RNA. J Virol 86:6503–6511

Naderi M, Gholipour N, Zolfaghari MR, Binabaj MM, Moghadam AY, Motalleb G (2014) Hepatitis C virus and vaccine development. Int J Mol Cell Med 3:207

Negahdaripour M et al (2017a) A novel HPV prophylactic peptide vaccine, designed by immunoinformatics and structural vaccinology approaches. Infect Genet Evol 54:402–416

Negahdaripour M, Golkar N, Hajighahramani N, Kianpour S, Nezafat N, Ghasemi Y (2017b) Harnessing self-assembled peptide nanoparticles in epitope vaccine design. Biotechnol Adv 35:575–596

Nezafat N, Karimi Z, Eslami M, Mohkam M, Zandian S, Ghasemi Y (2016) Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches. Comput Biol Chem 62:82–95

Nezafat N, Eslami M, Negahdaripour M, Rahbar MR, Ghasemi Y (2017) Designing an efficient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. Mol BioSyst 13:699–713

Ocal S, Selcuk H, Korkmaz M, Altun R, Yildirim AE, Akbas E (2014) Effect of HLA on hepatitis C virus clearance and persistence in anti-HCV-positive end-stage renal disease patients. Saudi J Gastroenterol 20:175

Olive C (2012) Pattern recognition receptors: sentinels in innate immunity and targets of new vaccine adjuvants. Expert Rev Vaccines 11:237–256

Pandey JP (2011) Comment on "Flagellin as an adjuvant: cellular mechanisms and potential". J Immunol 186:1299–1299

Pérez O et al (2013) Adjuvants are key factors for the development of future vaccines: lessons from the Finlay adjuvant platform. Front Immunol 4

Petruzziello A, Marigliano S, Loquercio G, Cozzolino A, Cacciapuoti C (2016) Global epidemiology of hepatitis C virus infection: an up-date of the distribution and circulation of hepatitis C virus genotypes. World J Gastroenterol 22:7824

Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinform 9:514

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738

Saha S, Raghava G (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34:W202–W209

Savar NS, Bouzari S (2014) In silico study of ligand binding site of toll-like receptor 5. Adv Biomed Res 3

Savar NS, Sardari S, Jahanian-Najafabadi A, Bouzari S (2013) In silico and in vitro studies of truncated forms of flagellin (FliC) of enteroaggregative *Escherichia coli* (EAEC). Mol Inform 32:707–716

Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385

Shahbazi M, Haghkhah M, Rahbar MR, Nezafat N, Ghasemi Y (2016) In silico sub-unit hexavalent peptide vaccine against an *Staphylococcus aureus* biofilm-related infection. Int J Pept Res Ther 22:101–117

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

Sivakumar S, Safhi MM, Kannadasan M, Sukumaran N (2011) Vaccine adjuvants—current status and prospects on controlled release adjuvancity. Saudi Pharm J 19:197–206

Smith KD et al (2003) Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat Immunol 4:1247–1253

Song WS, Yoon S-i (2014) Crystal structure of FliC flagellin from *Pseudomonas aeruginosa* and its implication in TLR5 binding and formation of the flagellar filament. Biochem Biophys Res Commun 444:109–115

Song WS, Jeon YJ, Namgung B, Hong M, Yoon S-i (2017) A conserved TLR5 binding and activation hot spot on flagellin. Sci Rep 7

Suresh R, Mosser DM (2013) Pattern recognition receptors in innate immunity, host defense, and immunopathology. Adv Physiol Educ 37:284–291

Torchala M, Bates PA (2014) Predicting the structure of protein–protein complexes using the SwarmDock web server. In: Protein structure prediction. Springer, New York, pp 181–197

Torresi J, Johnson D, Wedemeyer H (2011) Progress in the development of preventive and therapeutic vaccines for hepatitis C virus. J Hepatol 54:1273–1285

Vassilieva EV et al (2011) Enhanced mucosal immune responses to HIV virus-like particles containing a membrane-anchored adjuvant. mBio 2:e00328-00310

Verma A, Arora SK, Kuravi SK, Ramphal R (2005) Roles of specific amino acids in the N terminus of *Pseudomonas aeruginosa* flagellin and of flagellin glycosylation in the innate immune response. Infect Immun 73:8237–8246

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410

Xue J, Zhu H, Chen Z (2014) Therapeutic vaccines against hepatitis C virus. Infect Genet Evol 22:120–129

Yoon S-i, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA (2012) Structural basis of TLR5-flagellin recognition and signaling. Science 335:859–864

Yu CI, Chiang B-L (2010) A new insight into hepatitis C vaccine development. BioMed Res Int 10:548280–548292