Infection, Genetics and Evolution 58 (2018) 96–109

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Structural vaccinology considerations for *in silico* designing of a multi-epitope vaccine

Manica Negahdaripour[a,b], Navid Nezafat[b], Mahboobeh Eslami[b], Mohammad Bagher Ghoshoon[a,b], Eskandar Shoolian[c,d], Sohrab Najafipour[e], Mohammad Hossein Morowvat[a,b], Ali Dehshahri[a,b], Nasrollah Erfani[f], Younes Ghasemi[a,b,g,*]

[a] Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
[b] Pharmaceutical Sciences Research Center, Shiraz University of Medical Science, Shiraz, Iran
[c] Charité University of Medicine, Campus Research Hause of Clinical Chemistry and Biochemistry, Augustenburger Platz 1, 13353 Berlin, Germany
[d] Biotechnology Incubator Center, Shiraz University of Medical Science, Shiraz, Iran
[e] Microbiology Department, Fasa University of Medical Sciences, Fasa, Iran
[f] Cancer Immunology Group, Shiraz Institute for Cancer Research, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran
[g] Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Keywords:
Multi-epitope vaccine
String-of-beads vaccine
Structural vaccinology
Linker
HPV vaccine
Bioinformatics

ABSTRACT

Multi-epitope peptide vaccines, as a kind of fusion proteins, usually possess a string-of-beads structure, consisting of several peptidic epitopes, probably adjuvants and linkers. Very numerous options are possible in selecting the order of different segments and linkers. Such factors can affect the vaccine efficacy through impacting physicochemical characteristics and protein tertiary structure.

To investigate such relations, eleven different constructs were designed and studied as a multi-epitope prophylaxis vaccine for human papilloma virus (HPV). The vaccine contained two epitopes from the minor protein of virus capsid (L2) of HPV16, two TLR agonists as adjuvants (flagellin and RS09, as TLR5 and TLR4 agonists, respectively), and two universal T-helper epitopes. Since the used TLR4 agonist was inserted in the middle of the construct, its appropriate interaction with the bulky TLR4 was a serious concern. Thus, beyond evaluating the physicochemical properties, secondary and tertiary structures, and conformational B-cell epitopes of the designed constructs, TLR4 agonist exposability was also studied. Besides, the interaction between TLR4 and its agonist was investigated through docking and MD studies.

Consequently, one structure ("D") with proper physicochemical features, a high frequency of conformational B-cell epitopes, and appropriate interactions with TLR4 and TLR5 in docking and MD studies, was selected as a proper candidate.

Accordingly, for *in silico* designing of multi-epitope vaccines, structural concerns should be considered, and the linkers and arrangement of epitopes and adjuvants should be optimized. Considering the diversity of the possible structures, devising computational tools for such investigations would be very valuable.

1. Introduction

Recombinant DNA technology has created a major breakthrough in the pharmaceutical history, enabling production of diverse natural and fusion proteins. Genetic fusion of two or more genes that code for the production of different proteins can lead to the generation of fusion proteins with a variety of applications and functions (Berger et al., 2015). Fusin proteins that are used as drugs usually contain at least one domain with a main therapeutic function, such as binding with the cognate receptor. Other fused parts are added for different reasons including improving the molecule characteristics, such as stability and half-life, or supportive functions, for instance, adjuvants in vaccine. The biodistribution and delivery of the fused segments will be the same (Schmidt, 2013), unless they are degraded enzymatically to pieces. Additionally, novel structures not found in nature could be made through fusion (Yu et al., 2015).

Multi-epitope peptide vaccines, which are constituted from several epitopic regions, are attracting much attention compared to the traditional vaccines due to their advantages, including more safety, higher stability, less allergic and autoimmune responses, and a more

* Corresponding author at: Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, P.O. Box 71345-1583, Shiraz, Iran.
E-mail address: ghasemiy@sums.ac.ir (Y. Ghasemi).

https://doi.org/10.1016/j.meegid.2017.12.008
Received 7 November 2017; Received in revised form 5 December 2017; Accepted 11 December 2017
Available online 16 December 2017
1567-1348/

>A
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGPGPGAKFVAAWTLKAAAGPGPGILMQYIKANSKFIGIPMG
LPQSIALSSMVAQGPGPGAPPHALSAAYTKRASATQLYKTCKQAGTCPPDIIPKVAAYTGGRTGYIPLGTRP
PTATDTLAPVGPGPGSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNL
NSARSRIEDADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>B
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSAKFVAAWTLKAAAGGSILMQYIKANSKFIGIPMGLPQS
IALSSMVAQGGSAPPHALSGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGSTGGRTGYIPLGTRPPTATDT
LAPVGGSSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIED
ADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>C
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGSAKFVAAW
TLKAAAGGSAPPHALSGGSILMQYIKANSKFIGIPMGLPQSIALSSMVAQGGSTGGRTGYIPLGTRPPTATDT
LAPVGGSSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIED
ADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>D
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGSAPPHALSG
GSILMQYIKANSKFIGIPMGLPQSIALSSMVAQGGSAKFVAAWTLKAAAGGSTGGRTGYIPLGTRPPTATDT
LAPVGGSSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIED
ADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>E
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSAKFVAAWTLKAAAGGSILMQYIKANSKFIGIPMGLPQS
IALSSMVAQGGSAPPHALSSSLTKRASATQLYKTCKQAGTCPPDIIPKVSSLTGGRTGYIPLGTRPPTATDTL
APVGGSSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIED
ADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>F
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGSTGGRTGYI
PLGTRPPTATDTLAPVGGGAKFVAAWTLKAAAGGSAPPHALSGGGILMQYIKANSKFIGIPMGLPQSIALSS
MVAQGGGSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIE
DADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>G
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGSTGGRTGYI
PLGTRPPTATDTLAPVGGSAKFVAAWTLKAAAGGSAPPHALSGGSILMQYIKANSKFIGIPMGLPQSIALSS
MVAQGGSSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVTNLNSARSRIE
DADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>H
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGPGPGAKFVAAWTLKAAAGPGPGILMQYIKANSKFIGIPMG
LPQSIALSSMVAQGPGPGAPPHALSGGGSTKRASATQLYKTCKQAGTCPPDIIPKVGGGSTGGRTGYIPLGT
RPPTATDTLAPVGPGPGSTLINEDAAAAKKSTANPLASIDSALSKVDAVRSSLGAIQNRFDSAITNLGNTVT
NLNSARSRIEDADYATEVSNMSKAQILQQAGTSVLAQANQVPQNVLSLL

>I
MAQVINTNSLSLLTQNNLNKSQSSLSSAIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQASRNANDGIS
IAQTTEGALNEINNNLQRVRELSVQATNGTNSDSDLKSIQDEIQQRLEEIDRVSNQTQFNGVKVLSQDNQM
KIQVGANDGETITIDLQKIDVKSLGLDGFNVGPGPGAKFVAAWTLKAAAGPGPGILMQYIKANSKFIGIPMG
LPQSIALSSMVAQGPGPGAPPHALSKKTKRASATQLYKTCKQAGTCPPDIIPKVAAYTGGRTGYIPLGTRPP

Fig. 1. The primary sequences of the eleven designed vaccine candidate constructs - each construct consists of 7 segments, which are attached by linkers: 1 and 2) two epitopes shown in blue, driven from L2 (the minor HPV capsid protein) of HPV16, 3 and 4) the flagellin parts at the N- and C-termini as adjuvant, shown in black, 5) RS09, a TLR4 agonist adjuvant, shown in green, 6) PADRE (Pan HLA-DR reactive epitope), a universal T-helper epitope, shown in yellow, and 7) TpD, another universal T-helper epitope, shown in brown. The linkers are shown in red, which can be different in various constructs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

convenient production (Negahdaripour et al., 2017b). To conquer the major disadvantage of multi-epitope peptide vaccines, which is low immunogenicity, employment of adjuvants seems necessary. Based on the previous studies, incorporation of adjuvant in the vaccine structure, if possible, may lead to stronger immune responses compared to mixing the adjuvant and the vaccine, due to the simultaneous delivery of the adjuvant to the same antigen presenting cells (APC) and receptors (Mizel and Bates, 2010). Therefore, adjuvants can also be included as fused segments in such vaccines. In recent years, computer-aided vaccine design has been employed as a novel approach for designing different multi-epitope vaccines (Hajighahramani et al., 2017; Mahmoodi et al., 2016; Nezafat et al., 2017; Negahdaripour et al., 2017a).

Since multi-epitope vaccines are usually string-of-beads structures, which are constructed from several separate segments joined together directly or through linkers, diverse constructs may be made by changing the order of the segments in addition to using different linkers. Linker usage is usually recommended in development of multi-epitope peptide vaccines to avoid generation of junctional epitopes (neoepitopes) and promote the antigen presentation process (Livingston et al., 2002; Nezafat et al., 2016). Multi-epitope peptide vaccines can be considered as a kind of fusion vaccine (albeit with antigenic properties), for which the general concepts for the construction of fusion proteins may apply. Several important factors should be considered in designing synthetic fusion proteins, mainly the order of the fusing parts, which can affect the activity of the domains and the overall characteristics of the molecule. The other decisive factor is the properties of the employed linkers, such as length, composition, and structure (Yu et al., 2015). The structural stability of a vaccine is considered a critical aspect in its efficacy. Because, proper presentation of antigens, which can efficiently induce the immune system, is strongly dependent on the optimal structural stability of the vaccine construct (Scheiblhofer et al., 2017). Other physicochemical features of the vaccine construct, such as hydrophilicity-hydrophobicity, solubility, and pI, can also be important either for its production or efficacy.

Cervical cancer, caused by human papillomavirus (HPV) infection (Bruni et al., 2016), is one of the most prevalent cancers in women globally (Bruni et al., 2016; Ferlay et al., 2015). HPV16 is the most common cancer-causing type, responsible for about half of the HPV-related cervical cancers (Bruni et al., 2016). HPV carries two late proteins in its capsids, which are targeted for development of prophylactic vaccines: L1 (the major) and L2 (the minor) proteins (Wu et al., 2015). Three approved prophylactic vaccines for HPV: Cervarix® bivalent vaccine (GlaxoSmithKline) as well as Gardasil® 4-valent and 9-valent (Merck & Co. Inc.), are available, which all contain the L1 proteins in the form of virus like particles (VLP) (Panatto et al., 2015). In spite of high efficiency of these vaccines, due to their limitations, including type-restricted protection and affordability issues, L2-based vaccines have been investigated as a promising alternative in recent years. However, low immunogenicity has been identified as the major challenge of the L2-based vaccines. Therefore, in our attempt to develop a prophylactic L2-based HPV vaccine, two toll-like receptor (TLR) agonists (one TLR5 and one TLR4 agonists) and two universal T-helper epitopes were inserted in the construct to improve the vaccine immunogenicity of the two selected epitopic segments of HPV16 L2 (Negahdaripour et al., 2017a). Obviously, the order of these segments and the employed linker could impact the overall characteristics of the designed vaccine. Although a commercial toolkit has been developed for such investigations in vaccine design (Moise et al., 2015), to our knowledge, there is no freely available tool for evaluating and comparing the different possible structures and their suitability as a vaccine candidate. Therefore, in this study, the selection of the best sequence among the designed structures was performed based on a novel structural vaccinology approach for comparison of several generated molecules through bioinformatics tools. Beyond comparing the physicochemical properties, secondary and tertiary structures, and conformational B-cell epitopes, docking and molecular dynamics (MD) simulation of the designed constructs with TLR4 was also studied. The details of the steps undertaken for such evaluations are introduced in this article.

2. Methods

2.1. Fusion vaccine design

Two antigenic epitope segments were selected on the L2 sequence of HPV16 capsid protein (ID no.: P03107 at the UniProt database at http://www.uniprot.org/ (Consortium, 2014)), through extensive immunoinformatics and bioinformatics investigations explained in our previous study (Negahdaripour et al., 2017a). Several adjuvants, including: flagellin (the N-terminal head and C-terminal tail), as a TLR5 agonist, RS09, a short peptide TLR4 agonist, PADRE (Pan HLA-DR reactive epitope), and TpD, as two universal T-helper epitopes were also chosen. Finally, these seven peptide segments were joined together by different linkers, and eleven different constructs were built. The primary structures of these eleven constructs are shown in Fig. 1.

2.2. Structural analyses

2.2.1. Investigation of physicochemical properties

Some physicochemical characteristics of the designed structures, including instability index, grand average of hydropathicity (Gravy), theoretical pI, total no. of negatively charged residues (Asp + Glu), and total no. of positively charged residues (Arg + Lys) were evaluated using the ProtParam tool at http://web.expasy.org/protparam/ (Gasteiger et al., 2005).

Solubility of the different structures were predicted by the *cc*SOL *omics* server (http://service.tartaglialab.com/grant_submission/ccsol_omics), which works according to disorder, coil, and hydrophilicity propensities of the protein (Agostini et al., 2014).

2.2.2. Secondary structure and exposability of amino acids

The secondary structures of the designed sequences were predicted by five famous servers, including SPIDER2 (http://sparks-lab.org/server/SPIDER2/index.php) (Heffernan et al., 2016; Heffernan et al., 2015), PSIPRED v3.3 (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan et al., 2010), PORTER (http://distill.ucd.ie/porter/) (Pollastri and McLysaght, 2005), RaptorX-Property (http://raptorx2.uchicago.edu/StructurePropertyPred/predict/) (Wang et al., 2016a; Wang et al., 2016b), and GOR (Garnier Osguthorpe-Robson) IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html) (Garnier et al., 1996).

The exposability or solvent accessibility (ACC) of TLR4 agonist (RS09) in different designed constructs was predicted by three servers: PaleAle (http://distill.ucd.ie/paleale/) (Mooney and Pollastri, 2009; Pollastri et al., 2007), SPIDER2 (http://sparks-lab.org/server/SPIDER2/index.php) (Heffernan et al., 2016; Heffernan et al., 2015), and the RaptorX-Property server at http://raptorx2.uchicago.edu/StructurePropertyPred/predict/ (Wang et al., 2016a; Wang et al., 2016b).

The SPIDER2 server works based on a deep learning neural network. Using three hidden layers, it achieved the highest accuracy of 81.8% for an independent test on 1199 high-resolution proteins (< 2.0 Å) in CASP11 (Yang et al., 2016; Yang et al., 2017). The SPIDER2 results include relative accessible surface area (rASA) with a correlation coefficient of 0.76 between the server predicted and actual ASA values (Yang et al., 2017). The PSIPRED method uses a two-stage neural network to predict the secondary structure using PSI-BLAST output. In 2010, PSIPRED V3.0 achieved the highest then available three-state accuracy (Q3) of 81.4% (0.6%) for secondary structure prediction through simply updating sequences and structure databases and establishing minor algorithmic updates (Buchan et al., 2010; Yang et al., 2016). The Porter server predicts protein secondary structure using

Table 1
Physicochemical properties of the eleven designed vaccines by ProtParam and ccSOL omics.

Sequence Name	No. of amino acids	MW	Instability index	Gravy	Theoretical pI	Total no. of negatively charged residues	Total no. of positively charged residues	Solubility propensity (%)
A	407	42,421.66	28.99	− 0.256	8.94	31	35	34
B	399	41,556.53	32.56	− 0.252	8.97	31	35	91
C	399	41,556.53	32.56	− 0.252	8.97	31	35	91
D	399	41,556.53	32.56	− 0.252	8.97	31	35	91
E	399	41,728.80	33.60	− 0.233	8.97	31	35	98
F	399	41,466.45	32.58	− 0.249	8.97	31	35	91
G	399	41,556.53	32.56	− 0.252	8.97	31	35	97
H	409	42,327.46	30.31	− 0.275	8.97	31	35	55
I	406	42,372.67	29.25	− 0.281	9.21	31	37	63
J	405	41,718.68	34.36	− 0.248	8.97	31	35	99
K	407	42,413.72	29.20	− 0.238	8.95	31	35	34

Table 2
Secondary structure and exposability of the TLR4 agonist domain (RS09) predicted by several servers.

Sequence name	Relative accessible surface area	Solvent accessibility		Secondary structure				
	SPIDER2	RaptorX-ACC	PaleAle	Porter	RaptorX-SS3	psipred	SPIDER2	GOR IV
A	17	4E/2M/1B	3E/2e/1B/1b	3C/3H/1E	7C	7C	5C/2E	4C/3H
B	19	5E/1M/1B	4e/1B/2b	4C/3H	3C/4E	7C	3C/4H	7C
C	19	5E/2M	5e/2b	2C/5E	7C	7C	4C/3H	7C
D	22	4E/2M/1B	3E/1e/1B/2b	4C/2H/1E	3C/4E	7C	7C	7C
E	22	4E/2M/1B	1B/6b	1C/6H	3C/4E	7C	4C/3H	5C/2H
F	19	6E/1M	1E/2e/2B/2b	2C/5E	6C/1E	7C	5C/2H	7C
G	22	5E/1M/1B	2E/2e/1B/2b	2C/5E	7C	7C	3C/4H	7C
H	17	5E/1M/1B	1e/3b/3B	7E	3C/4E	7C	4C/3H	5C/2E
I	19	6E/1B	1e/2B/4b	3C/4E	7C	7C	5C/2H	4C/3H
J	19	4E/1M/2B	3E/2e/1B/1b	3C/3H/1E	4C/3E	4C/3H	3C/4H	3C/4E
K	18	7E	4e/2B/1b	2C/5H	7C	7C	5C/2E	4C/3H

In the secondary structure results: E = sheet, C = coil, H = helix. In the exposability and solvent accessibility results, RaptorX-Property: E = exposed, M = medium, B = buried. PaleAle: E = completely exposed (50 + % exposed), e = partly exposed (25–50% exposed), b = partly buried (4–25% exposed), B = completely buried (0–4% exposed). SPIDER2: the higher number shows higher exposability. (The number is the sum of numbers the server indicated for each residue.)

Table 3
The C-score of the modeled 3-D structures by I-Tasser, conformational B-cell epitopes, proper docking result of TLR4 agonist with its receptor of the eleven designed vaccines.

Sequence Name	I-tasser C-score values	No. of conformational B cell epitopes/total no. of amino acids
A	0.09	71/407
B	− 0.46	44/399
C	0.06	30/399
D	0.08	64/399
E	0.01	44/399
F	0.08	42/399
G	0.07	39/399
H	0.09	78/409
I	0.10	73/406
J	0.10	47/405
K	0.11	51/407

machine learning systems through homology to proteins of known structure, where available (Pollastri et al., 2007). Using three runs of PSI-BLAST alignment, Porter achieved 79% correct classification on the "hard" CASP 3-class assignment tested by a rigorous 5-fold cross validation procedure (Pollastri and McLysaght, 2005). The PaleAle is developed by the same team who designed the Porter server by using the same architecture. Its accuracy, measured on the same large, non-redundant set adopted to train Porter, exceeds 55% correct in a 4-class classification, and 80% in a 2-class classification (Buried *vs* Exposed, with 25% threshold) (Pollastri et al., 2007). The RaptorX-Property (DeepCNF) server, which predicts structure properties of a protein sequence without using any templates, has outperformed some other servers, especially for proteins without close homologs in PDB. It employs a powerful in-house deep learning model DeepCNF (Deep Convolutional Neural Fields) to predict secondary structure (SS) and solvent accessibility (ACC). This server performs secondary structure prediction through two methods: 3-state secondary structure (SS3) and 8-state secondary structure (SS8). In this study, the SS3 method, which has shown a higher accuracy (~84% Q3 accuracy) than SS8, was used, considering that the three secondary structure classification was enough for our purpose. The GOR IV algorithm works based on both information theory and Bayesian statistics, but not evolutionary information, by employing a sliding window of neighboring residues (Sen et al., 2005).

2.3. Protein modeling and refinement

The 3D structure of the eleven designed vaccine constructs were modeled by the I-Tasser software at http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Yang and Zhang, 2015), ranked at the top sever in the 12th community-wide CASP (critical assessment of methods of protein structure prediction) experiment for automated protein structure prediction. This server goes through several steps: 1) threading (identifying template proteins with similar structure to the query protein sequence from PDB using multiple threading alignment approaches), 2) structural assembly (fragment assembly through a modified replica-exchange Monte Carlo simulation method, in addition to *ab initio* modeling for some regions, mostly loops and tails), 3) model selection and refinement (selection of the model through employing clustering structure decoys and refinement by fragment-guided molecular dynamics simulation (FG-MD) or ModRefiner), and finally 4) structure-based functional annotation by COACH approach. C-score is a number defined by the I-Tasser as a confidence score for estimating the model overall accuracy. Higher value C-scores (C-score is in the range of [5 to 2])

Table 4
The overall results of different steps of investigation obtained for each designed construct.

Sequence Name	ITasser C-score values	Physicochemical properties	Solubility	Coil probability for RS09 based on secondary structure prediction	Exposability of RS09	Accessibility and steric hindrance based on 3D structure	Conformational B cell epitopes	Docking of TLR4 agonist
A	0.09	+	−	26	26	+	+	−
B	− 0.46	+	+	24	28	−	−	−
C	0.06	+	+	29	29	−	−	−
D	0.08	+	+	29	30	+	+	+
E	0.01	+	+	20	26	+	−	±
F	0.08	+	+	27	28	−	−	−
G	0.07	+	+	26	31	−	−	−
H	0.09	+	−	19	23	+	+	±
I	0.10	−	−	26	26	−	+	−
J	0.10	+	+	17	28	−	−	−
K	0.11	+	−	25	29	−	−	−

The positive and negative scores show the overall results obtained by construct in each step. The number in the secondary structure shows the probability for coil structure of RS09 based on the sum of all predictions, with a higher number showing a greater probability. The exposability also is the sum of the results of the three server used for solvent accessibility. " ± " in the docking results show that though there was a close to correct orientation in their docked models, closer observation of the docked models showed that the interactions were not ideal.

indicate models with higher confidence (Yang and Zhang, 2015). The 3D modeled structures were visualized by the Discovery studio 3.5 and the UCSF chimera programs.

2.4. Conformational B-cell epitopes prediction

Once the 3D structure of the protein was modeled, the discontinuous B-cell epitopes were predicted by the DiscoTope 2.0 server (http://www.cbs.dtu.dk/services/DiscoTope/). This server uses two methods for calculation of the final scores: contact numbers derived from surface accessibility and a novel epitope propensity amino acid score (Kringelum et al., 2012). The default threshold (− 3.7) was used, at which the sensitivity = 0.47 and specificity = 0.75.

2.5. Molecular docking studies: investigation of the interaction of TLR4 agonist with the receptor

The interaction of the TLR4 agonist (RS09) in each structure with TLR4 was investigated through docking studies, using the ClusPro server at https://cluspro.org/home.php (Kozakov et al., 2017).

This server performs a rigid-body protein-protein docking. Its fully automated, fast algorithm filters docked conformations with good surface complementarity based on desolvation and electrostatic energies, and ranks them based on their clustering properties. Clustering is used to smooth the local minima and to select the ones with the broadest energy wells-a property associated with the free energy at the binding site (Comeau et al., 2004).

In order to dock RS09 with TLR4, the mouse TLR4 structure was driven from the UniProt database (Consortium, 2014), with ID number of Q9QUK6. Its 3D structure was also obtained from PDBe under code 3fxi.

2.6. Antigenicity and allergenicity evaluation

The best construct (structure "D"), based on the results of docking studies, was analyzed for antigenicity and allergenicity evaluation as was described previously (Negahdaripour et al., 2017a). Briefly, antigenicity analysis was done by the ANTIGENpro server (http://scratch.proteomics.ics.uci.edu/) (Magnan et al., 2010) and the VaxiJen v2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) (Doytchinova and Flower, 2007). The allergenicity evaluation was performed by four servers: the Allerdictor at http://allerdictor.vbi.vt.edu/home/ (Dang and Lawrence, 2014), the PREAL (Prediction of Allergenic Proteins) server at http://gmobl.sjtu.edu.cn/PREAL/index.php (Wang et al., 2013), the AllergenFP v.1.0 at http://ddgpharmfac.net/AllergenFP/ (Dimitrov et al., 2013), and the Algpred at http://www.imtech.res.in/raghava/algpred/ (Saha and Raghava, 2006).

2.7. Refinement and validation of the 3D modeled structure

Structure "D" underwent a refinement process by the GalaxyRefine server at http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE (Shin et al., 2014), as previously described (Negahdaripour et al., 2017a). Then, the 3D refined structures were compared and the best model was validated by the ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php) (Wiederstein and Sippl, 2007), ERRAT (http://services.mbi.ucla.edu/ERRAT/), and RAMPAGE (Ramachandran Plot Assessment) (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (Lovell et al., 2003) servers, as previously explained (Negahdaripour et al., 2017a). The best refined model was selected for the following step.

2.8. MD simulation studies

In order to ensure the binding stability of the designed vaccine to TLR4/MD2 complex, the selected docking model (related to structure "D") was used as the initial configuration in the molecular dynamics simulation (MD) process. MD simulation gives TLR4/MD2-vaccine complex an appropriate opportunity to obtain the best orientation and interactions relative to each other.

All the steps and applied conditions of MD simulation were similar to our previous works (Nezafat et al., 2016; Hajighahramani et al., 2017; Nezafat et al., 2017; Negahdaripour et al., 2017a). GROMACS 5.0.1 was employed to perform MD simulation on TLR4/MD2-vaccine complex for 30 ns.

3. Results and discussion

In recent years, bioinformatics tools have been valuable resources for researchers to save time and energy by performing predictions in different biological fields (Karimi et al., 2015; Negahdaripour et al., 2016; Negahdaripour et al., 2017c; Negahdaripour et al., 2017d; Rahmatabadi et al., 2016; Rahmatabadi et al., 2017), including vaccine design, in which antigenic domains are identified through immunoinformatics tools and linked together to form a multi-epitope vaccine (Negahdaripour et al., 2017b). In multi-epitope vaccines, as string-of-beads structures, not only the nature of the selected domains such as epitopes and adjuvants, but also their order and position in the protein sequence as well as linkers (compositions and length) are important. Because the construct sequence determines the

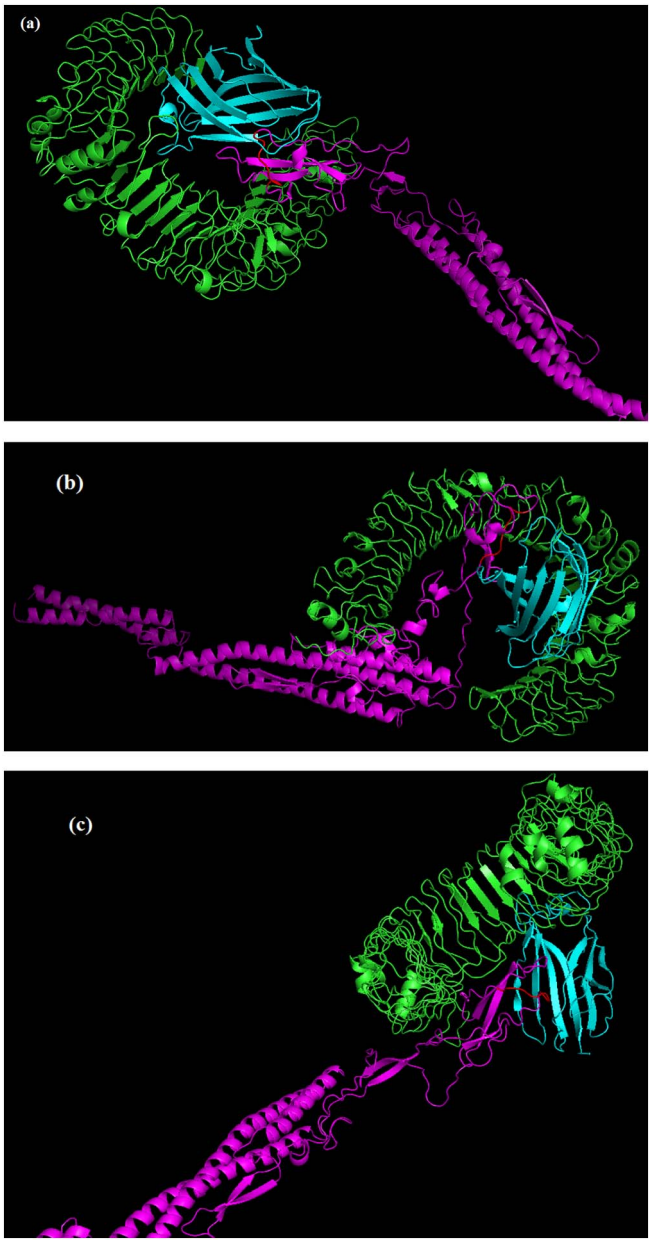

Fig. 2. Docking of structure "D", "H" and "E" with TLR4 - these are the only docked models in which RS09 seemed to interact in an almost proper way with TLR4. TLR4 is shown in green, MD2 in blue, the vaccine construct in magenta, while RS09 is colored in red.
(a) The docked model of structure "E", the part of the vaccine that has entered the MD2 pocket is the adjacent residues to RS09, and the RS09 residues are mainly positioned outside MD2 pocket. (b) The docked "H" structure, the vaccine is not entered as much as structure "D" into MD2 pocket, but has remained close to the MD2 opening. (c) The docked "D" structure, whose engagement with TLR4 seems the best, with RS09 entry into the opening of MD2 pocket. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

physicochemical features and the conformational structure of the designed vaccine and consequently, affects the vaccine efficiency.

In this study, eleven different constructs were built by incorporation of seven peptidic segments (two epitope segments and several adjuvants) (Fig. 1). These eleven structures differed in the order of the segments and the used linkers. The flagellin N-terminal and C-terminal segments that were used as adjuvants were kept as free head and tail segments in all the tested constructs to follow the natural flagellin conformation and allow their convenient interaction with TLR5 (Negahdaripour et al., 2017a). Other segments (antigens and other adjuvants) were located between the head and tail, substituting the deleted parts of the original flagellin. Given the fact that in CTL and linear B-cell epitopes, only the sequence of the epitope is important, there were numerous options for arrangement of the other five segments (except flagellin parts) between the N- and C-termini. Analysis of all possible structures, one by one, was impossible by such an approach, due to the very huge volume of evaluations. Therefore, these eleven structures were designed and studied as samples. The main goal was obtaining a stable protein structure with proper physicochemical properties, and suitable functionality, in regard to induction of proper immune responses. To achieve such an objective, efficient interaction of the two employed TLR agonists with their cognate TLRs would count.

Difference in linker length can impact protein stability, domain-domain orientation, and folding rate (George and Heringa, 2002; Robinson and Sauer, 1998; van Leeuwen et al., 1997). Therefore, several different linkers with various lengths and composition were selected. Linker sequences were chosen based on the linkers reported in the multi-epitope peptide vaccines in the published literature and linker databases. The used linkers included GPGPG (Livingston et al., 2002; Nezafat et al., 2014), AAY, SSL (Levy et al., 2007; Schubert and Kohlbacher, 2016), GGG (Rosenthal et al., 2017), KK (Nezafat et al., 2017; Yano et al., 2005), GGGG, GGS, and GGGS (SynLinker at http://synlinker.syncti.org/index.php) (Klement et al., 2015; Yu et al., 2015). Some rational approaches were followed in their selection. For example, as the pI of the selected antigenic domains were somewhat basic (> 8), linkers that contain basic amino acids would increase the pI, such as KK in structure "I" (Fig. 1). Thus, basic linkers were avoided in other structures in order not to increase the pI of the protein. On this basis, glycine-based linkers were commonly chosen for joining the different domains. Glycine-rich linkers, which are the most frequent used linkers to date, are flexible due to the small size of glycine, usually improve solubility, and are believed to be proteolysis resistant. Accordingly, they may allow the adjoining domains to be accessible and act freely (Kavoosi et al., 2007). All these characteristics seemed appropriate to be used in our designed constructs. To add variety and evaluate more diverse structures, a few other linkers were also employed, including AAY and SSL (Fig. 1). Some tendencies have been mentioned previously for linkers toward a specific immunological response, such as a higher CTL immunogenicity by insertion of basic, amide, or small residues immediately following the C-terminal of the epitope in DNA vaccines (Livingston et al., 2002). Since in our designed vaccine, the epitopes were selected based on the overlapping results of several different servers, and consequently, each selected epitope can play different roles in induction of immune responses (Negahdaripour et al., 2017a), the selection of linker in order to induce a specific immunological reaction could not be a selective strategy here.

To sum for designing these eleven sequences, some criteria were initially set regarding the order of the segments and linker selection, namely the fixed location of flagellin and usage of flexible linkers with a report of efficiency in vaccines. Then, through the performed studies that will be explained in the following, some clues to reach a proper structure were gradually found, which we tried to consider in the rest of the designing process. For instance, using linkers: GPGPG, AAY, and ALL (in structures A, H, I, and K) seemed to cause a reduction in the solubility propensity (as will be discussed later). Therefore, linkers with G and S residues could be preferred. For the proper interaction of RS09 with TLR4, which was a main concern in our designed vaccine, an almost middle location for RS09 was chosen in most structures, in order to obtain a structure with less steric hindrance. However, the 3D structure of the protein and position of RS09 in the cognate conformation as well as its appropriate engagement with TLR4 was not completely predictable, unless the constructs were built and analyzed.

After construction of the sequences, their physicochemical properties were studied and compared. The results of the investigations by Protparam and ccSOL omics are shown in Table 1. The number of amino acids varied in a range of 399–409, based on the selected linkers;

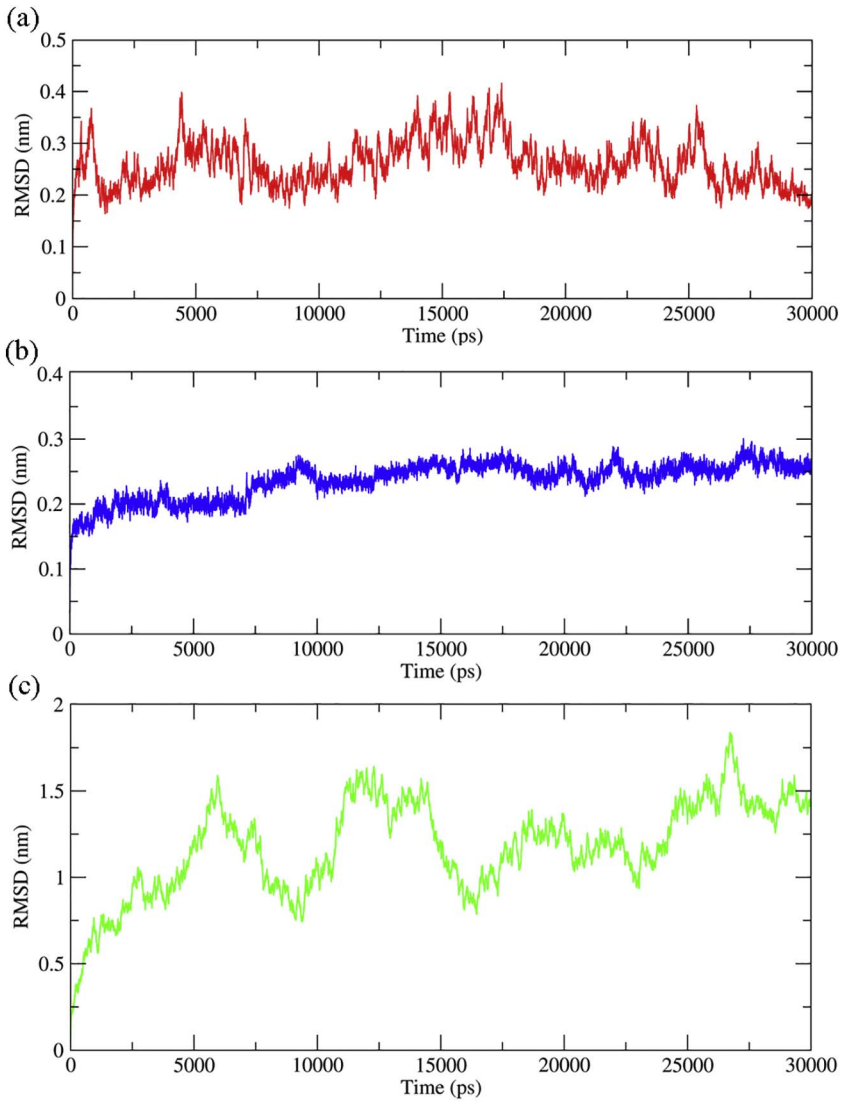

Fig. 3. Backbone RMSD plots of: (a) TLR4 protein, (b) MD2 protein, and (c) vaccine molecule.

thus, the molecular weights were also slightly different. As seen in Table 1, most structures showed similar Gravy values. Gravy is a measure of hydrophobicity or hydrophilicity of the structures (Gasteiger et al., 2005). Gravy value for all structures was negative, representing their slightly hydrophilic nature. Their theoretical pI values were also very close, except for structure "I" ("406"), whose theoretical pI was calculated as 9.21. Since the designed vaccine would be for injection, a closer pI to the natural pH of the blood and body fluids or neutral pH, is preferred. Accordingly, structure "I" was inferior to others in this regard. The predicted instability index showed slight differences among the 11 structures in a range of 28.99 to 34.36. According to the server threshold, an instability index below 40 is indicative of protein stability, and a lower value demonstrates a more stable protein (Gasteiger et al., 2005). In our tested sequences, the variations were not significant, and all the structures were predicted to be stable. The total numbers of negatively charged amino acids were the same (31). The total numbers of positively charged amino acids were 35 in all structures, except structure "I", whose pI also showed the highest value due to the addition of two basic amino acids (KK) as a linker. The predicted solubility propensity of the designed vaccines, by ccSOL omics server, showed variations in a range of 34% (for structure "A" and "K") to 99% (for structure "J"). The ccSOL algorithm predicts protein solubility based on physicochemical properties and discriminates soluble and insoluble proteins obtained from heterologous expression experiments with an accuracy of 78% (Agostini et al., 2014). Considering the fact that a higher solubility can be an advantage, structures "A", "H", "I", and "K", with predicted solubilities below 70%, seemed inappropriate structures. These four structures all contained "GPGPG" linkers. In structure "H" (solubility = 55%), all the linkers were "GPGPG". Structure "A" and "K", with the lowest solubility (= 34%) both contained "AAY" linkers; structure "A" had two "AAY" linkers, while in structure "K", one "AAY" and one "ALL" were used. Structure "I" (solubility = 63%) also contained one "AAY" and one "KK" linkers. The other structures mainly consisted of linkers with G or S residues. The amino acid contents of the structures were similar and only differed due to the selected linkers. Therefore, no major alteration in the physicochemical properties was observed, except for solubility. On this basis, a decrease in the solubility of these three sequences can be due to the linker composition. Linkers composed of small neutral amino acids, such as glycine or serine tend to increase the protein solubility (Kavoosi et al., 2007), with no major impact on the physical characteristics of the constructs. Thus, such linkers appeared to be the preferred optional linkers in our constructs. Based on the results of this step, structures "B", "C", "D", "E", "F", and "J" seem better than others (Tables 1 and 4). However, further evaluations were still continued for all the structures.

In the next step, the secondary structure of the constructs was studied. Since the accuracy of the predictions of protein secondary

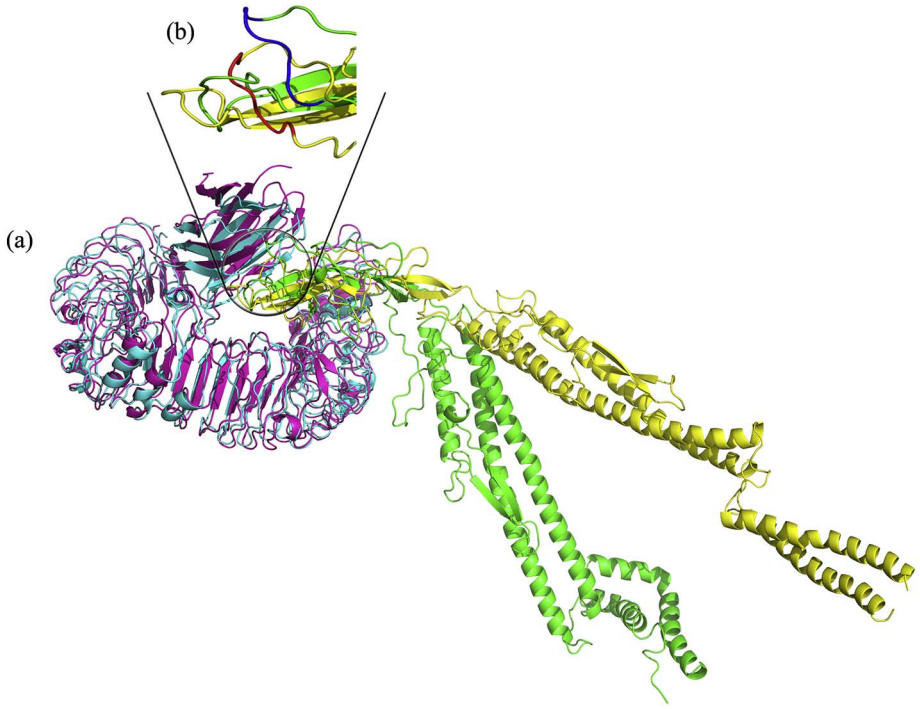

Fig. 4. Superimposition of: (a) TL4/MD2-vaccine complex at the end of the MD simulation time over its obtained structure by the docking process. TLR4/MD2 is shown in magenta and cyan in the MD simulation and docking process, respectively. Vaccine molecule is also colored in green and yellow in MD simulation and docking process, respectively. (b) Adjuvant position in the obtained structure by docking process (red) over the obtained complex at the end of simulation time (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

structures are reported to be up to 84%, and the tested constructs were unnatural structures that only partly resembled natural proteins, five different servers were employed to obtain a more precise prediction (Table 2). RS09 was positioned in the middle of the protein sequences, and its accessibility could play a crucial role for its proper interaction with the bulky receptor (TLR4). Therefore, the main focus of the secondary structure analyses was put on RS09. Moreover, the exposability of RS09 (TLR4 agonist domain) was investigated through three servers (RaptorX-property, SPIDER2, and PaleAle). On the other hand, TLR5 agonist (flagellin) was inserted at head and tail positions to resemble natural flagellin, and no special issue was expected for its interaction with TLR5, as was also confirmed by the results of docking and MD simulation evaluations (Negahdaripour et al., 2017a).

According to Yang et al., RaptorX-property, Porter 4.0, and SPIDER2 are the top three servers for protein secondary structure prediction, and PSIPRED3.3 is also among the top five methods. All these four servers belong to the third generation methods, which use evolutionary information obtained from alignment of multiple homologous sequences, in addition to some novel computational approaches that have been employed through these years (Yang et al., 2016). The GOR IV server, as a popular second-generation method, was also used. In the RaptorX-property server, SS3, H, E, and C represent alpha-helix, beta-sheet, and coil, respectively.

The RaptorX-property was also used for prediction of solvent accessibility, which divides the results into three states: Buried (B) for < 10%, exposed (E) for larger than 40%, and medium (M) for between 10% and 40%. In the SPIDER2 results, rASA of each residue is indicated by a number between zero and nine. A higher number shows more accessibility of the residue, while a rASA < 20% (number zero or one) demonstrates buried residues. The predicted values for the seven residues in RS09 were added to each other and the sum is shown in Table 2. Thus, a higher number indicates a higher relative accessibility of this domain. The PaleAle server predicts protein relative solvent accessibility by classification of each amino acid in one of the four groups: B = completely buried (0–4% exposed), b = partly buried (4–25% exposed), e = partly exposed (25–50% exposed), and E = completely exposed (50 + % exposed).

The PSIPRED server predicted coil structure for RS09 in all the tested sequences, except for structure “J”. Since PSIPRED is very accurate in coil prediction and over-predicts coil residues, some of the structures may be in fact helix or beta-sheets. On account of more flexibility of coils than helixes and beta-sheets, we assumed that a coil structure probably is in favor of a proper interaction between RS09 and TLR4.

The steric hindrance of the other parts of the molecule may limit the interaction of RS09 with TLR4 agonist. On this basis, it was assumed that if RS09 will be more accessible in the protein conformation, its interaction with TLR4 would be probably more convenient. Although, it cannot be inferred as a direct proportional rule. Comparison of the results of these different servers (Table 2) reveals the differences of the predicted structures and exposability scores probably due to the variety in the algorithm and applied databases of these methods. Thus, reaching a certain consensus was not possible, and no single structure could be selected as the best candidate in this step. However, these data may help in the final selection, as is briefed in Table 4.

At the following step, the 3D structures of the eleven sequences were modeled by the I-Tasser server (Fig. S1). The best model was selected for each sequence based on the highest C-scores value, as shown in Table 3. C-score is usually in the range of − 5 to 2, and a higher C-score represents a model with a higher confidence. Except for the 3D models of structures “B” (C-score = − 0.46) and “E” (C-score = 0.01), the other models were presented with close C-score values, in a range of 0.06–0.11. A closer look at the 3D models may help to understand in which models RS09 would be more accessible for interacting with TLR4. As seen in Fig. S1, in all the sequences, the flagellin parts of the molecule at the N- and C-terminals, formed some helixes, and the N- and C-termini bound to each other to form a rod-shape format, similar to natural flagellin. RS09 domain is colored in yellow and shown by a violet oval shape in the I-Tasser modeled 3D conformations in Fig. S1. According to the modeled structures, RS09 in constructs “A”, “B”, “D”, “E”, and “J” formed a coil (Fig. S1a, b, d, and e), while structures “C”, “F”, “G”, “H”, and “K” contained a beta-sheet in RS09 domain (Fig. S1c, f, g, h, and k). Sequence “I” was the only structure that possessed a helix in its RS09 domain (Fig. S1i), which is in accordance with the PSIPRED prediction (Table 2). The results of the predicted exposability of RS09 based on the 3D models are shown in Table 4, which is partly similar to the results of secondary structure

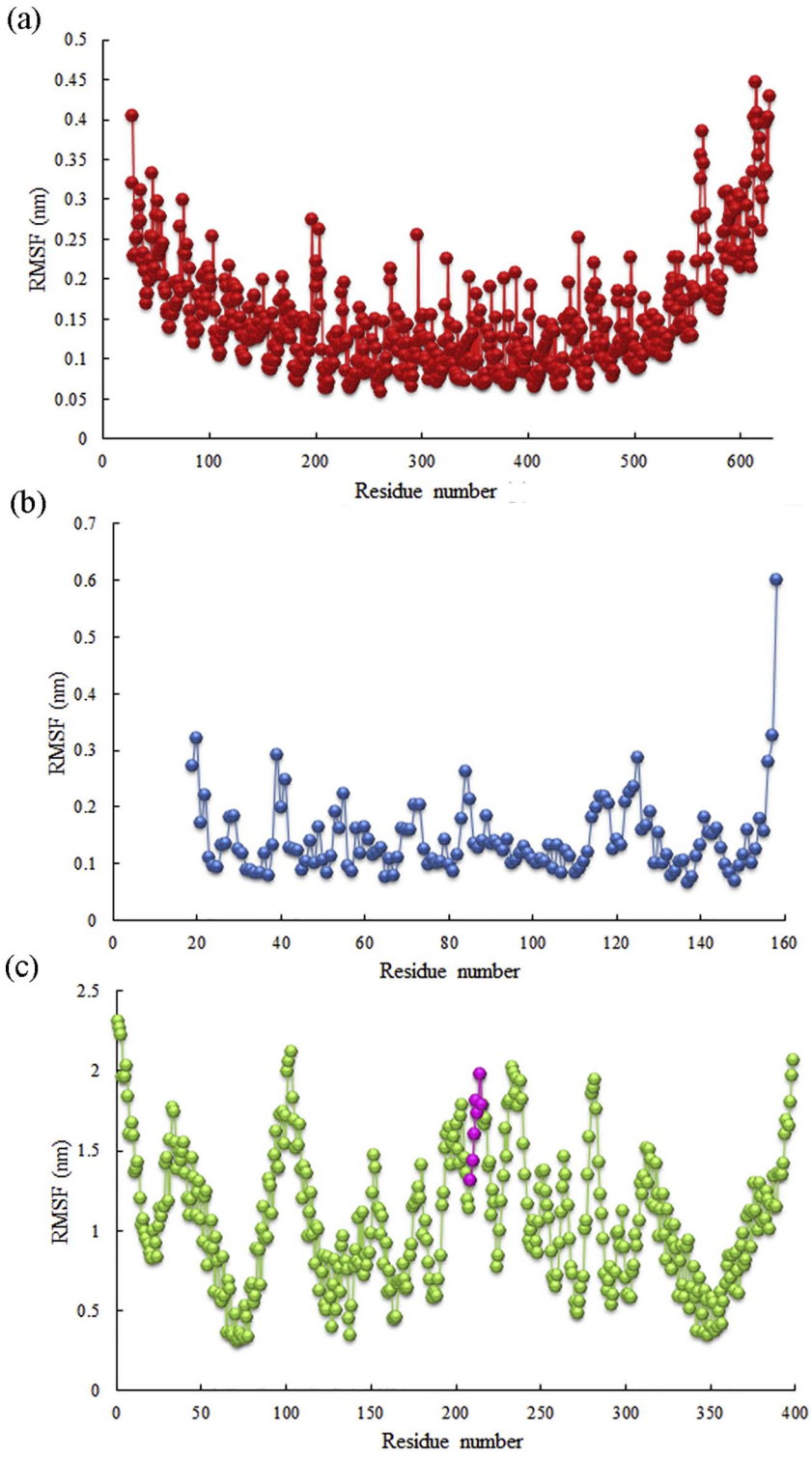

Fig. 5. RMSF plot of all residues of: (a) TLR4 protein, (b) MD2 protein, and (c) vaccine molecule (The adjuvant residues (residues 209–215) of the vaccine protein are colored in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

prediction by the previously mentioned servers. Variation in the predictions again may be due to the difference in the I-Tasser algorithm *versus* the other servers. For example, for structure "D", RS09 secondary structure is predicted as a coil by three of the 5 used servers (Table 2), and in the 3D I-Tasser model, this domain also is shown with a coil conformation. Therefore, it can be inferred that RS09 domain has a coil structure in construct "D" with a high probability.

Based on the previous researches, TLR4 agonist (naturally a lipopolysaccharide or LPS) is recognized by TLR4 through an adapter protein, called myeloid differentiation factor 2 (MD2). MD2 has a cylindrical shape, and the interacting part of the LPS usually enters into the MD2 hydrophobic large pocket (Park et al., 2009). Therefore, for interaction of RS09 with TLR4 in our designed vaccine, the shape of the vaccine structure should be in a way that allows entry of RS09 into the MD2. On this basis, RS09 position in constructs "B" and "I" seem not suitable for such an entry (Fig. S1b and i). Meanwhile, the steric hindrance in structures "C", "F", "G", "J", and "K" can be easily observed (Fig. S1c, f, g, j, and k). Hence, it seems that only structures "A", "D", "E", and "H" might be able to establish a proper interaction with TLR4 (Fig. S1a, d, e, and h). However, in construct "H" (Fig. S1h), RS09 location has a distance with

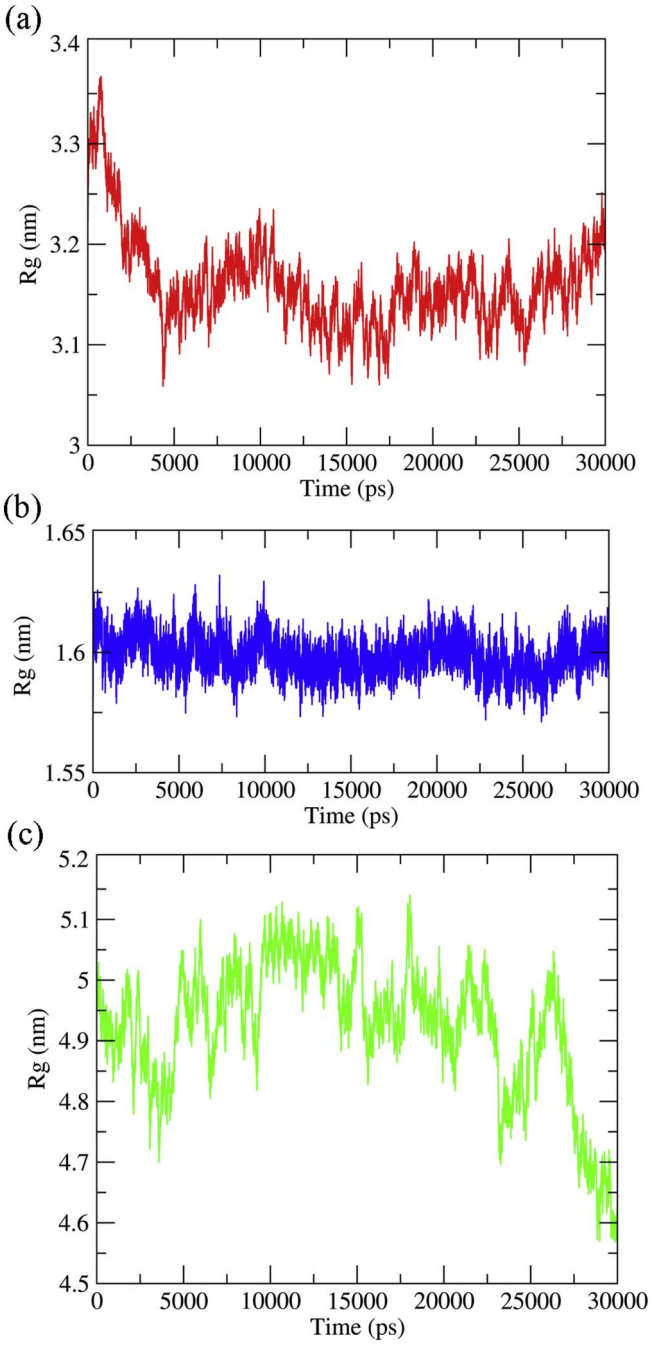

Fig. 6. R_g plot of: (a) TLR4 protein, (b) MD2 protein, and (c) vaccine molecule.

the tip of the bulging loop, which may cause a barrier for the proper interaction. Additionally, in structure "A" (Fig. S1a), the tip of the bulging part is bulky; and RS09, which also contains a beta-sheet, is located in a small distance from the tip. In sum, these predictions could be confirmed in docking studies in the later steps.

Following the 3D modeling, the conformational B-cell epitopes were identified by the DiscoTope 2.0 server. Since the majority of the B-cell epitopes are conformational and only about 10% are linear (Tong and Ranganathan, 2013), a higher number of the identified B-cell epitopes was considered a positive factor for selection of the construct. As seen in Table 3, the number of predicted conformational B-cell epitopes varied in different structures, from the lowest (30) in structure "C" to the highest (78) in structure "H". Though no rule exists in this regard, we assumed a number higher than 55 epitopes as a positive score for this criterion (Table 4).

At the next step, the molecular docking studies were done by the ClusPro server to study a proper engagement of TLR4 with the vaccine. The ClusPro server identifies ten best models in the results. These ten best docked models were investigated for each designed construct by the PyMOL viewer and UCSF Chimera programs (structures not shown due to publication limitation). As seen in Table 4, in most cases, the proper interaction was not established between the receptor and ligand. In other words, the two molecules could not interact properly through RS09 and MD2 of the TLR4 receptor. The proper interaction was only identified in the docked results of structure "D", "H" and probably "E" (Fig. 2). In fact, in the "E" docked model, the part of the vaccine that has entered into the MD2 pocket is the adjacent residues to RS09, and the RS09 residues are mainly positioned outside MD2 pocket (Fig. 2a). Additionally, in the "H" docked structure, the vaccine is not entered as much as structure "D" into MD2 pocket, but has remained close to the MD2 opening (Fig. 2b); hence, a looser interaction can be assumed. The interaction of structure "D" seems the best, with RS09 entry into the opening of MD2 pocket (Fig. 2c).

To be able to select a proper candidate structure, all the results were put together and compared, as seen in Table 4. The positive and negative signs show the sum of the results of constructs in each step. The chance of a coil conformation of RS09, driven from all the five used servers, was added to reach a number shown in Table 4. A higher number means more probability of being coil for RS09. However, the results of this step were not that much decisive in the final decision, because the 3D models and docking results provided a better tool for the final selection. The exposability number was also obtained by adding the results of the numbers assigned by the three servers used for solvent accessibility. For PaleAle results, both "E" and "e", which showed complete and partly exposability of higher than 25%, were counted. The " ± " in the docking results also indicated that though there was a close to correct orientation in their docked models, a closer observation of the docked models showed that the interactions were not ideal. The "H" construct was shown with a low solubility, and its docking was also not proper enough, so could not be a good candidate. Structure "E" also had got a negative point in the previous steps, because the number of its identified conformational B-cell epitopes was low (44). Additionally, the position of RS09 in its docked model was not ideal. Considering all the results shown in Table 4, structure "D" was selected as the best candidate, which got positive scores in all the steps. The antigenicity and allergenicity analyses on structure "D", as reported previously (Negahdaripour et al., 2017a), identified this structure as non-allergen and antigen. Our previous docking and MD studies verified the proper interaction of flagellin segments with TLR5 in the structure "D" (Negahdaripour et al., 2017a). Here, the MD simulations study was done to further confirm the proper engagement of TLR4 with RS09 motif in the structure "D". MD simulation studies are helpful for evaluating of biological processes and interactions between molecules during a timeframe (Sakhteman et al., 2016; Boehr et al., 2009).

In order to ensure the correctness of the MD simulation process, initially, some parameters such as temperature, pressure, density, and volume energy value (kinetic, potential, and total energy) were assessed at the equilibration steps and the end of simulation time. These parameters approved that all steps of MD simulation have been carried out correctly.

In the next step, the final trajectory was used to analyze some important parameters. The first parameter was root mean square deviation (RMSD). The stability of the three protein chains (TLR4 protein, MD2 protein, and the designed vaccine molecule) during MD simulation were evaluated by measuring RMSD values. Although TLR4 protein does not act as the main receptor of the designed vaccine, this protein also established some interactions with the vaccine molecule. These interactions were strengthened during the MD simulation time. As seen in Fig. 3a, these interactions led to reduce backbone RMSD fluctuations of TLR4 protein over the MD simulation. The designed vaccine interacted with MD2 protein directly, so the backbone atoms of this protein were almost stable over the MD simulation time, especially since the 10th ns (Fig. 3b).

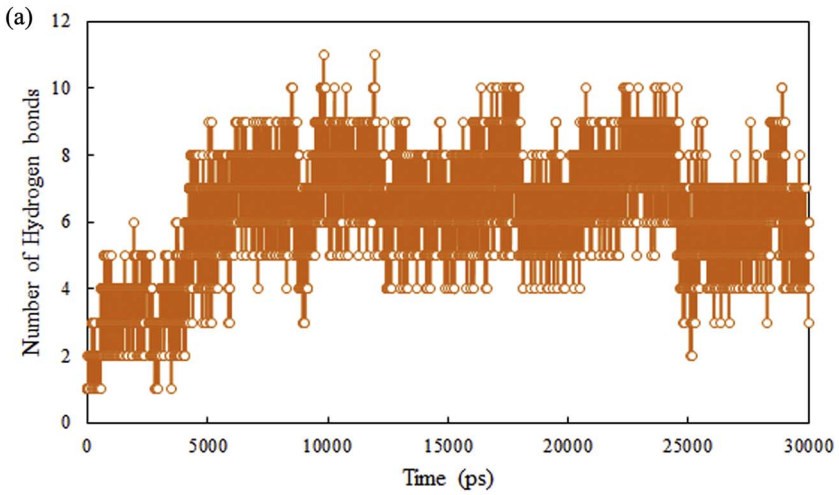

Fig. 7. (a) Changes in the number of hydrogen bonds between the MD2 protein and vaccine molecule during MD simulation. (b) Changes in the number of hydrogen bonds between TLR4/MD2 complex and vaccine molecule during the MD simulation.

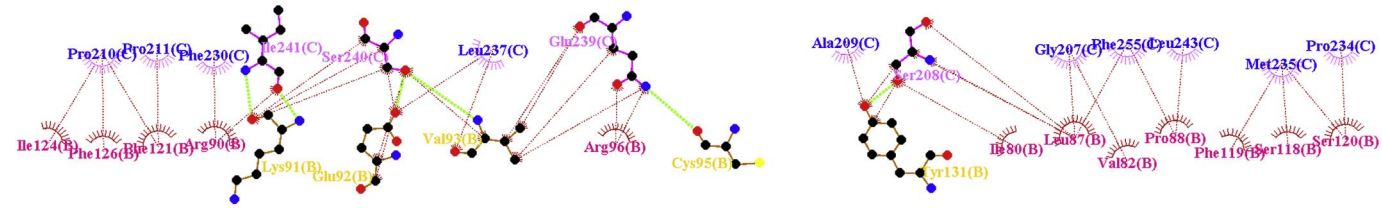

Fig. 8. Two-dimensional representations of the observed interactions between MD2 protein and vaccine at the end of the MD simulation time; hydrogen bonds are shown by green dashed lines between MD2 (orange) and vaccine (pink) residues, and hydrophobic interactions are shown by brick red dashed lines between spoked arcs representing residues of MD2 (deep pink) and vaccine (blue). Chain B and chain C identify MD2 and vaccine protein chains, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As shown in Fig. 3c, the designed vaccine molecule endured a significant displacement during the MD simulation. In order to evaluate the displacement value of protein chains during the MD simulation time, superimposition of the initial configuration of the protein-vaccine complex (its obtained structure at docking process) over the obtained structure at the end of the MD simulation time is shown in Fig. 4. According to Fig. 4a, although TLR4/MD2 complex did endure significant displacement, the vaccine molecule moved considerably. Fig. 4b shows that the TLR4 agonist adjuvant part of this molecule (residues 209–215 of the vaccine sequence) were displaced, so that these residues could reinforce their interactions with the MD2 protein chain. This trend demonstrated its appropriate bindings and interactions inside MD2 protein. On the other hand, those parts of the vaccine molecule that did not interact with MD2 or TLR4 proteins suffered a significant displacement. These parts included flagellin adjuvant that was used as the adjuvant of TLR5 protein. The interactions between these parts of the designed vaccine with TLR5 protein were investigated with details in our previous work (Negahdaripour et al., 2017a).

The root-mean-square fluctuations (RMSF) of all residues were computed for TLR4, MD2, and vaccine protein chains. Due to the presence of strong intramolecular interactions in TLR4 structure and its interactions with MD2 and vaccine molecules, most of the TLR4 residues had small fluctuations (Fig. 5). However, the N/C terminal residues and a few residues located far from MD2 and vaccine molecules showed bigger fluctuations. All the residues of MD2 expect N158 tolerated small fluctuations (Fig. 5). N158 was the furthest residue from the other proteins that was exposed to solvent, so it suffered a significant fluctuation compared to the other MD2 residues. On the other hand, most of the vaccine residues underwent significant fluctuations that were in compliance with their RMSD plots, as previously explained (Fig. 5).

In order to investigate the compactness of the protein chains during

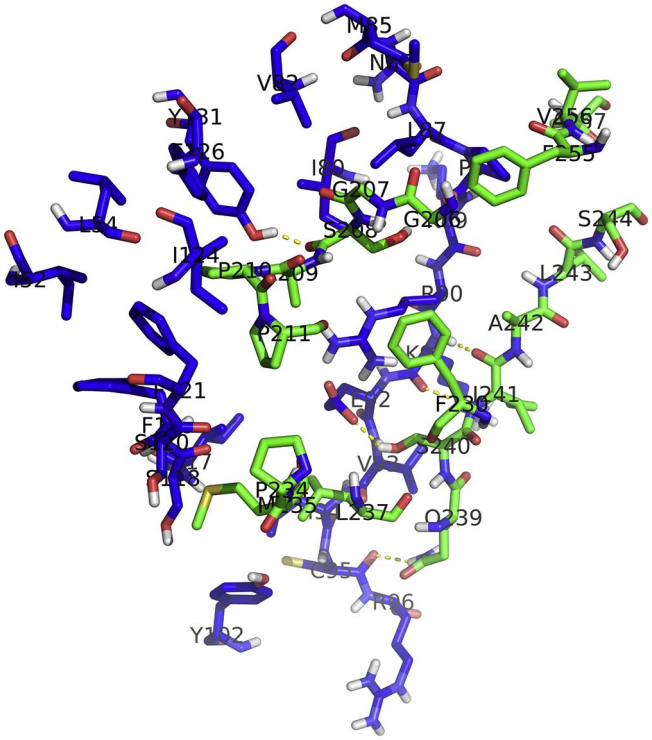

Fig. 9. Three-dimensional representations of the observed interactions between MD2 (blue residues) and vaccine (green residues) proteins at the end of MD simulation time. Hydrogen bonds are shown by yellow dashed lines between MD2 and vaccine residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

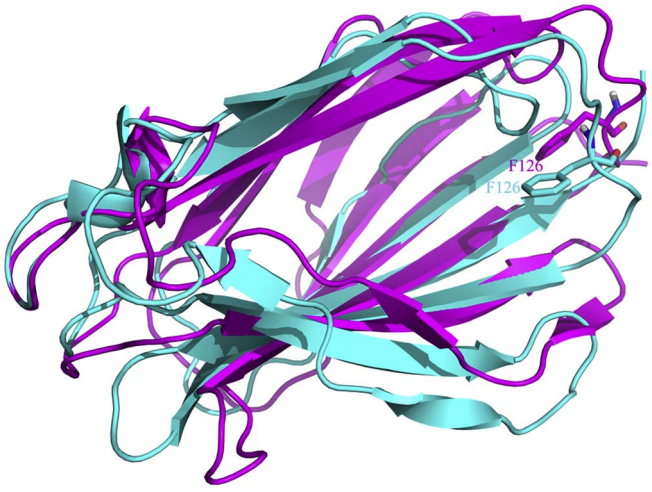

Fig. 10. Superimposition of the MD2 protein chain at the end of simulation time (magenta chain) over its structure obtained during the docking process (cyan chain). F126 residue is shown in stick form in both structures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the MD simulation time, changes in radius of gyration (R_g) was also evaluated. The R_g values of TLR4 protein were decreased somewhat at the end of the simulation time, which resulted from strong intramolecular interactions (Fig. 6a). The MD2 protein had almost constant R_g values over the MD simulation time (Fig. 6b). In fact, the structure of this protein included a lot of permanent intramolecular interactions. Moreover, the two other proteins also established many strong intermolecular interactions with MD2. These interactions preserved the relative position of MD2 atoms to some extent, as was confirmed by RMSF and RMSD plots. Rg plot of the vaccine molecule demonstrated a significant decrease, showing that the vaccine protein gained a more compact and stable form during the MD simulation (Fig. 6c).

Hydrogen bond formation/deformation was investigated over the MD simulation time as well. Firstly, the number of hydrogen bonds between the MD2 protein chain and the designed vaccine were evaluated. As seen in Fig. 7a, they were increased clearly. Moreover, the number of hydrogen bonds between TLR4/MD2 complex and the vaccine molecule were also calculated. According to Fig. 7b, these numbers also augmented significantly. These trends confirmed that the appropriate interactions between the designed vaccine and TLR4/MD2 complex led to improve the number of hydrogen bonds, so that the TLR4/MD2/vaccine complex became more stable during the MD simulation time.

LPS agonist engages some of the MD2 residues, which have been already mentioned in some papers (Park et al., 2009; Billod et al., 2016; Carpenter and O'Neill, 2009). Some of these residues include F126, I54, Y131, I124, R90, K91, S118, K122. LigPlot + v.1.4.5 (Wallace et al., 1995) was used to analyze and present the two-dimensional representation of the interactions between the MD2 protein and the designed vaccine (Fig. 8). The three-dimensional representation of the observed interactions is also shown in Fig. 9. It clearly shows that the designed vaccine molecule could interact with a majority of the mentioned residues. In fact, this molecule could imitate the LPS behavior appropriately. F126 residue and its conformational change are very important for the binding between a TLR4 agonist and the MD2 protein chain. F126 gets a closed state in the presence of a TLR4 agonist (Billod et al., 2016; Carpenter and O'Neill, 2009). The state of this residue is clearly specified in the crystallographic structure and the obtained structure at the end of the simulation time (Fig. 10). This residue maintained its closed state in the presence of the designed vaccine during 30 ns MD simulation time. All in all, the designed vaccine (structure "D") showed the ability to establish a stable interaction with TLR4, as well as TLR5.

However, given the limitations of the *in silico* tools in prediction of quaternary structures (Yu et al., 2015), adding to the lack of complete knowledge about the immunological pathways in the body (Scheiblhofer et al., 2017), the efficacy of the designed vaccine should be further proven in wet-lab experiments.

In order to prepare the candidate vaccine for experimental studies, the selected construct was reverse translated and codon optimized (Negahdaripour et al., 2017a).

4. Conclusions and future perspective

Based on the performed analyses, one construct that showed proper characteristics and desired interaction with TLR4 and TLR5, was selected from a group of eleven designed proteins, which should be further validated in experimental investigations. In designing a string-of-beads multi-epitope vaccine, there are numerous options for insertion of different functional segments and motifs in the sequence as well as the linker selection. This, obviously, can affect the physicochemical properties, the 3D structure, and thereby the final efficiency of the designed vaccine. In this study, a structural vaccinology approach was followed to discover the best construct, albeit with limited initially designed sequences. Such a strategy seems necessary in designing similar multi-epitope vaccines. However, due to the high number of possible constructs, investigating all the possible structures through such an approach is almost impossible. To solve this problem, only computational algorithms may help. Hence, designing bioinformatics tools that would be able to execute all the necessary evaluations sequentially on large numbers of sequences is highly recommended.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2017.12.008.

Acknowledgment

This study was supported by Grant no. 13435 from the Research Council of Shiraz University of Medical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran.

Conflict of interests

Authors declare no conflict of interests.

References

Agostini, F., Cirillo, D., Livi, C.M., Delli Ponti, R., Tartaglia, G.G., 2014. ccSOL omics: a webserver for solubility prediction of endogenous and heterologous expression in *Escherichia coli*. Bioinformatics 30, 2975–2977.

Berger, S., Lowe, P., Tesar, M., 2015. In: Schmidt, Stefan R. (Ed.), Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges. MAbs 7. pp. 456–460.

Billod, J.-M., Lacetera, A., Guzmán-Caldentey, J., Martín-Santamaría, S., 2016. Computational approaches to Toll-like receptor 4 modulation. Molecules 21, 994.

Boehr, D.D., Nussinov, R., Wright, P.E., 2009. The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. 5, 789–796.

Bruni, L., Barrionuevo-Rosas, L., Albero, G., Aldea, M., Serrano, B., Valencia, S., Brotons, M., Mena, M., Cosano, R., Muñoz, J., Bosch, F., de Sanjosé, S., Castellsagué, X., 2016. Human papillomavirus and related diseases in the world. In: Summary Report. ICO Information Centre on HPV and Cancer (HPV Information Centre).

Buchan, D.W.A., Ward, S.M., Lobley, A.E., Nugent, T.C.O., Bryson, K., Jones, D.T., 2010. Protein annotation and modelling servers at University College London. Nucleic Acids Res. 38, W563–W568.

Carpenter, S., O'Neill, L.A., 2009. Recent insights into the structure of Toll-like receptors and post-translational modifications of their associated signalling proteins. Biochem. J. 422, 1–10.

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004. Cluspro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20, 45–50.

Consortium, T.U., 2014. Uniprot: a hub for protein information. Nucleic Acids Res. 43, D204–D212.

Dang, H.X., Lawrence, C.B., 2014. Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics 30, 1120–1128.

Dimitrov, I., Naneva, L., Doytchinova, I., Bangov, I., 2013. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 30, 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, 856–866.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. Int. J. Cancer 136, E359–386.

Garnier, J., Gibrat, J.F., Robson, B., 1996. Gor method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. 266, 540–553.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the EXPASY server. In: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, pp. 571–607.

George, R.A., Heringa, J., 2002. An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng. Des. Sel. 15, 871–879.

Hajighahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi, S.S., 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.

Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., Wang, J., Sattar, A., Yang, Y., Zhou, Y., 2015. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci. Rep. 5, 11476.

Heffernan, R., Dehzangi, A., Lyons, J., Paliwal, K., Sharma, A., Wang, J., Sattar, A., Zhou, Y., Yang, Y., 2016. Highly accurate sequence-based prediction of half-sphere exposures of amino acid residues in proteins. Bioinformatics 32, 843–849.

Karimi, Z., Nezafat, N., Negahdaripour, M., Berenjian, A., Hemmati, S., Ghasemi, Y., 2015. The effect of rare codons following the atg start codon on expression of human granulocyte-colony stimulating factor in *Escherichia coli*. Protein Expr. Purif. 114, 108–114.

Kavoosi, M., Creagh, A.L., Kilburn, D.G., Haynes, C.A., 2007. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. Biotechnol. Bioeng. 98, 599–610.

Klement, M., Liu, C., Loo, B.L., Choo, A.B., Ow, D.S., Lee, D.Y., 2015. Effect of linker flexibility and length on the functionality of a cytotoxic engineered antibody fragment. J. Biotechnol. 199, 90–97.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S., 2017. The ClusPro web server for protein-protein docking. Nat. Protoc. 12, 255–278.

Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M., 2012. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol. 8, e1002829.

Levy, A., Pitcovski, J., Frankenburg, S., Elias, O., Altuvia, Y., Margalit, H., Peretz, T., Golenser, J., Lotem, M., 2007. A melanoma multiepitope polypeptide induces specific CD8 + T-cell response. Cell. Immunol. 250, 24–30.

Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., Sette, A., 2002. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J. Immunol. 168, 5499–5506.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins Struct. Funct. Bioinf. 50, 437–450.

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, 2936–2943.

Mahmoodi, S., Nezafat, N., Barzegar, A., Negahdaripour, M., Nikanfar, A.R., Zarghami, N., Ghasemi, Y., 2016. Harnessing bioinformatics for designing a novel multiepitope peptide vaccine against breast cancer. Curr. Pharm. Biotechnol. 17, 1100–1114.

Mizel, S.B., Bates, J.T., 2010. Flagellin as an adjuvant: cellular mechanisms and potential. J. Immunol. 185, 5677–5682.

Moise, L., Gutierrez, A., Kibria, F., Martin, R., Tassone, R., Liu, R., Terry, F., Martin, B., De Groot, A.S., 2015. IVAX: an integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines. Hum. Vaccin. Immunother. 11, 2312–2321.

Mooney, C., Pollastri, G., 2009. Beyond the twilight zone: automated prediction of structural properties of proteins by recursive neural networks and remote homology information. Proteins 77, 181–190.

Negahdaripour, M., Nezafat, N., Ghasemi, Y., 2016. A panoramic review and in silico analysis of IL-11 structure and function. Cytokine Growth Factor Rev. 32, 41–61.

Negahdaripour, M., Eslami, M., Nezafat, N., Hajighahramani, N., Ghoshoon, M.B., Shoolian, E., Dehshahri, A., Erfani, N., Morowvat, M.H., Ghasemi, Y., 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 (5), 575–596.

Negahdaripour, M., Nezafat, N., Hajighahramani, N., Rahmatabadi, S.S., Ghasemi, Y., 2017c. Investigating CRISPR-CAS systems in *Clostridium botulinum* via bioinformatics tools. Infect. Genet. Evol. 54, 355–373.

Negahdaripour, M., Nezafat, N., Hajighahramani, N., Soheil Rahmatabadi, S., Hossein Morowvat, M., Ghasemi, Y., 2017d. In silico study of different signal peptides for secretory production of interleukin-11 in *Escherichia coli*. Curr. Proteomics 14, 112–121.

Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M.J., Omidinia, E., 2014. A novel multi-epitope peptide vaccine against cancer: an in silico approach. J. Theor. Biol. 349, 121–134.

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, M.R., Ghasemi, Y., 2017. Designing an efficient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. Mol. BioSyst. 13 (4), 699–713.

Panatto, D., Amicizia, D., Bragazzi, N.L., Rizzitelli, E., Tramalloni, D., Valle, I., Gasparini, R., 2015. Chapter eight - human papillomavirus vaccine: state of the art and future perspectives. In: Rossen, D. (Ed.), Advances in Protein Chemistry and Structural Biology. vol. 101. Academic Press, pp. 231–322.

Park, B.S., Song, D.H., Kim, H.M., Choi, B.-S., Lee, H., Lee, J.-O., 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458, 1191–1195.

Pollastri, G., McLysaght, A., 2005. Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics 21, 1719–1720.

Pollastri, G., Martin, A.J.M., Mooney, C., Vullo, A., 2007. Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information. BMC Bioinf. 8.

Rahmatabadi, S.S., Nezafat, N., Negahdaripour, M., Hajighahramani, N., Morowvat, M.H., Ghasemi, Y., 2016. Studying the features of 57 confirmed CRISPR loci in 29strains of *Escherichia coli*. J. Basic Microbiol. 56, 645–653.

Rahmatabadi, S.S., Sadeghian, I., Nezafat, N., Negahdaripour, M., Hajighahramani, N., Hemmati, S., Ghasemi, Y., 2017. In silico investigation of pullulanase enzymes from various *Bacillus* species. Curr. Proteomics 14, 175–185.

Robinson, C.R., Sauer, R.T., 1998. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. Proc. Natl. Acad. Sci. U. S. A. 95, 5929–5934.

Rosenthal, K.S., Stone, S., Koski, G., Zimmerman, D.H., 2017. Leaps vaccine incorporating her-2/neu epitope elicits protection that prevents and limits tumor growth and spread of breast cancer in a mouse model. J. Immunol. Res. 2017, 8.

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

Sakhteman, A., Khoddami, M., Negahdaripour, M., Mehdizadeh, A., Tatar, M., Ghasemi, Y., 2016. Exploring 3D structure of human gonadotropin hormone receptor at antagonist state using homology modeling, molecular dynamic simulation, and cross-docking studies. J. Mol. Model. 22, 225.

Scheiblhofer, S., Laimer, J., Machado, Y., Weiss, R., Thalhamer, J., 2017. Influence of protein fold stability on immunogenicity and its implications for vaccine design. Expert Rev. Vaccines 16, 479–489.

Schmidt, S.R., 2013. Fusion Proteins: Applications and Challenges, Fusion Protein Technologies for Biopharmaceuticals. John Wiley & Sons, Inc., pp. 1–24.

Schubert, B., Kohlbacher, O., 2016. Designing string-of-beads vaccines with optimal spacers. Genome Med. 8, 9.

Sen, T.Z., Jernigan, R.L., Garnier, J., Kloczkowski, A., 2005. Gor V server for protein secondary structure prediction. Bioinformatics 21, 2787–2788.
Shin, W.-H., Lee, G.R., Heo, L., Lee, H., Seok, C., 2014. Prediction of protein structure and interaction by GALAXY protein modeling programs. Biodesign 2 (1), 1–11.
Tong, J.C., Ranganathan, S., 2013. 2 - Immunoglobulins and B Cell Responses, Computer-aided Vaccine Design. Woodhead Publishing, pp. 13–20.
van Leeuwen, H.C., Strating, M.J., Rensen, M., de Laat, W., van der Vliet, P.C., 1997. Linker length and composition influence the flexibility of OCT-1 DNA binding. EMBO J. 16, 2043–2053.
Wallace, A.C., Laskowski, R.A., Thornton, J.M., 1995. Ligplot: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. Des. Sel. 8, 127–134.
Wang, J., Zhang, D., Li, J., 2013. PREAL: prediction of allergenic protein by maximum Relevance Minimum Redundancy (mRMR) feature selection. BMC Syst. Biol. 7 (Suppl. 5), S9.
Wang, S., Li, W., Liu, S., Xu, J., 2016a. RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res. 44, W430–W435.
Wang, S., Peng, J., Ma, J., Xu, J., 2016b. Protein secondary structure prediction using deep convolutional neural fields. Sci. Rep. 6, 18962.
Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–410.
Wu, W.H., Alkutkar, T., Karanam, B., Roden, R.B., Ketner, G., Ibeanu, O.A., 2015. Capsid display of a conserved human papillomavirus L2 peptide in the adenovirus 5 hexon protein: a candidate prophylactic hpv vaccine approach. Virol. J. 12, 140.
Yang, J., Zhang, Y., 2015. I-Tasser server: new development for protein structure and function predictions. Nucleic Acids Res. 43, W174–W181.
Yang, Y., Gao, J., Wang, J., Heffernan, R., Hanson, J., Paliwal, K., Zhou, Y., 2016. Sixty-five years of the long march in protein secondary structure prediction: the final stretch? Brief. Bioinform., bbw129.
Yang, Y., Heffernan, R., Paliwal, K., Lyons, J., Dehzangi, A., Sharma, A., Wang, J., Sattar, A., Zhou, Y., 2017. SPIDER2: a package to predict secondary structure, accessible surface area, and main-chain torsional angles by deep neural networks. Methods Mol. Biol. 1484, 55–63.
Yano, A., Onozuka, A., Asahi-Ozaki, Y., Imai, S., Hanada, N., Miwa, Y., Nisizawa, T., 2005. An ingenious design for peptide vaccines. Vaccine 23, 2322–2326.
Yu, K., Liu, C., Kim, B.-G., Lee, D.-Y., 2015. Synthetic fusion protein design and applications. Biotechnol. Adv. 33, 155–164.