Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Molecular and Cellular Probes 48 (2019) 101446

Contents lists available at ScienceDirect

Molecular and Cellular Probes

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

Designing self-assembled peptide nanovaccine against Streptococcus pneumoniae: An in silico strategy

Hesam Dorosti[a,b], Mahboobeh Eslami[a], Navid Nezafat[a,b,*], Fardin Fadaei[a,b], Younes Ghasemi[a,b,**]

[a] Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[b] Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Keywords:
Streptococcus pneumoniae
Epitope
Nanovaccine
Structural evaluation
Immunoinformatics

ABSTRACT

Streptococcus pneumoniae is the main cause of diseases such as meningitis, pneumoniae and sepsis, especially in children and old people. Due to costly antibiotic treatment, and increasing resistance of pneumococcus, developing high-efficient protective vaccine against this pathogen is an urgent need. Although the pneumoniae polysaccharide vaccine (PPV) and pneumonia conjugate vaccines (PCV) are the efficient pneumococcal vaccine in children and adult groups, but the serotype replacement of *S. pneumoniae* strains causes the reduction in efficacy of such vaccines. For overcoming the aforesaid drawbacks epitope-based vaccines are introduced as the relevant alternative. In our previous research, the epitope vaccine was designed based on immunodominant epitopes from PspA, CbpA antigens as cellular stimulants and PhtD, PiuA as humoral stimulants. Because the low immunogenicity is the main disadvantage of epitope vaccine, in the current study, we applied coiled-coil self-assembled structures for developing our vaccine. Recently, self-assembled peptide nanoparticles (SAPNs) have gained much attention in the field of vaccine development due to their multivalency, self-adjuvanticity, biocompatibility, and size similarity to pathogen. In this regard, the final designed vaccine is comprised of cytotoxic T lymphocytes (CTL) epitopes from PspA and CbpA, helper T lymphocytes (HTL) epitopes from PhtD and PiuA, the pentamer and trimmer oligomeric domains form 5-stranded and 3-stranded coiled-coils as self-assembled scaffold, Diphtheria toxoids (DTD) as a universal T-helper, which fused to each other with appropriate linkers. The four different arrangements based on the order of above-mentioned compartments were constructed, and each of them were modeled, and validated to find the 3D structure. The structural, physicochemical, and immunoinformatics analyses of final vaccine construct represented that our vaccine could stimulate potent immune response against *S. pneumoniae*; however, the potency of that should be approved via various *in vivo* and *in vitro* immunological tests.

1. Introduction

Streptococcus pneumoniae is the leading cause of bacterial diseases such as sepsis, meningitis and pneumonia. Due to high mortality rate and the global burden of disease, especially in children under 5 and elderly, the National Institutes of Health (NIH) has paid more attention to pneumococcal disease in recent years [1]. The high-cost of antibiotic therapy and increasing resistance of pneumococcus to current antibiotics, suggesting vaccination against *S. pneumoniae* as the best method for prevention of disease, according to Centers for Disease Control and Prevention (CDC) report (https://www.cdc.gov/pneumococcal/about/prevention.html). [2]. The pneumococcal vaccine development is mainly based on surface antigens, including immunogenic proteins and carbohydrates [3]. The 23-valent streptococcus pneumoniae polysaccharide vaccine (PPV) and 7, 10, or 13-valent pneumonia conjugated vaccines (PCV-7, 10, −13) are carbohydrate-based vaccines and stimulate serotype-specific immunity [4]. Up to now, the 23-valent pneumoniae polysaccharide vaccine is still applied in elderly and the Pcv-7 as a routine vaccine for children in some countries; however, the vaccine does not cover some serotypes, especially in elderly. PCV10 and PCV13 as the next generation of vaccine could cover more *S. pneumoniae* serotypes. However, the serotype replacement strategy of pneumococcal infections causes reducing the efficacy of this group of vaccine [5]. Thus, it is necessary to design an efficient novel vaccine. The epitope-based pneumococcal vaccines are the last generation of vaccines that are in the pipeline [2]. In this context, pneumococcal surface

* Corresponding author. Pharmaceutical Sciences Research Center, Shiraz University of Medical Science, Shiraz, Iran.
** Corresponding author. Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, P.O. Box 71345-1583, Shiraz, Iran.
E-mail addresses: n.nezafat@srbiau.ac.ir (N. Nezafat), ghasemiy@sums.ac.ir (Y. Ghasemi).

https://doi.org/10.1016/j.mcp.2019.101446
Received 10 August 2019; Received in revised form 6 September 2019; Accepted 10 September 2019
Available online 11 September 2019
0890-8508/

protein A (PspA) that is located on the surface of bacteria, and choline binding protein A (CbpA), as the most crucial protective surface antigens, have been introduced as vaccine candidate [6]. Another important vaccine candidate is pneumococcal histidine triad (Pht) protein family, including PhtE and PhtD as the conserved member of Pht expressed in pneumococcal pathogenesis [7]. Other pneumococcal proteins that have shown potential as an efficient vaccine candidate and are needed for full virulence of *S. pneumoniae* in animal models are pneumococcal iron ABC transporters, including two lipoprotein components, PiuA and PiaA [8]. In general, the majority of epitope-based vaccines can cause induction of cellular or/and humoral immunity as well as offer a safer and more easy-produced alternative for the prevention of diseases; however, the low immunogenicity is the main disadvantage of such types of vaccines. For solving this problem, several methods have been used to enhance the immunogenicity epitopes vaccines, like prescribing the multiple doses of vaccine, adding immunostimulatory agents such an adjuvant to the vaccine [9], and the recent strategy is applying self-assembled peptide motif including beta-sheets or coiled-coil structures in the epitope vaccine [10]. Besides above-mentioned strategies, incorporating universal T-helper epitopes like diphtheria toxoids (DTD) into vaccine can enhance vaccine efficacy. DTD is the most essential components of diphtheria vaccines that are widely used as a universal T-helper [11]. In some cases, the low efficacy of epitope peptide-based vaccines may be according to their fast extracellular degradation, rapid distribution from the injection site and inefficient uptake by antigen-presenting cells (APCs) [12,13]. Several investigations have shown that linking the epitope vaccine to certain carriers (e.g., polymeric or lipidic nanostructures) can develop the efficacy of vaccine via conferring to aforesaid obstacles [13–15]. In this context, designing epitope vaccine bases on self-assembled peptide nanoparticles (SAPN) scaffold, which acts as an antigen-presenting system, is a novel strategy. Moreover, the SAPN vaccines have self-adjuvanticity, without any toxicity; in this regard, it can bypass the low immunogenicity of the epitope peptide vaccine. In this study, CD4 epitopes were selected from PiuA, PhtD antigens, and CD8 epitopes were chosen from PspA, CbpA antigens of *S. pneumoniae*; moreover, the DTD was determined as universal T-helper compartment. All selected CD4 and CD8 epitopes were inserted between the pentamer and trimer oligomeric domains, which can construct self-assembled structures. Additionally, all segments were fused to each other by the appropriate linkers. The immunoinformatics and structural evaluations of the designed vaccine, showed that it can induce proper immunity against *S. pneumoniae*; however, experimental immunological tests should confirm the efficacy of our vaccine.

2. Methods

2.1. Designing fusion protein

At the first step, the amino acid sequences of PspA (Accession no. FJ668667), CbpA (Accession no. LN847353), PhtD (AE007317), PiuA (Accession no. CNWI02000107) of *Streptococcus pneumonia* (AE014074.1) were retrieved from National Center Biotechnology Information (NCBI) at www.ncbi.nlm.nih.gov in FASTA format. The PspA and CbpA were applied for selecting cytotoxic T lymphocytes (CTL) epitopes, PhtD and PiuA were used for selectin helper epitopes, moreover DTD was employed as a universal T-helper. Moreover, to generate our self-assembled vaccine the pentamer and trimer oligomeric domains, which respectively form 5-stranded and 3-stranded coiled-coils were added to vaccine construct. In the next step, each part of the vaccine, including $CD4^+$ epitopes $CD8^+$ epitopes, universal T-helper, and the pentamer and trimer oligomeric domains were fused by proper amino acid linkers. According to the order of vaccine parts, four different constructs were built.

The schematic structures of the four designed vaccine constructs are shown in Fig. 1.

2.2. Structural evaluates

2.2.1. Physicochemical characterization

The instability index, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY), extinction coefficients, estimated half-life, and aliphatic index of the designed vaccine construct were evaluated by the ProtParam server at http://web.expasy.org/protparam/ [16]. The solubility of the different proteins was predicted by the Protein–Sol server at http://protein-sol.manchester.ac.uk. Protein–Sol employs a fast sequence-based method for predicting the solubility of protein based on available data for *Escherichia coli* protein solubility [17]. The solubility value greater than 0.45 is defined to have a higher solubility than the average soluble *E. coli* protein. In addition, fold propensity, and net charge for per amino acid can be profiled in calculation along the sequence [18]. SOLpro at http://scratch.proteomics.ics.uci.edu was another server used for solubility prediction based on overexpression in *E. coli*; its prediction method is based on a two-stage SVM architecture, and the accuracy of the server is about 74%.

2.2.2. Exposability of amino acids to solvent and evaluation of the secondary structure

The secondary structure of the protein sequence was defined by RaptorX-Property at http://raptorx2.uchicago.edu/StructurePropertyPred/predict/. The RaptorX-Property server can evaluate the structure-property of a protein sequence via the template-free method. In this server, the secondary structure (SS) and solvent accessibility (ACC) of the input sequence are predicted through a powerful machine learning model called Deep Convolutional Neural Fields (DeepCNF) [19]. The ~84% Q3 accuracy for 3-state SS, ~72% Q8 accuracy for 8-state SS, ~66% Q3 accuracy for 3-state solvent accessibility, obtained for the server.

2.3. Protein modeling

The I-TASSER software at http://zhanglab.ccmb.med.umich.edu/I-TASSER/, and the GalaxyTBM software at http://galaxy.seoklab.org/softwares were used for generating the 3D structure of the four designed vaccine constructs and based on the obtained results, the best models were selected.

According to CASP12 (critical assessment of methods of protein structure prediction) data, I-TASSER is ranked at the top server for automated protein structure modeling. The I-TASSER prediction method includes four successive steps: 1) threading (secondary structure projection and template selection) 2) structural fragment assembly applying a modified replica-exchange Monte Carlo simulation method. 3) Selection of the model through employing clustering structure decoys and refinement by fragment-guided molecular dynamics simulation (FG-MD) or ModRefiner, and lastly, 4) structure-based biological function annotation. I-TASSER introduces C-score as a confidence score for assessing the overall quality of the model [20]. UCSF chimera and Discovery studio 3.5 software were employed for visualizing a 3D model of the protein. GalaxyTBM (template-based modeling) predicts protein structure from the sequence when experimental structures of homologous proteins are available as templates and refines loop or terminus regions by ab initio modeling.

2.4. Refinement and validation of the 3D modeled structure

All structures were refined by the GalaxyRefine server at http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE [21,22]. The server refines the protein model via loop modeling and overall structural relaxation that is performed through molecular dynamics simulation [23]. The server is among the best available web servers for structure refinement according to the CASP10 assessment. Then all 3D refined structures were compared, and the models were validated by the ERRAT server at http://nihserver.mbi.ucla.edu/ERRATv2/ [24],

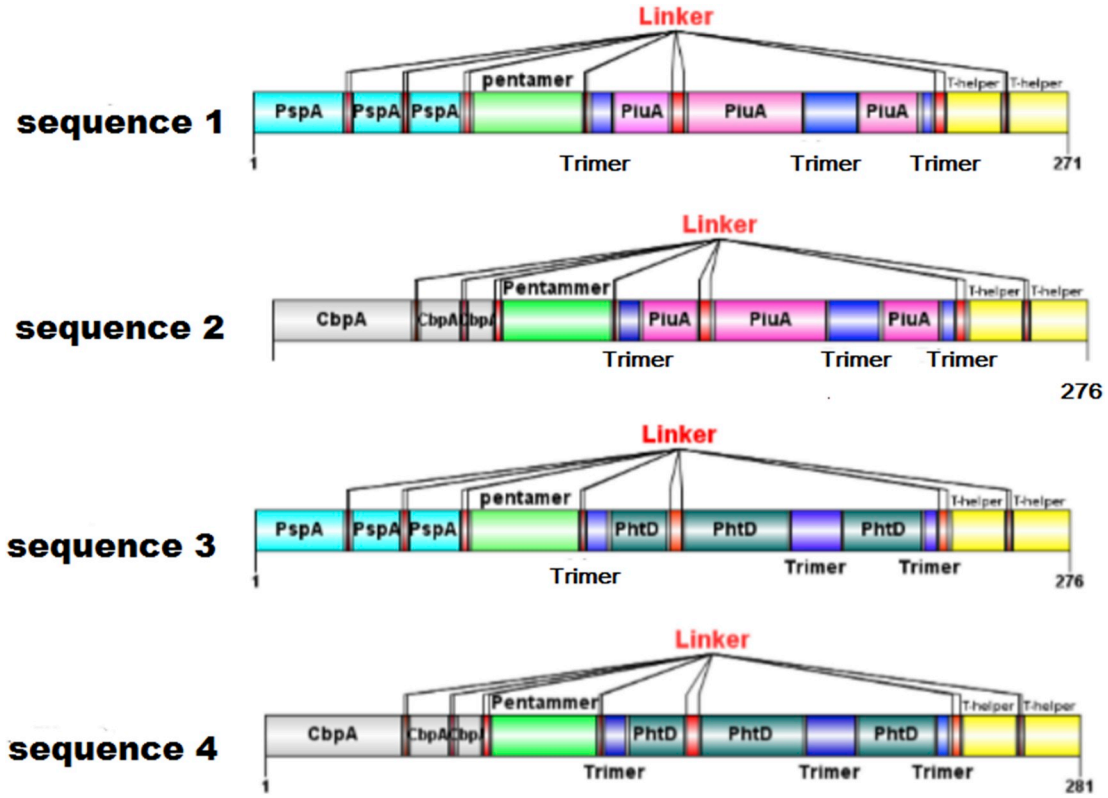

Fig. 1. The schematic diagram of the four designed vaccine construct consists of helper epitopes (PhtD and PiuA), CTL epitopes (CbpA and PspA), the pentamer and trimer oligomeric domains form 5-stranded and 3-stranded coiled-coils and conserved amino acid residues of DTD as a universal T-helper, which fused together by proper linkers.

Ramachandran Plot Assessment at (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [25], Verify-3D at http://services.mbi.ucla.edu/Verify_3D/ [26] and proSA-web at https://prosa.services.came.sbg.ac.at [27]. The overall quality score is known as an indicator of Prosa-Web and calculated by analyzing atomic coordinates of the model. The ProSA-web z-score is represented on a plot, that contains the z-score of experimentally identified structures, which are available in PDB. The Ramachandran plot validates the protein model through evaluating the residue-by-residue stereo chemical qualities of models. In ERRAT server, the specific atomic interactions are considered to determine the correct and incorrect defined regions of protein structures, and plos the value of the error function versus position of a 9-residue sliding window, analyzed by comparison with statistics from a database of reliable high-resolution crystallography structures.

2.5. Protein homo-oligomer structure prediction

The GalaxyHomomer server at http://galaxy.seoklab.org/homomer was applied to predict protein homo-oligomers structure from its monomer structure subunit. Either a sequence or PDB of the monomer can be used as input. The server, according to the accessibility of suitable oligomer templates, predicts five homo-oligomer structures through three strategies, 1) sequence similarity-based approach, 2) structure similarity-based approach, 3) ab initio docking approach. Recently the developed model refinement methods, GalaxyLoop, GalaxyRefineComplex, have been integrated into the GalaxyHomomer that can improve the quality of the model [23,28,29]. The performance of this server was better than other accessible methods verified on benchmark sets [30,31].

2.6. Conformational B-cell epitopes prediction

After the 3D modeling of protein was done, to predict the conformational B-cell epitopes DiscoTope 2.0 server at http://www.cbs.dtu.dk/services/DiscoTope/was applied. In this server the final score was calculated by two methods: a) contact numbers derived from surface accessibility, b) a novel epitope partiality amino acid score [32]. The default threshold, sensitivity and specificity of the server were −3.7, 0.47, 0.75 respectively.

2.7. The antigenicity and allergenicity evaluation

The ANTIGENpro server at http://scratch.proteomics.ics.uci.edu/ [33] and the VaxiJen v2.0 server at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html were applied for antigenicity analysis [34]. ANTIGENpro is a sequence-based, alignment-free and pathogen-independent predictor, with about 76% accuracy. Moreover, VaxiJen applies alignment-independent method that is based on auto cross-covariance transformation of protein sequences into uniform vectors of principal amino acid features. The accuracy of the server is in the range of 70%–89% depending on the selected target organism (Bacteria, tumor, fungal, parasite, virus).

Two servers, the AllergenFP v.1.0 at http://ddgpharmfac.net/AllergenFP/ [35], and the Algpred at http://www.imtech.res.in/raghava/algpred/ [36] were used for evaluating of allergenicity. In AllergenFP, to discriminate allergens from non-allergens, an alignment-independent descriptor-based fingerprint method is used; the descriptors are hydrophobicity, size, relative abundance, helix and beta-strand of protein. The sensitivity, specificity, and accuracy of the server are about 87%, 90%, and 88%. AlgPred employs a combination of different approaches (SVMc + IgEepitope + ARPs BLAST + MAST) to predict the allergenicity. The server accuracy is estimated 85% at a threshold −0.4.

Table 1
The four different arrangements of designed sequences as self-assembled epitope peptide vaccine candidates consist of PspA, PiuA, linkers, pentamer and trimmer structures, and universal T helper (DTD).

Sequence 1 [1]	WYYLNSNGAMATGWLQYNGSWYYLNANGAMAGSLQYNGSWYYLNANGA MAGSLQYNGSWYYLNANGAMATYGSWQTWNAKWDQWSNDWNAWESD WQAWKDDWAEWRALWMGGRLLLRLERVTFDLGAADTIRALGFEKNGPGP GNVGSMKEPDLEAIAALEPDLIIASPRTQKFVDKFKEIAPEALEALARFVANLS MRLATQKAKEELAKLDKSIQEVATRNESRGGSGPVFAGANYAAWAVNVAQ VIGSQSIALSSLMVAQAIPLVGEL
Sequence 2 [2]	KYLRELNVLEEKSKKEELTSKTKKEIDAAFEQFNKDTLKPGEKVEEAQGSNY PTNTYKTLELEIGSKEPQNEEKIYGSWQTWNAKWDQWSNDWNAWESDWQA WKDDWAEWRALWMGGRLLLRLERVTFDLGAADTIRALGFEKNGPGPGNVG SMKEPDLEAIAALEPDLIIASPRTQKFVDKFKEIAPEALEALARFVANLSMRLA TQKAKEELAKLDKSIQEVATRNESRGGSGPVFAGANYAAWAVNVAQVIGSQ SIALSSLMVAQAIPLVGEL
Sequence 3 [3]	WYYLNSNGAMATGWLQYNGSWYYLNANGAMAGSLQYNGSWYYLNANGA MAGSLQYNGSWYYLNANGAMATYGSWQTWNAKWDQWSNDWNAWESD WQAWKDDWAEWRALWMGGRLLLRLERNGKVPYDAIISEELLMKDPNGPGP GENGVPRYIPAKDLSAETAAGIDSKLAKQESLSHKLGEALEALARFVANLSM RLAKKDSLSEAERAAAQAYAKEKGLTPPSTRNESRGGSGPVFAGANYAAWA VNVAQVIGSQSIALSSLMVAQAIPLVGEL
Sequence 4 [4]	KYLRELNVLEEKSKKEELTSKTKKEIDAAFEQFNKDTLKPGEKVEEAQGSNY PTNTYKTLELEIGSKEPQNEEKIYGSWQTWNAKWDQWSNDWNAWESDWQA WKDDWAEWRALWMGGRLLLRLERNGKVPYDAIISEELLMKDPNGPGPGEN GVPRYIPAKDLSAETAAGIDSKLAKQESLSHKLGEALEALARFVANLSMRLA
	KKDSLSEAERAAAQAYAKEKGLTPPSTRNESRGGSGPVFAGANYAAWAVNV AQVIGSQSIALSSLMVAQAIPLVGEL

[1] PspA (red), Linker (black), Pentamer (green), Trimmer (blue), PiuA (violent), Universal T-helper (yellow)

[2] CbpA (red), Linker (black), Pentamer (green), Trimmer (blue), PiuA (violent), Universal T-helper (yellow)

[3] PspA (red), Linker (black), Pentamer (green), Trimmer (blue), PhtD (violent), Universal T-helper (yellow)

[4] CbpA (red), Linker (black), Pentamer (green), Trimmer (blue), PhtD (violent), Universal T-helper(yellow)

2.8. MD simulation studies

MD simulation is known as a unique technique for better understanding of biological systems. The use of this tool has recently expanded significantly [37,38]. In this work, MD simulation was carried out on the monomeric chain and homodimer structure, which chosen by docking process (GalaxyHomomer server). MD simulations were performed for the intended systems using Gromacs 5.0.7 package [39] with Amber99sb-ildn force field [40]. Both systems were solvated in a triclinic box of TIP3P water molecules with a minimal distance of 12 Å from any edge of the box. Moreover, the concentration of 150 mM of NaCl (physiological conditions) was added to neutralize the systems. The Particle Mesh Ewald (PME) method and the LINCS algorithm were used to calculate long-range electrostatic interactions and constrain all bonds, respectively. The non-bonded cutoff of 10 Å was used. Firstly, the prepared systems were minimized using the steepest descent algorithm. The first equilibration step was done under NVT ensemble at 300 K during 500 ps. This step was followed by a 3000 ps NPT equilibration run at 1 bar pressure and 300 K temperature. The velocity rescaling with $\tau_t = 0.1$ ps and Parrinello-Rahman barostat algorithms with $\tau_p = 0.5$ ps were used for the temperature and pressure coupling, respectively. Finally, MD production run was performed for 150 ns under equivalent conditions at 1 bar and 300 K. The time step of 2.0 fs was applied, and the output files were saved every two ps for the further analyses.

3. Results

3.1. Fusion vaccine design

According to our previous study, the final CD4 helper epitopes were selected from PiuA antigen (aa 29-47), (aa 69-107), (aa 138-157), and PhtD antigen (aa 91-110), (aa 401-436), and (aa 567-593). Moreover, the CD8 CTL epitopes were selected from PspA antigen (aa 400-430), (aa 454-470), and (aa 494-511), and CbpA antigen (aa 61-108), (aa 283- 298), and (aa 316- 324) [41]. Additionally, the fragments (aa 271-290) and (aa 331-350) of DTD was chosen as T-helper. The pentamer sequence, WQTWNAKWDQWSNDWNAWESDWQAWKDDWAEWRALWM, and trimer sequence, RLLLRLER, EALEALARFVANLSMRLA, and RNESR were chosen as the self-assembled motifs. According to the sequence of CD8 CTL epitopes (PspA, CbpA) and CD4 helper epitopes (PiuA, PhtD) 4 following sequences were designed (Fig. 1), and Table 1.

3.2. Structural analyses of constructs

3.2.1. Physicochemical characterization

Theoretical PI, number of amino acid, estimated protein half-life in mammalian, yeast, and *E. coli*, aliphatic index, GRAVY, stability, and solubility of the designed structures are shown in Table 2.

3.2.2. Exposability of amino acids to solvent and evaluation of the secondary structure

The secondary structure elements, solvent accessibility, and disordered position of vaccine construct were predicted by RAPTORX (Table 3); moreover, details of prediction results are shown in Fig. 2.

Table 2
Evaluation of physicochemical features of four designed sequences.

Sequences	Number of amino acids	pI	Estimated half-life	Aliphatic index	extinction coefficients	GRAVY	Stability	Solubility	
								Protein-sol server	Sol-pro server
Seq. 1	271	4.94	2.8 h in mammalian 3 min in yeast 2 min in *E. coli*	78.71	112870 Abs 0.1% (=1 g/l) 3.765	−0.264	20.74 Stable	0.460	0.61981
Seq. 2	276	5.00	1.3 h in mammalian 3 min in yeast 3 min in *E. coli*	80.72	73450 Abs 0.1% (=1 g/l) 2.377	−0.537	25.78 Stable	0.690	0.79912
Seq. 3	276	5.34	2.8 h in mammalian 3 min in yeast 2 min in *E. coli*	74.13	117340 Abs 0.1% (=1 g/l) 3.871	−0.394	26.82 Stable	0.493	0.70609
Seq. 4	281	5.22	1.3 h in mammalian 3 min yeast 3 min *E. coli*	76.19	77920 Abs 0.1% (=1 g/l) 2.494	−0.659	31.66 Stable	0.663	0.87598

Table 3
Evaluation of predicted Secondary structure by RAPTORX.

Sequence Number	Secondary structure	Solvent access	Positions predicted as disordered
Sequence 1	24%H, 26%E, 49%C[a]	31%E, 36%M, 31%B[b]	2(0%)
Sequence 2	55%H, 9%E, 35%C	46%E, 21%M, 32%B	52(18%)
Sequence 3	14%H, 35%E, 50%C	40%E, 35%M, 24%B	12(4%)
Sequence 4	56%H, 3%E, 39%C	62%E, 17%M, 20%B	59(20%)

[a] H, E, and C stand for alpha-helix, beta-sheet and coil, respectively.
[b] B, M and E stand for Buried, Medium and Exposed residues, respectively.

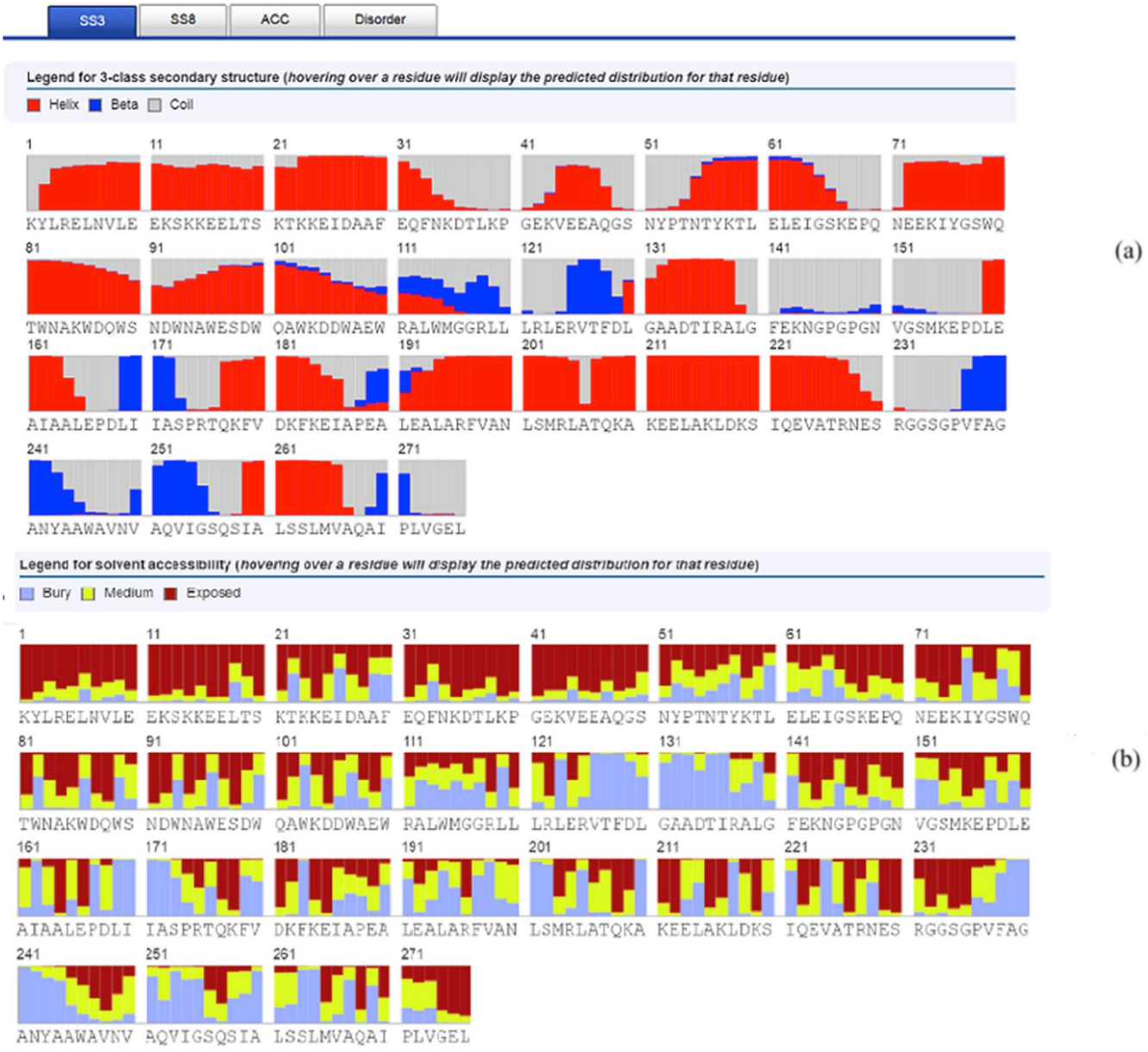

Fig. 2. Detailed prediction results for (a) Secondary structure prediction (SS) for 3-state: alpha-helix, beta-sheet and coil (b) The relevant solvent accessibility(ACC) is divided into three states by 2 cutoff values: 10% and 40% so that the three states have equal distribution. Buried for less than 10%, exposed for larger than 40% and medium for between 10% and 40% Exposability of amino acids to solvent and evaluation of secondary structure.

3.3. Protein modeling, refinement, and validation of the 3D modeled structure

The 3D structure of the four designed vaccine constructs were modeled by the I-TASSER and Galaxyweb software. Galaxyweb and I-TASSER provided five and two models respectively for each sequence. All structures were refined by the GalaxyRefine. Finally, the refined structures were compared, and the best model was validated by the ERRAT server, Ramachandran Plot, Verify-3D [26] and proSA-web. The results are presented in supplementary Tables (Tables S1 and S2). According to GalaxyWeb results, model-1 of sequence-1, model-5 of sequence-2, model-5 of sequence-3, model-4 of sequence-4 were selected; Additionally, based on I-TASSER results, model-2 of sequence-1, model-1 of sequence-2, model-1 of sequence-3, and model-2 of sequence-4

Table 4
The high-score 3D models obtained from four designed sequences by GalaxyWeb server.

Validation method	Sequence 1 Model 1	Sequence 2[a] Model 5	Sequence 3 Model 5	Sequence 4 Model 4
Ramachandran plot				
Favored regions%	93.9	**94.9**	93.1	93.9
Allowed regions%	4.1	**3.6**	6.6	5.4
Outlier regions%	2.2	**1.5**	0.4	0.7
ERRAT (quality factor)	88.235	**94.4882**	72.5869	86.4469
Verify 3D score%	61.99	**60.87**	83.33	56.23
Prosa-web	−2.98	**−3.19**	−2.95	−3.62

[a] The best 3D model among all modeled sequences (sequence 2 Model 5) is in bold font.

Table 5
The best docking models among different homo-oligomer structures (2, 3, 4, 5, 6, 7-mer).

Model No	Number of subunits	Interface area	Docking score
1	2-mer[a]	2005.5	1220.741
2	3-mer	2417.3	614.203
3	4-mer	4918.5	633.435
4	5-mer	6167.4	726.400
5	6-mer	7548.3	860.877
6	7-mer	9546.9	804.326

[a] The 2-mer structure has the best docking score among different homo-oligomer structures, and is represented in bold font.

were chosen. Finally, the model-5 of sequence-2 obtained from GalaxyWeb was selected as the best model for the following steps (Table 4).

3.4. Protein homo-oligomer structure prediction

The homo-oligomer of our protein was predicted by the GalaxyHomomer.

Five homo-oligomer structures were predicted with the given oligomeric states (Tables S3 and S4). Based on docking results, the 2-mer structure of the protein possessed the best score (docking score = 1220.742) between 2, 3, 4, 5, 6, 7-mer structures (Table 5). Afterwards, the best 2-mer models were subjected to refinement process. Finally, the initial homomer model and the best-refined homomer model were validated by Ramachandran plot, Verify3D, and Prosa-web. The quality of the protein geometry was evaluated by applying Ramachandran plot, Verify 3D, and ProSA z-score. In the primary model, 94.9%, 4.7%, and 0.4% of residues were located in favored, allowed, and outlier regions, respectively, whereas, in the refined model, 95.6%, 4%, and only 0.4% of residues were placed in favored, allowed, and outlier regions, respectively. The Verify 3D result showed that in the initial model 52.61% of the residues had an average 3D–1D score, while in the refined model the Verify 3D score was 58.12. The Prosa-web z score of initial model was −3.69 that reach to −3.94 in final model.

3.5. Conformational B-cell epitopes prediction

Conformational B-cell epitopes play a significant role in humoral immunity, thus the tertiary structure of the final vaccine was exploited as an input for conformational epitope prediction through DiscoTope. There were 97 B-Cell epitope residues identified out of 276 total residues at a DiscoTope score threshold of −3.7 that are shown in Table 6.

3.6. Antigenicity and allergenicity evaluation

The probability of the whole vaccine antigenicity was estimated 0.922525% by the ANTIGENpro and 0.4807 by the VaxiJen (threshold is 0.4). Based on the AllergenFP result, the vaccine was probable Non-allergen; additionally, the Algpred showed the protein was a non-allergen.

3.7. Molecular dynamic simulation

To ensure that the conditions have been well-controlled during MD simulation, some important parameters were evaluated before analyzing the results. In this regard, pressure, temperature, density, and energy values were assayed. The obtained values for these parameters indicated MD simulation is done correctly.

Firstly, the analysis of MD trajectories was performed for the monomeric system. Root-mean-square deviation (RMSD) parameter was evaluated to investigate the stability of the protein chain throughout MD simulation. As seen in Fig. 3a, RMSD plot of backbone atoms of this protein chain shows a significant increase up to 15 ns, but it reaches to steady-state at the rest of simulation time. Since RMSD parameter cannot determine the flexible and rigid regions of a protein chain, Root-mean-square fluctuation (RMSF) values of all residues were assayed as well. All residues in the N- and C-terminal regions show the high fluctuations (Fig. 3b). They have fewer intra-molecular interactions than other residues so that they, and their neighbor residues showed high mobility during MD simulation. Moreover, some residues in the other regions of protein have endured a high fluctuation as well. The majority of these residues are placed in the linker or T-helper regions. The most side chains of these residues have been exposed to solvent molecules, so they have also established a few intra-molecular interactions with other residues over MD simulation time.

The Radius of Gyration (Rg) of a protein is a measure of its compactness [42]. This parameter is estimated as follows:

$$R_g = \left(\frac{\sum_i |r_i|^2 \, m_i}{\sum_i m_i} \right)^2$$

Where m_i is the mass of atom i and r_i the position of atom i with respect to the center of mass of the molecule.

Changes in the R_g values of the monomeric chain are presented in Fig. 3c. It is clear that the displacements in RMSD and RMSF plots have been led to decrease Rg values during MD simulation. According to aforesaid equation, decreasing in Rg values indicates the reduction in distance of atoms from the mass center of the protein so that the protein chain has been folded during this time. Superimposition of initial configuration over the obtained structure at the end of MD simulation time confirms this trend (Fig. 4).

In the second step, trajectory analysis was done for homo-dimer system. Firstly, RMSD parameter was evaluated for the backbone atoms of two protein chains. Both chains had suitable stability over MD simulation, although chain A was slightly more stable. They reached to steady state after about 10 ns (Fig. 5a). RMSF plots of both chains are almost similar although the residues of chain B have endured little more fluctuations than the chain A. All residues of the both protein chains showed few fluctuations in homodimer system in comparison with the monomeric system, especially in Pentamer, Trimer regions and their neighbors (Fig. 5b).

Rg parameters were also measured for both chains over MD simulation. According to Fig. 5c, Rg values of chain A have decreased throughout 100 ns MD simulation, although it suffered some fluctuations during this time. Howsoever Rg values of chain B have increased

Table 6
Conformational B-cell epitopes determined from protein vaccine using DiscoTope 2.0 server.

Number	Residue Number	Amino Acid	DiscoTope score
1	1	LYS	2.991
2	2	TYR	1.657
3	3	LEU	1.58
4	4	ARG	2.993
5	5	GLU	2.569
6	6	LEU	1.478
7	7	ASN	1.984
8	8	VAL	2.667
9	9	LEU	2.477
10	10	GLU	3.262
11	11	GLU	3.495
12	12	LYS	3.975
13	13	SER	2.964
14	14	LYS	2.928
15	15	LYS	3.601
16	16	GLU	2.583
17	17	GLU	3.044
18	18	LEU	3.297
19	19	THR	3.366
20	20	SER	0.878
21	21	LYS	1.217
22	22	THR	2.626
23	23	LYS	1.873
24	24	LYS	1.832
25	25	GLU	1.021
26	26	ILE	1.722
27	27	ASP	−0.338
28	28	ALA	−1.617
29	29	ALA	1.354
30	30	PHE	0.374
31	31	GLU	−0.405
32	32	GLN	0.511
33	33	PHE	0.888
34	34	ASN	0.639
35	35	LYS	2.649
36	36	ASP	2.668
37	37	THR	0.578
38	38	LEU	1.066
39	39	LYS	1.513
40	40	PRO	0.39
41	41	GLY	−1.787
42	42	GLU	−2.362
43	43	LYS	−0.272
44	44	VAL	−0.992
45	45	GLU	0.922
46	46	GLU	−1.02
47	47	ALA	−2.808
48	48	GLN	−0.423
49	49	GLY	−0.912
50	50	SER	−3.451
51	52	TYR	−0.71
52	53	PRO	−2.683
53	57	TYR	−2.71
54	58	LYS	−2.405
55	60	LEU	−1.962
56	61	GLU	−1.744
57	62	LEU	0.104
58	63	GLU	2.732
59	64	ILE	2.38
60	65	GLY	2.212
61	66	SER	4.435
62	67	LYS	4.989
63	68	GLU	4.361
64	69	PRO	5.054
65	70	GLN	3.018
66	71	ASN	5.09
67	72	GLU	5.409
68	73	GLU	4.592
69	74	LYS	4.426
70	75	ILE	3.961
71	76	TYR	5.392
72	77	GLY	4.143
73	78	SER	5.244
74	79	TRP	5.89
75	80	GLN	5.022
76	81	THR	4.11
77	82	TRP	3.964
78	83	ASN	5.244
79	84	ALA	4.131
80	85	LYS	4.137
81	86	TRP	4.953
82	87	ASP	4.152
83	88	GLN	4.062
84	89	TRP	5.577
85	90	SER	5.237
86	91	ASN	4.556
87	92	ASP	5.122
88	93	TRP	3.793
89	94	ASN	3.653
90	95	ALA	4.566
91	96	TRP	3.836
92	97	GLU	5.674
93	98	SER	2.668
94	99	ASP	2.504
95	100	TRP	0.613
96	101	GLN	−0.292
97	102	ALA	−1.929

during the first half of simulation time, Rg plot of this chain has shown a decreasing trend at the rest of simulation time (Fig. 5c). In fact, MD simulation prepared a suitable opportunity for two protein chains to get the best configuration. In other words, they tried to have more stable conformation over this time. RMSD, RMSF, and Rg plots of chain B indicate that it has tolerated further displacement in order to improve its configuration and interactions relative to chain A.

The number of hydrogen bonds were also evaluated between chain A and chain B over MD simulation time. As seen in Fig. 5d, it has endured some fluctuations up to 60 ns, and finally it has reached to steady state. Obviously, the obtained structure at the end of simulation time has more hydrogen bonds in comparison with the initial structure.

Distance between the mass center of two protein chains in the homodimer system was also measured during MD simulation time (Fig. 5e). It is obvious that this distance has decreased significantly up to 60 ns. It has been almost stable from the 100th ns. The similar trend between Fig. 5d and e shows the distance between the center mass of two protein chains has affected by the number of hydrogen bonds between two chains.

Solvent accessible surface (SASA) parameter was also measured during MD simulation (Fig. 6). Hydrophilic SASA was presented for the epitope parts of protein chains, whereas hydrophobic SASA was shown for Pentamer, and Trimer parts. The increasing trend of the hydrophilic SASA values of the epitope parts (Fig. 6a), and hydrophobic SASA values of Pentamer, and Trimer parts indicate that the epitope parts prefer to interact with solvent molecules during 150 ns MD simulation while Pentamer, and Trimer parts incline to establish hydrophobic interactions in the course of this time (Fig. 6b).

Eventually, the final snapshot of homodimer system at the end of the simulation was superimposed over the initial structure that has obtained from docking process. According to Fig. 7, the structure of this system has not changed significantly over MD simulation while protein folding has happened during this time to achieve a more sustainable state.

4. Discussion

According to NIH reports, *Streptococcus pneumoniae* is still one of the main sources of morbidity and mortality worldwide, especially in children and elderly and categorized into high burden disease [43]. The capsular polysaccharide (CPS) is recognized as an important virulence

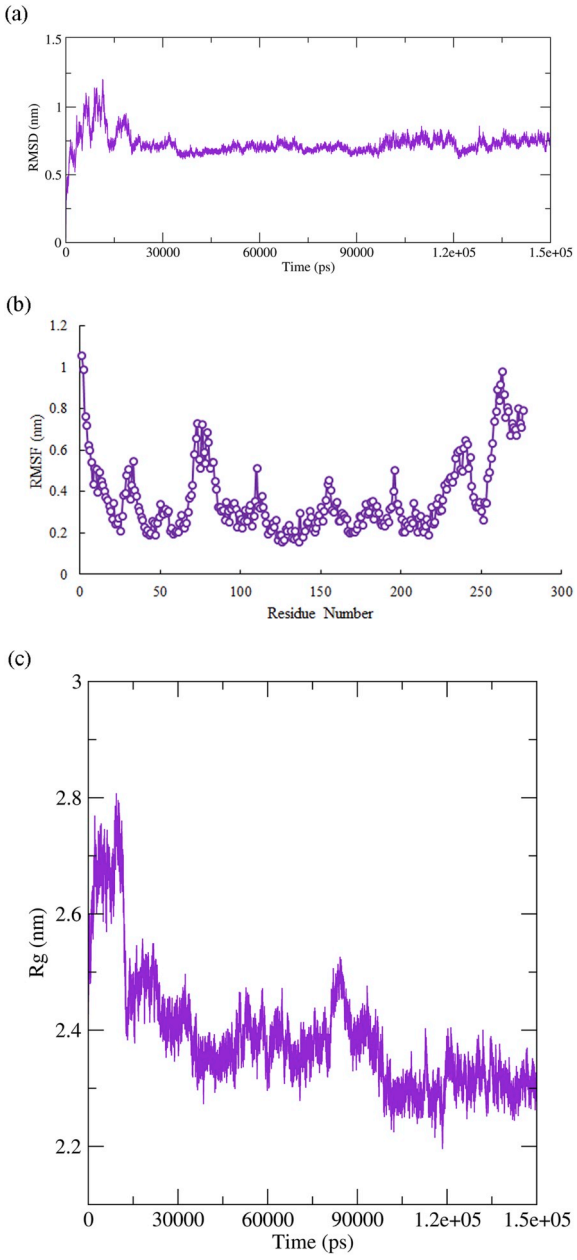

Fig. 3. (a) Backbone RMSD plots of the monomeric chain, (b) RMSF plots of all the residues of monomeric chain, (c) Rg plot of the monomeric chain.

factor of *S. pneumoniae*; moreover, based on CPSs the *S. pneumoniae* is categorized into 95 distinct serotypes. There are ten common serotypes, 4, 6 (6A/6B/6C/6D), 9V/9A, 14, 15F/15A, 15B/15C, 18 (18F/18A/18B/18C), 19F, 19A and 23F, that are the main reason for most pediatric pneumococcal infections worldwide [44]. In the context of antibiotic therapy, common medication to fight bacterial infections, can lead to the emergence of antibiotics resistance *S. pneumoniae* strains; in this regard, designing the efficient preventive vaccine can be the rational choice [45]. Recently, in some country, child and elderly pneumococcal vaccination have been added to the national immunization program [46]. Now, the PPV and PCV pneumococcal vaccines are available in the market and could prevent individuals from different types of pneumococcal disease, but in some cases, serotype replacement caused the failure of vaccination [47]. Complications and high-cost of PCVs production have caused this vaccine to become unaffordable, especially in developing countries [48]. There are some strategies to overcome serotype replacement phenomena; firstly, the addition of new

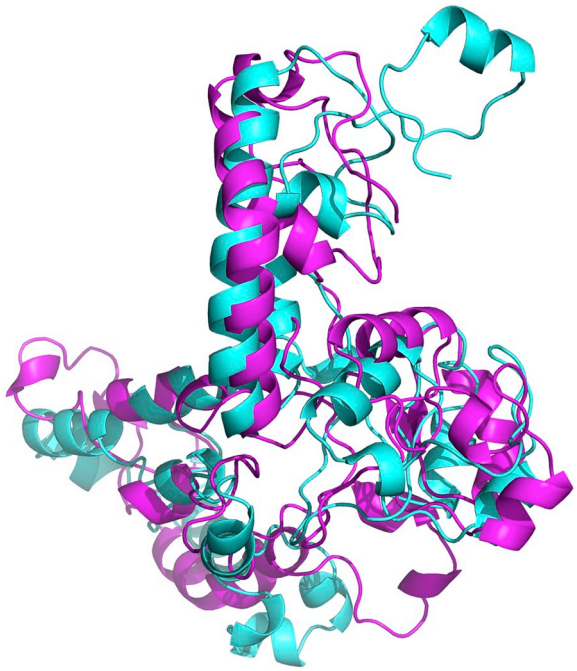

Fig. 4. Superimposition of initial configuration of monomeric chain (cyan) over its structure obtained at the end of MD simulation time (magenta).

pneumococcal serotypes e.g. 22F, and 33F to PCV-15, this approch is currently on phase 3 clinical trial [49]. The second approach is employing the whole-cell pneumococcal vaccines; in this method, the attenuated or inactivated pathogen can be used [50]. An appropriate alternative to polysaccharide-based vaccines, is applying pneumococcal protective antigen in vaccine construct. There are several conserved proteins [51] among different *S. pneumoniae* serotypes, most of them are surface proteins such as PspA, CbpA, and PhtD [52]. In this regard, in our previous immunoinformatic study, the immunodominant epitopes were selected from four antigens, PspA, CbpA, PhtD, and PiuA [53]. Recently, epitope-based vaccines have been introduced as a safer alternative to previous types of vaccines. However, they suffer from some disadvantages such as low immunogenicity. Along with low immunogenicity of peptide vaccines, the variability of immunodominant regions of antigens among different serotypes, is another obstacle in designing efficient peptide-based pneumococcal vaccines [54]. If the suitable conserve epitope was selected, the vaccine would have high population coverage [55]. As mentioned-above low immunogenicity is one of the key drawbacks of epitope-based peptide vaccines; to overcome this issue several approaches are employed. Incorporating universal T-helper epitopes into a vaccine is one of the rational strategies for enhancing immunogenicity. Previous studies recommended that diphtheria toxoids (DTD) contain "universal" epitopes for human $CD4^+$ cells (residues 632-651 and 950-969 of tetanus toxoids, and residues 271-290, 321-350, 351-370, 411-430, and 431-450 of DTD); in this context, Okita BM et al., showed The DTD peptides (residues 271-290 and 331-350) are recognized by over than 80% of the toxoid-sensitized subjects [56]. In this regard, the sequences 271-290, and 331-350 were selected as universal T-helper epitopes in our study. The efficacy of nanovaccines were confirmed in different investigations, including influenza nanovaccine [57], chitosan nanoparticle against H1N1 influenza virus [58], and chaitosan-formulated polyepitope vaccine against pulmonary mycobacterial [59]. Another important approach for increasing the immunogenicity of peptide vaccine is the use of self-assembled peptide structures, including alpha helical or beta sheet motifs, which can obtain from natural or de novo structures. Coiled-coil and β-sheet bio-structural motifs are the most common scaffold for constructing SAPN vaccine.

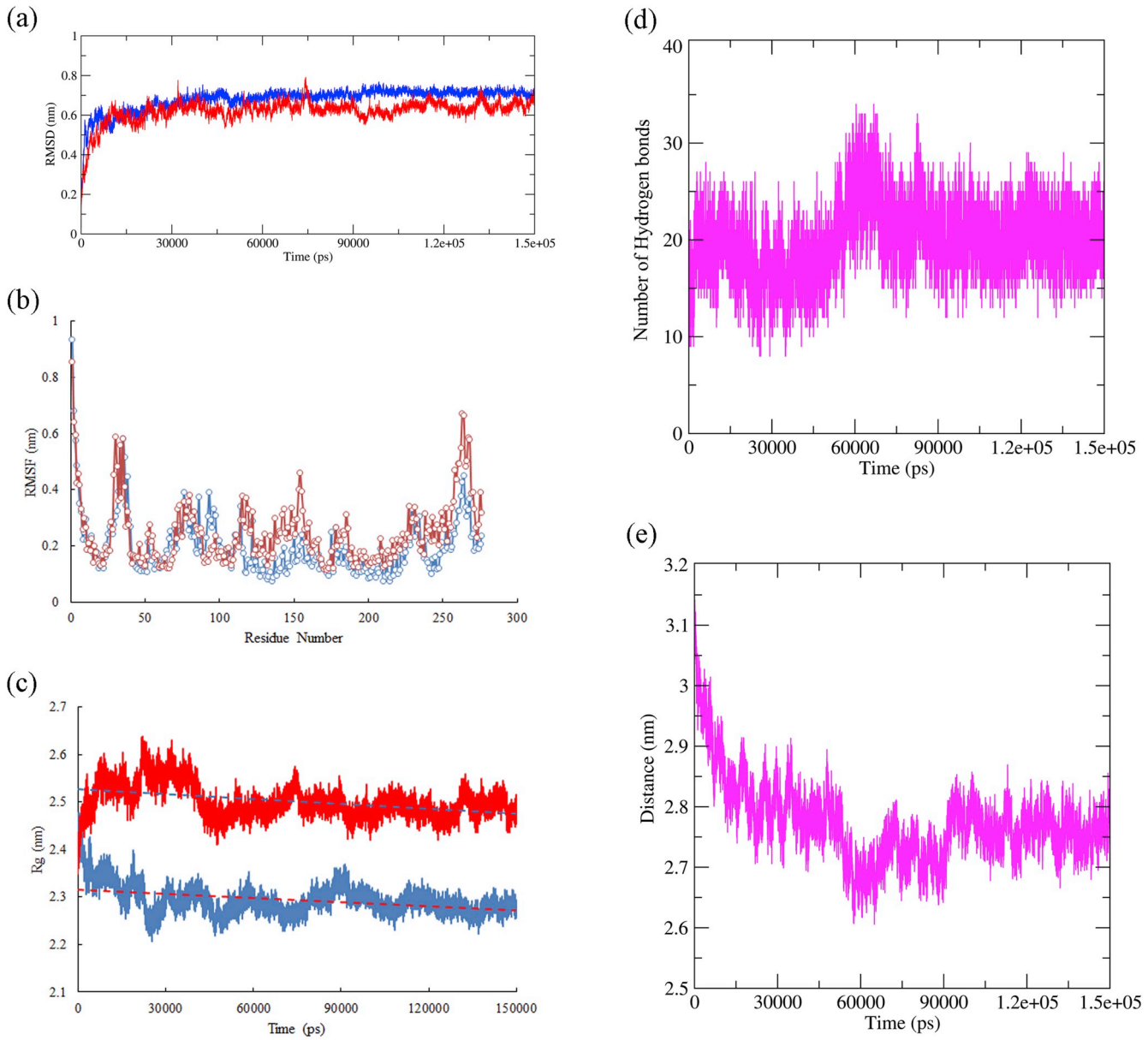

Fig. 5. (a) Backbone RMSD plots of two protein chains in homodimer structure, (b) RMSF plots of all the residues of two protein chains in homodimer structure, (c) Rg plot of two protein chains in homodimer structure, (d) The number of hydrogen bonds between two protein chains in homodimer structure during MD simulation, (e) Distance between the center of mass of two protein chains in homodimer structure during MD simulation.

Coiled-coil motifs, due to their high stability and well-defined structures have gained more attention in developing SAPN vaccine [60]. In recent years, because of unique characterizations of SAPN, including biodegradability, biocompatibility, size similarity to pathogen [61], self-adjuvanticity, and containing repetitive structure [62], they have been proposed as prominent structure in designing epitope-based peptide vaccine [10,63]. Each SAPN monomer, the subunit that oligomerizes to form the nanoparticle, contains two coiled-coil domains fused together by proper linkers, which are able to oligomerize, and consequently, form nanoparticles. In the context of SAPN formation mechanism, two α-helical coiled-coil domains, the pentameric coiled-coils that is derived from an N-terminal peptide of cartilage oligomeric matrix protein, and the trimeric coiled-coils that are made through de novo designed C-terminal leucine zipper peptide are linked by two glycine residues to form higher-order nanoparticles; these domains are self-assembled by a combination of hydrophobic and electrostatic interactions existing within the residues of coiled-coil domains [64,65]. In recent years, different SAPN-based vaccines have been designed and evaluated against various pathogens, including *Toxoplasma gondii*, human immunodeficiency viruses (HIV), *Plasmodium falciparum*, severe acute respiratory syndrome (SARS) coronavirus [66–71]. Although Chen et al. study, showed the trivalent antigen combination of YLN (protective peptide epitopes from CbpA) and L460D (fragment of PspA) as epitope vaccine elicited the strong and broad protection in diverse pneumococcal challenge models [72], but there isn't any report on the production of epitope-based SAPN vaccine against *S. pneumoniae*.

In our study, for the first time, the antigenic B-cell epitopes are added to either the N or C terminal ends of the gene encoding the monomer. In the same way T-cell epitopes were incorporated into the vaccine construct that can activate both the humoral and cellular arms of the immune system. Besides, a trimmer and pentamer domains were used as the scaffold of SAPN (Fig. 1) and the selected T- and B-cell epitopes were fused to monomer by proper linkers; the applied linkers can enhance proteasome-mediated cleavage, binding to TAP transporter processes, and playing a role as helical linker for increasing the

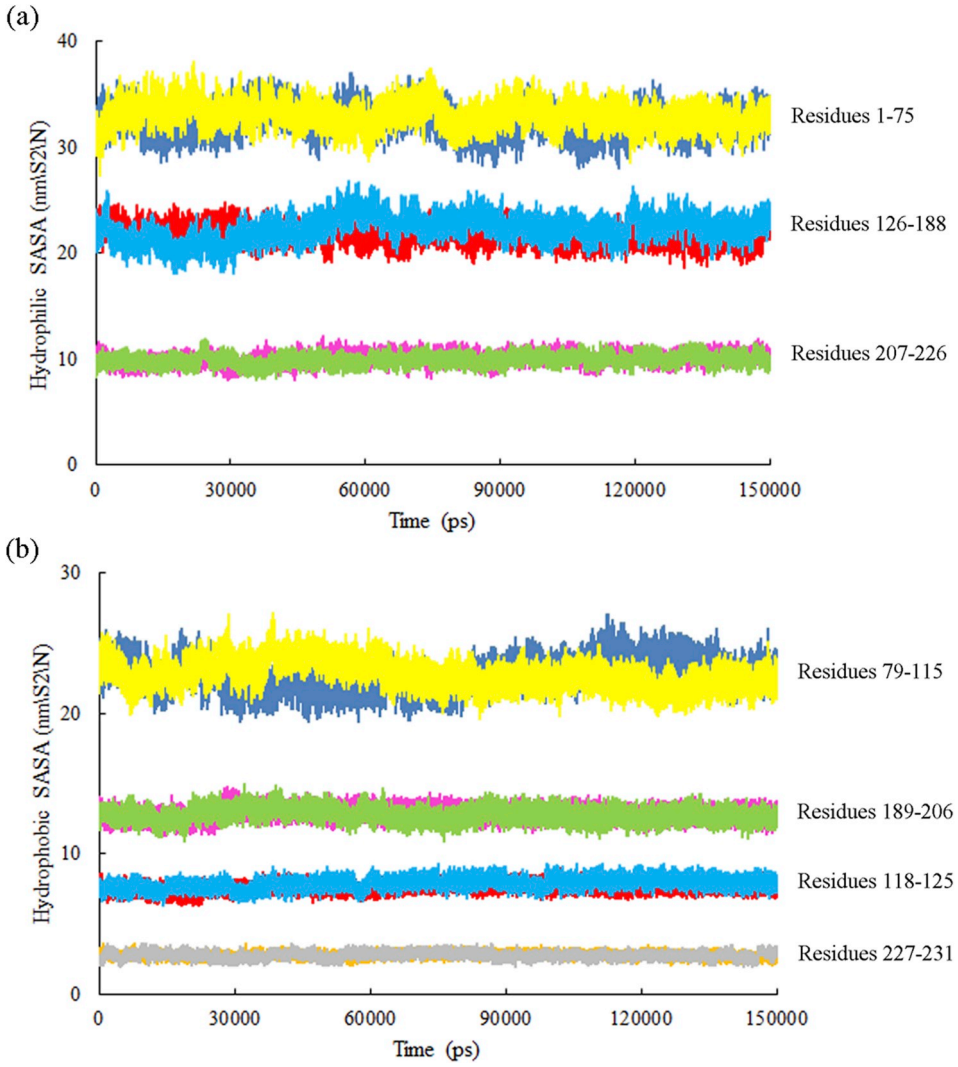

Fig. 6. (a) Hydrophilic solvent accessible surface area (SASA) of the epitope parts of protein chains in homodimer structure, (b) Hydrophobic solvent accessible surface area (SASA) of PENTAMER, and TRIMER parts of protein chains in homodimer structure. Dark blue, red, magenta, orange colors have been used for chain A. Yellow, blue, green, grey colors have been used for chain B.

immunogenicity of vaccine construct [73–75].

The physicochemical and immunological properties of our peptide vaccine was evaluated by different computational tools, which indicated that it is the structurally stable, antigenic and non-allergenic protein. The population average of the scaled solubility value (Query Sol) in protein-sol server was 0.45, and any Query Sol greater than 0.45 is predicted as a high soluble protein [76]. Therefore, as shown in Table 2, the selected protein is soluble (Query sol = 0.69), and the result was also confirmed by sol-pro (0.79912). The aliphatic index of protein was 80.72, showed that our protein is thermostable. The instability index of the protein is 25.78, that is less than 40, which shows our protein is stable in the test tube. The GRAVY index of our vaccine is −0.537 that presents the protein has an efficient interaction with water. The four different arrangements of vaccine were designed (Table 1), afterwards they modeled and refined by various servers to achieve high-quality models. Among all models, the best model was chosen through validation by Ramachandran Plot, ERRAT, VERIFY 3D, and Prosa-web servers. The high-quality 3D model, was applied for the identification of conformational B-cell epitopes and docking process. According to GalaxyHomomer docking score = 1220.742, the 2-mer subunit is the best number of the subunit for our protein; however, the best model was refined and subjected to MD simulation studies. According to the result of DiscoTope server, the final 3D model of vaccine had 97 conformational B-cell epitope residues (Table 6), which confirmed the vaccine is highly potent in stimulating humoral immunity. The study of secondary structure and exposability of amino acids to solvent demonstrated T and B-cell epitopes regions in final construct are highly exposed to solvent than trimmer and pentamer regions (Fig. 2), which can be concluded the pentamer and trimmer segments are located in the core region of vaccine, and the epitopes parts are located on the surface of oligomer. Additionally, evaluation of the solvent accessible surface (SASA) parameter that measured during MD simulation (Fig. 6) indicates epitope parts have preferred to interact with solvent molecules (hydrophilic part) during 150 ns MD simulation while pentamer, and trimmer parts have inclined to establish hydrophobic interactions during this time. Evaluation of the R_g values during MD simulation indicated decreasing Rg values that mean decreasing the distance of atoms from the center of mass of the protein, so that confirmed the protein chain has been folded during this time. However, superimposition of initial configuration and end of simulation time of MD simulation confirmed the stability of the designed monomeric chain during 150 ns MD simulation time. On the other hand, homodimer structure that was chosen by docking process showed appropriate stability in the course of this time as well. To obtain the best position relative to each other, two protein chains of homodimer structure tried to establish more stable intermolecular interactions and get better

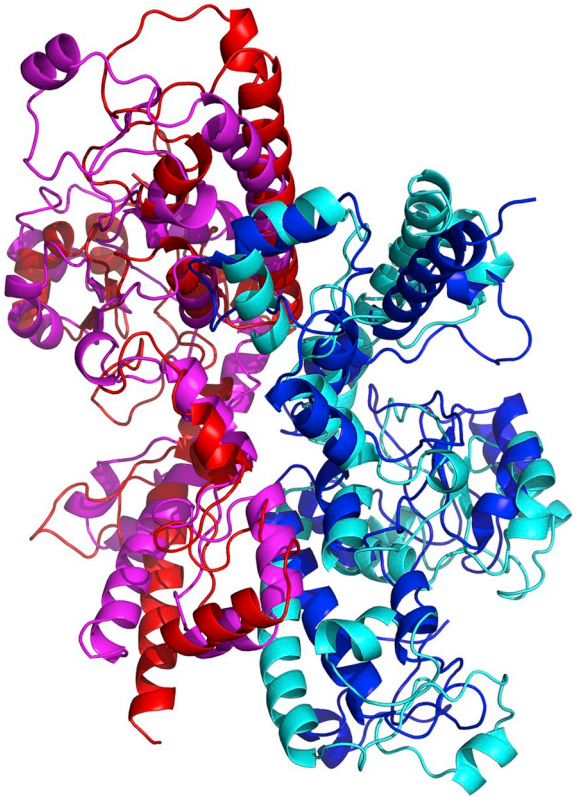

Fig. 7. Superimposition of docking model of homodimer structure (chain A and chain B are shown in cyan and magenta, respectively) over its structure obtained at the end of MD simulation time (chain A and chain B are shown in blue and red, respectively).

configuration during MD simulation time.

5. Conclusion

In this study, in order to design a self-assembled peptide nanovaccine against pneumoniae, the immunodominant CD4 and CD8 epitopes were selected from PspA, CbpA, PiuA, PhtD of *S. pneumoniae* antigens. DTD as T-helper was incorporated to vaccine to increase the immunogenicity of vaccine; moreover, the pentamer and trimer oligomeric domains as the coiled-coil self-assembled structures were employed in vaccine construct. In sum, in our study different strategies were applied to develop protective SAPN vaccine against *S. pneumoniae*; however, the potency of designed vaccine should be confirmed via different *in vitro* and *in vivo* immunological assays.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors wish to thank Shiraz University of Medical Sciences (13435) for supporting the conduct of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mcp.2019.101446.

References

[1] C. Von Mollendorf, S. Tempia, A. Von Gottberg, S. Meiring, V. Quan, C. Feldman, J. Cloete, S.A. Madhi, K.L. O'Brien, K.P. Klugman, Estimated severe pneumococcal disease cases and deaths before and after pneumococcal conjugate vaccine introduction in children younger than 5 years of age in South Africa, PLoS One 12 (7) (2017) e0179905.
[2] C.C. Daniels, P.D. Rogers, C.M. Shelton, A review of pneumococcal vaccines: current polysaccharide vaccine recommendations and future protein antigens, J. Pediatr. Pharmacol. Ther. : JPPT : the official journal of PPAG 21 (1) (2016) 27–35.
[3] A.E. Bridy-Pappas, M.B. Margolis, K.J. Center, D.J. Isaacman, Streptococcus pneumoniae: description of the pathogen, disease epidemiology, treatment, and prevention, Pharmacotherapy 25 (9) (2005) 1193–1212.
[4] M.E. Pichichero, M.N. Khan, Q. Xu, Next generation protein based Streptococcus pneumoniae vaccines, Hum. Vaccines Immunother. 12 (1) (2016) 194–205.
[5] A. Vila-Corcoles, Advances in pneumococcal vaccines: what are the advantages for the elderly? Drugs Aging 24 (10) (2007) 791–800.
[6] E.N. Miyaji, C.F. Vadesilho, M.L. Oliveira, A. Zelanis, D.E. Briles, P.L. Ho, Evaluation of a vaccine formulation against Streptococcus pneumoniae based on choline-binding proteins, Clin. Vaccine Immunol. : CVI 22 (2) (2015) 213–220.
[7] J. Hamel, N. Charland, I. Pineau, C. Ouellet, S. Rioux, D. Martin, B.R. Brodeur, Prevention of pneumococcal disease in mice immunized with conserved surface-accessible proteins, Infect. Immun. 72 (5) (2004) 2659–2670.
[8] R.H. Whalan, S.G. Funnell, L.D. Bowler, M.J. Hudson, A. Robinson, C.G. Dowson, Distribution and genetic diversity of the ABC transporter lipoproteins PiuA and PiaA within Streptococcus pneumoniae and related streptococci, J. Bacteriol. 188 (3) (2006) 1031–1038.
[9] T. Farhadi, N. Nezafat, Y. Ghasemi, Z. Karimi, S. Hemmati, N. Erfani, Designing of complex multi-epitope peptide vaccine based on omps of Klebsiella pneumoniae: an in silico approach, Int. J. Pept. Res. Ther. 21 (3) (2015) 325–341.
[10] M. Negahdaripour, N. Golkar, N. Hajighahramani, S. Kianpour, N. Nezafat, Y. Ghasemi, Harnessing self-assembled peptide nanoparticles in epitope vaccine design, Biotechnol. Adv. 35 (5) (2017) 575–596.
[11] R.J. Collier, Diphtheria toxin: mode of action and structure, Bacteriol. Rev. 39 (1) (1975) 54–85.
[12] S. Weijzen, S.C. Meredith, M.P. Velders, A.G. Elmishad, H. Schreiber, W.M. Kast, Pharmacokinetic differences between a T cell-tolerizing and a T cell-activating peptide, J. Immunol. 166 (12) (2001) 7151–7157.
[13] N.K. Mehta, K.D. Moynihan, D.J. Irvine, Engineering new approaches to cancer vaccines, Cancer immunology research 3 (8) (2015) 836–843.
[14] W. Chen, L. Huang, Induction of cytotoxic T-lymphocytes and antitumor activity by a liposomal lipopeptide vaccine, Mol. Pharm. 5 (3) (2008) 464–471.
[15] D.J. Irvine, M.C. Hanson, K. Rakhra, T. Tokatlian, Synthetic nanoparticles for vaccines and immunotherapy, Chem. Rev. 115 (19) (2015) 11109–11146.
[16] E. Gasteiger, C. Hoogland, A. Gattiker, S.e. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, Totowa, NJ, 2005, pp. 571–607.
[17] M. Hebditch, R. Curtis, S. Charonis, M.A. Carballo-Amador, J. Warwicker, Protein–Sol: a web tool for predicting protein solubility from sequence, Bioinformatics 33 (19) (2017) 3098–3100.
[18] M. Hebditch, M.A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwicker, Protein-Sol: a web tool for predicting protein solubility from sequence, Bioinformatics 33 (19) (2017) 3098–3100.
[19] S. Wang, W. Li, S. Liu, J. Xu, RaptorX-Property: a web server for protein structure property prediction, Nucleic Acids Res. 44 (W1) (2016) W430–W435.
[20] J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Res. 43 (W1) (2015) W174–W181.
[21] J. Ko, H. Park, L. Heo, C. Seok, GalaxyWEB server for protein structure prediction and refinement, Nucleic Acids Res. 40 (2012) W294–W297 (Web Server issue).
[22] W.-H. Shin, G.R. Lee, L. Heo, H. Lee, C. Seok, Prediction of protein structure and interaction by GALAXY protein modeling programs, Bio Design 2 (1) (2014) 1–11.
[23] L. Heo, H. Lee, C. Seok, GalaxyRefineComplex: refinement of protein-protein complex model structures driven by interface repacking, Sci. Rep. 6 (2016) 32153-32153.
[24] T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of < em > Escherichia coli < /em > proteins, Proc. Natl. Acad. Sci. 106 (11) (2009) 4201–4206.
[25] S.C. Lovell, I.W. Davis, W.B. Arendall 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: phi,psi and Cbeta deviation, Proteins 50 (3) (2003) 437–450.
[26] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396–404.
[27] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407–W410 Web Server issue.
[28] H. Park, G.R. Lee, L. Heo, C. Seok, Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments, PLoS One 9 (11) (2014) e113811.
[29] J. Lee, D. Lee, H. Park, E.A. Coutsias, C. Seok, Protein loop modeling by using fragment assembly and analytical loop closure, Proteins 78 (16) (2010) 3428–3436.
[30] M. Baek, T. Park, L. Heo, C. Park, C. Seok, GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure, Nucleic Acids Res. 45 (W1) (2017) W320–w324.

[31] H. Lee, H. Park, J. Ko, C. Seok, GalaxyGemini: a web server for protein homo-oligomer structure prediction based on similarity, Bioinformatics 29 (8) (2013) 1078–1080.

[32] J.V. Kringelum, C. Lundegaard, O. Lund, M. Nielsen, Reliable B cell epitope predictions: impacts of method development and improved benchmarking, PLoS Comput. Biol. 8 (12) (2012) e1002829.

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

[34] I.A. Doytchinova, D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinf. 8 (2007) 4.

[35] I. Dimitrov, L. Naneva, I. Doytchinova, I. Bangov, AllergenFP: allergenicity prediction by descriptor fingerprints, Bioinformatics 30 (6) (2014) 846–851.

[36] S. Saha, G.P.S. Raghava, AlgPred: prediction of allergenic proteins and mapping of IgE epitopes, Nucleic Acids Res. 34 (2006) W202–W209 (Web Server issue).

[37] M.G. Pikkemaat, A.B. Linssen, H.J. Berendsen, D.B. Janssen, Molecular dynamics simulations as a tool for improving protein stability, Protein Eng. 15 (3) (2002) 185–192.

[38] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Mol. Biol. 9 (9) (2002) 646–652.

[39] M. Abraham, D. Van Der Spoel, E. Lindahl, B. Hess, The GROMACS development team, GROMACS User Manual Version 5 (2) (2014).

[40] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D.E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, Proteins: Structure, Function, and Bioinformatics 78 (8) (2010) 1950–1958.

[41] M. Eslami, M. Negahdaripour, M.B. Ghoshoon, A. Gholami, R. Heidari, A. Dehshahri, N. Erfani, N. Nezafat, Y. Ghasemi, Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine AU - dorosti, Hesam, J. Biomol. Struct. Dyn. (2019) 1–12.

[42] M.Y. Lobanov, N. Bogatyreva, O. Galzitskaya, Radius of gyration as an indicator of protein structure compactness, Molecular Biology 42 (4) (2008) 623–628.

[43] K.L. O'Brien, L.J. Wolfson, J.P. Watt, E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mulholland, O.S. Levine, T. Cherian, Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates, The Lancet 374 (9693) (2009) 893–902.

[44] L. Wu, X. Yin, L. Zheng, J. Zou, P. Jin, Y. Hu, T. Kudinha, F. Kong, X. Chen, Q. Wang, Practical prediction of ten common Streptococcus pneumoniae serotypes/serogroups in one PCR reaction by multiplex ligation-dependent probe amplification and melting curve (MLPA-MC) assay in shenzhen, China, PLoS One 10 (7) (2015) e0130664.

[45] R. Cherazard, M. Epstein, T.L. Doan, T. Salim, S. Bharti, M.A. Smith, Antimicrobial resistant Streptococcus pneumoniae: prevalence, mechanisms, and clinical implications, Am. J. Therapeut. 24 (3) (2017) e361–e369.

[46] T.U. Yang, E. Kim, Y.-J. Park, D. Kim, Y.H. Kwon, J.K. Shin, O. Park, Successful introduction of an underutilized elderly pneumococcal vaccine in a national immunization program by integrating the pre-existing public health infrastructure, Vaccine 34 (13) (2016) 1623–1629.

[47] D.M. Weinberger, R. Malley, M. Lipsitch, Serotype replacement in disease after pneumococcal vaccination, The Lancet 378 (9807) (2011) 1962–1973.

[48] J.A. Herbert, E.J. Kay, S.E. Faustini, A. Richter, S. Abouelhadid, J. Cuccui, B. Wren, T.J. Mitchell, Production and efficacy of a low-cost recombinant pneumococcal protein polysaccharide conjugate vaccine, Vaccine 36 (26) (2018) 3809–3819.

[49] A. Sobanjo-ter Meulen, T. Vesikari, E.A. Malacaman, S.A. Shapiro, M.J. Dallas, P.A. Hoover, R. McFetridge, J.E. Stek, R.D. Marchese, J. Hartzel, W.J. Watson, L.K. Musey, Safety, tolerability and immunogenicity of 15-valent pneumococcal conjugate vaccine in toddlers previously vaccinated with 7-valent pneumococcal conjugate vaccine, Pediatr. Infect. Dis. J. 34 (2) (2015) 186–194.

[50] K.L. Moffitt, P. Yadav, D.M. Weinberger, P.W. Anderson, R. Malley, Broad antibody and T cell reactivity induced by a pneumococcal whole-cell vaccine, Vaccine 30 (29) (2012) 4316–4322.

[51] S.S. Tai, Streptococcus pneumoniae protein vaccine candidates: properties, activities and animal, Studies 32 (3) (2006) 139–153.

[52] S.S. Tai, Streptococcus pneumoniae protein vaccine candidates: properties, activities and animal studies, Crit. Rev. Microbiol. 32 (3) (2006) 139–153.

[53] H. Dorosti, M. Eslami, M. Negahdaripour, M.B. Ghoshoon, A. Gholami, R. Heidari, A. Dehshahri, N. Erfani, N. Nezafat, Vaccinomics Approach for Developing Multi-Epitope Peptide Pneumococcal Vaccine, (2019), pp. 1–12.

[54] G. Gámez, A. Castro, A. Gómez-Mejia, M. Gallego, A. Bedoya, M. Camargo, S. Hammerschmidt, The variome of pneumococcal virulence factors and regulators, BMC Genomics 19 (1) (2018) 10.

[55] W.P. Hanage, N.J. Croucher, F. Yang, K.P. Klugman, C. Chewapreecha, J. Parkhill, S.R. Harris, S.D. Bentley, B. Beall, L. McGee, M. van der Linden, J.-H. Song, K.S. Ko, R. Sá-Leão, H. de Lencastre, C. Turner, P. Turner, Evidence for soft selective sweeps in the evolution of pneumococcal multidrug resistance and vaccine escape, Genome Biology and Evolution 6 (7) (2014) 1589–1602.

[56] B.M. Diethelm-Okita, D.K. Okita, L. Banaszak, B.M. Conti-Fine, Universal epitopes for human CD4+ cells on tetanus and diphtheria toxins, J. Infect. Dis. 181 (3) (2000) 1001–1009.

[57] M. Qi, X.E. Zhang, X. Sun, X. Zhang, Y. Yao, S. Liu, Z. Chen, W. Li, Z. Zhang, J. Chen, Intranasal nanovaccine confers homo-and hetero-subtypic influenza protection, Small 14 (13) (2018) 1703207.

[58] C. Sawaengsak, Y. Mori, K. Yamanishi, A. Mitrevej, N. Sinchaipanid, Chitosan nanoparticle encapsulated hemagglutinin-split influenza virus mucosal vaccine, AAPS PharmSciTech 15 (2) (2014) 317–325.

[59] W. Ai, Y. Yue, S. Xiong, W. Xu, Enhanced protection against pulmonary mycobacterial challenge by chitosan-formulated polyepitope gene vaccine is associated with increased pulmonary secretory IgA and gamma-interferon+ T cell responses, Microbiol. Immunol. 57 (3) (2013) 224–235.

[60] Y. Yang, P. Ringler, S.A. Müller, P. Burkhard, Optimizing the refolding conditions of self-assembling polypeptide nanoparticles that serve as repetitive antigen display systems, J. Struct. Biol. 177 (1) (2012) 168–176.

[61] A.S. Cordeiro, M.J. Alonso, M. de la Fuente, Nanoengineering of vaccines using natural polysaccharides, Biotechnol. Adv. 33 (6) (2015) 1279–1293 Part 3.

[62] S. Howorka, Rationally engineering natural protein assemblies in nanobiotechnology, Curr. Opin. Biotechnol. 22 (4) (2011) 485–491.

[63] S. Raman, G. Machaidze, A. Lustig, U. Aebi, P. Burkhard, Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles, Nanomed. Nanotechnol. Biol. Med. 2 (2) (2006) 95–102.

[64] D.L. Lee, S. Ivaninskii, P. Burkhard, R.S. Hodges, Unique stabilizing interactions identified in the two-stranded α-helical coiled-coil: crystal structure of a cortexillin I/GCN4 hybrid coiled-coil peptide, Protein Sci. 12 (7) (2003) 1395–1405.

[65] M. Meier, J. Stetefeld, P. Burkhard, The many types of interhelical ionic interactions in coiled coils–an overview, J. Struct. Biol. 170 (2) (2010) 192–201.

[66] P. Burkhard, D.E. Lanar, Malaria vaccine based on self-assembling protein nanoparticles, Expert Rev. Vaccines 14 (12) (2015) 1525–1527.

[67] C.P. Karch, T. Doll, S.M. Paulillo, I. Nebie, D.E. Lanar, G. Corradin, P. Burkhard, The use of a P. falciparum specific coiled-coil domain to construct a self-assembling protein nanoparticle vaccine to prevent malaria, 15 (1) (2017) 62.

[68] S.A. Kaba, C.P. Karch, L. Seth, K.M.B. Ferlez, C.K. Storme, D.M. Pesavento, P.Y. Laughlin, E.S. Bergmann-Leitner, P. Burkhard, D.E. Lanar, Self-assembling protein nanoparticles with built-in flagellin domains increases protective efficacy of a Plasmodium falciparum based vaccine, Vaccine 36 (6) (2018) 906–914.

[69] K. El Bissati, Y. Zhou, D. Dasgupta, D. Cobb, J.P. Dubey, P. Burkhard, D.E. Lanar, R. McLeod, Effectiveness of a novel immunogenic nanoparticle platform for Toxoplasma peptide vaccine in HLA transgenic mice, Vaccine 32 (26) (2014) 3243–3248.

[70] T.A. Pimentel, Z. Yan, S.A. Jeffers, K.V. Holmes, R.S. Hodges, P. Burkhard, Peptide nanoparticles as novel immunogens: design and analysis of a prototypic severe acute respiratory syndrome vaccine, Chem. Biol. Drug Des. 73 (1) (2009) 53–61.

[71] N. Wahome, T. Pfeiffer, I. Ambiel, Y. Yang, O.T. Keppler, V. Bosch, P. Burkhard, Conformation-specific display of 4E10 and 2F5 epitopes on self-assembling protein nanoparticles as a potential HIV vaccine, Chem. Biol. Drug Des. 80 (3) (2012) 349–357.

[72] A. Chen, B. Mann, G. Gao, R. Heath, J. King, J. Maissoneuve, M. Alderson, A. Tate, S.K. Hollingshead, R.K. Tweten, Multivalent pneumococcal protein vaccines comprising pneumolysoid with epitopes/fragments of CbpA and/or PspA elicit strong and broad protection, Clin. Vaccine Immunol. 22 (10) (2015) 1079–1089.

[73] K. Guruprasad, S. Rajkumar, gβ-and gγ-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials, J. Biosci. 25 (2) (2000) 143–156.

[74] B. Livingston, C. Crimi, M. Newman, Y. Higashimoto, E. Appella, J. Sidney, A. Sette, A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes, J. Immunol. 168 (11) (2002) 5499.

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

[76] T. Niwa, B.W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins, Proc. Natl. Acad. Sci. U.S.A. 106 (11) (2009) 4201–4206.