

RESEARCH

Open Access



Immunoinformatics-aided design of a new multi-epitope vaccine adjuvanted with domain 4 of pneumolysin against *Streptococcus pneumoniae* strains

Mona Shafaghi^{1,2,3†}, Zohreh Bahadori^{1,2,3†}, Hamid Madanchi^{1,2,4}, Mohammad Mehdi Ranjbar⁵, Ali Akbar Shabani^{1,2*} and Seyed Fazlollah Mousavi^{3*}

[†]Mona Shafaghi and Zohreh Bahadori have contributed equally to this work.

*Correspondence: shabani@semums.ac.ir; mousavi@pasteur.ac.ir

¹ Department of Medical Biotechnology, Faculty of Medicine, Semnan University of Medical Sciences, Semnan, Iran

² Research Center of Biotechnology, Semnan University of Medical Sciences, Semnan, Iran

³ Department of Bacteriology, Pasteur Institute of Iran, Tehran, Iran

⁴ Drug Design and Bioinformatics Unit, Department of Medical Biotechnology, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran

⁵ Agricultural Research, Education, and Extension Organization (AREEO), Razi Vaccine and Serum Research Institute, Karaj, Iran

Abstract

Background: *Streptococcus pneumoniae* (Pneumococcus) has remained a leading cause of fatal infections such as pneumonia, meningitis, and sepsis. Moreover, this pathogen plays a major role in bacterial co-infection in patients with life-threatening respiratory virus diseases such as influenza and COVID-19. High morbidity and mortality in over one million cases, especially in very young children and the elderly, are the main motivations for pneumococcal vaccine development. Due to the limitations of the currently marketed polysaccharide-based vaccines, non-serotype-specific protein-based vaccines have received wide research interest in recent years. One step further is to identify high antigenic regions within multiple highly-conserved proteins in order to develop peptide vaccines that can affect various stages of pneumococcal infection, providing broader serotype coverage and more effective protection. In this study, immunoinformatics tools were used to design an effective multi-epitope vaccine in order to elicit neutralizing antibodies against multiple strains of pneumococcus.

Results: The B- and T-cell epitopes from highly protective antigens PspA (clades 1–5) and PhtD were predicted and immunodominant peptides were linked to each other with proper linkers. The domain 4 of Ply, as a potential TLR4 agonist adjuvant candidate, was attached to the end of the construct to enhance the immunogenicity of the epitope vaccine. The evaluation of the physicochemical and immunological properties showed that the final construct was stable, soluble, antigenic, and non-allergenic. Furthermore, the protein was found to be acidic and hydrophilic in nature. The protein 3D-structure was built and refined, and the Ramachandran plot, ProSA-web, ERRAT, and Verify3D validated the quality of the final model. Molecular docking analysis showed that the designed construct via Ply domain 4 had a strong interaction with TLR4. The structural stability of the docked complex was confirmed by molecular dynamics. Finally, codon optimization was performed for gene expression in *E. coli*, followed by in silico cloning in the pET28a(+) vector.



Conclusion: The computational analysis of the construct showed acceptable results, however, the suggested vaccine needs to be experimentally verified in laboratory to ensure its safety and immunogenicity.

Keywords: Immunoinformatics, Multi-epitope vaccine, Pneumococcal surface protein A (PspA), Pneumococcal histidine triad protein D (PhtD), Domain 4 of pneumolysin (Ply4), Protein TLR agonist adjuvant

Background

Streptococcus pneumoniae, a commensal Gram-positive bacterium, is one of the leading causes of death worldwide, causing a variety of infections such as pneumonia, meningitis, and bacteremia [1, 2]. The high-risk populations for these severe diseases include children under two years old, the elderly above 65 years old, and immunocompromised patients [3, 4]. Furthermore, viral infections such as influenza and coronavirus pose a high risk of developing severe invasive pneumococcal infections, resulting in high mortality rates [5, 6]. The high cost of antibiotic treatment and increasing the antibiotic resistance make vaccination against pneumococcus a particularly attractive intervention [7]. Currently, there are two kinds of pneumococcal vaccines based on a limited number of serotype-specific capsular polysaccharides; the 23-valent pneumococcal polysaccharide vaccine (PPV), which is consisted of 23 different capsular polysaccharides; and the 7-, 10-, or 13-valent pneumococcal conjugate vaccine (PCV), which is composed of the prevalent polysaccharides conjugated to a carrier protein [8, 9]. The PPV provides good coverage but does not protect at-risk groups under the age of two years, while the PCV induces efficient protective immunity in newborns but presents limited coverage and has a high production cost, and also requires several injections [10, 11]. However, the increase in pneumococcal infections caused by non-vaccine serotypes, due to serotype replacement, reduces the effectiveness of this group of vaccines [12, 13]. In response to these problems, it is necessary to design a novel vaccine that is more affordable and provides broad protection across all pneumococcal serotypes [13, 14]. Protein-based pneumococcal vaccines, containing a combination of the conserved proteins common in most or all of the strains, provide a promising alternative to the existing vaccines [14, 15]. In this context, in recent years, various protein factors involved in colonization and virulence have been investigated, and each of them could induce significantly different levels of protection in animal models and humans [16–18].

The pneumococcal surface protein A (PspA) is one of the most studied virulence factors that has been found in all clinical isolates of *S. pneumoniae* [19]. This antigen helps the pneumococcus escape the defense system of the host by interfering with the deposition of complement molecules on the surface of the bacteria and blocking the bactericidal activity of lactoferrin peptides [20, 21]. The PspA protein has three major domains (Additional file 1: Fig. S1); including (i) amino acids 1–288, an alpha helical domain (α -HD) consisting of A, A' and B regions; (ii) amino acids 289–370, a proline-rich domain (C region); and (iii) amino acids 371–571, a choline-binding domain responsible for surface attachment [22]. The α -HD and C regions in N-terminal of the protein are surface-exposed and can interact with the host immune system [23, 24]. The α -HD is variable and highly immunogenic, and it appears that protection is caused by epitopes in the 100 amino acids at its N-terminus (A region) and \sim 100 residues at its C-terminus

(B region) [24–26]. PspA is classified into six clades composing three families based on the sequence variability at B region designated as the clade-defining region (CDR) [22]. Family 1 consists of two clades (1 and 2), family 2 consists of three clades (3, 4 and 5), and family 3 has only one clade (6) which is found in 0.1–4% of strains [22]. According to different studies, more than 95% of pneumococcal isolates are in family 1 and family 2, and therefore efforts to develop PspA-based vaccines are focused on these two families [27, 28]. The level of cross-reactivity differs between different PspAs based on the degree of sequence similarities so that there is a greater cross-reactivity within the same clade [29]. Since in some studies it has been found that immune responses triggered by two major families of PspA are clade dependent [29–31], it is proposed that high antigenic regions from all their clades, having highest effect on cross-reactivity, should be included in PspA-based subunit vaccines [32]. Furthermore, the regions A and C in the PspA protein possess conserved epitopes which have the impact on cross-reactivity [31, 33].

Another important vaccine candidate is the Pneumococcal histidine triad protein D (PhtD), that belongs to the polyhistidine triad family and is characterized by the presence of five copies of the His triad motif (HxxHxH) [34]. This highly-conserved protein, which is expressed by all pneumococcal strains, inhibits complement deposition and mediates bacterial adherence by zinc binding [16]. The studies have revealed that the C-terminal fragment of PhtD (PhtD-C) is more surface-exposed and hence could be a more immunogenic region than other regions of the protein [35, 36]. The immunization with the truncated derivatives of PhtD-C was found to be more able to induce antibody responses and protective immunity than the immunization with the full length protein [37].

Adjuvants, which are important factors in vaccine development, are used to induce a faster, more efficient, and longer-lasting immune response [38]. Over the past few decades, protein toll-like-receptor (TLR) agonists have been studied as promising vaccine adjuvant candidates [39]. It has been shown that various bacterial protein TLR agonists exhibit adjuvant properties, such as the activation of TLR signaling, the production of pro-inflammatory cytokines, and maturation of antigen presenting cells. The peptide nature of these potential protein adjuvants provides unique properties, including the capability to modify the structure/function as necessary for minimal toxicity and high immunogenicity. Moreover, they can be genetically fused to peptide antigens that ensure the co-delivery of antigen-adjuvant simultaneously to the same cell, resulting to more effective activation of immune system [40]. Many pneumococcal proteins have been proven to be sensed by toll-like receptors, such as pneumolysin (Ply), DnaJ, RrgA pneumococcal pilus type 1 protein, and so on [40–42]. Studies have demonstrated that the C-terminal domain 4 of pneumolysin (Ply4) alone possesses TLR4 agonist activity [43]. Therefore, it can be suggested that Ply4 could be considered a potential vaccine adjuvant candidate, provided that modifications could be made to eliminate its virulence.

Bioinformatics approaches can assist researchers in various biological fields [44, 45]; particularly, for prediction of potentially immunoprotective epitopes to design candidate vaccines [46, 47]. Actually, the immuno-bioinformatics tools represent advantages over conventional vaccinology methods, including faster results and lower costs [48]. The present study is aimed at designing of a new multi-epitope based subunit vaccine to

elicit neutralizing antibody responses against *S. pneumoniae* using immunoinformatics approaches. In view of the importance of the proteins PspA from clades 1–5 (PspA1-5) and PhtD, the subunit vaccine containing proper B- and helper T-cell epitopes of these antigens may serve as an efficient vaccine candidate for a wide range of pneumococcal strains. In this study, it has been focused on the sequences of reference strains that are expected to exhibit a greater diversity than the other isolates [22]. The dominant epitopes were predicted and then the selected peptides were fused together using suitable linkers to construct the main core of the multi-epitope vaccine. Since the epitope vaccine may be quickly degraded by peptidases, the TLR agonist Ply4 was attached to the N-terminal of the construct as a potential adjuvant candidate to overcome this weakness.

Results

The overall workflow utilized in this study for developing the multi-epitope vaccine against *S. pneumoniae* is summarized in Fig. 1.

Retrieval of the protein sequences

The amino acid sequences of PspA1-5, PhtD and Ply were obtained from GenBank in FASTA format (Additional file 1: Table S1). A region of PspA2 (PspA2-A), B region of PspA1-5 (PspA1-5-B), C region of PspAs (PspA-C) and C-terminal of PhtD (from amino acid 383–853) were selected for B-cell and helper T-cell epitope predictions. Domain 4 of pneumolysin (from amino acid 360–471) was used as a potential TLR4 ligand-based adjuvant to enhance the immunogenicity of the construct. Three mutations Asp385Asn, Cys428Gly and Trp433Phe were done in the domain 4 to eliminate the virulence of pneumolysin.

Transmembrane domain prediction

The results of the TMHMM server demonstrated that the candidate sequences had no transmembrane domain (Additional file 1: Fig. S2).

3D structure prediction and validation of candidate antigens

The 3D structure of A region of PspA2 or PhtD-C was predicted through I-TASSER, and the model No. 1 with the highest score (C-score = - 0.70 or - 0.64, respectively) was refined with ModRefiner and GalaxyRefine (Additional file 1: Fig. S3). The validation results obtained using the SAVES server on the basis of PROCHECK and ERRAT are shown in Additional file 1: Fig. S4. Ramachandran plot of the refined model of PspA2 or PhtD-C indicated 82.1% or 87.2% of amino acids in favored regions, respectively. The ERRAT plot showed that the structure of A region of PspA2 or PhtD-C with overall quality factor 88.69% or 85.23%, respectively, can be acceptable. The tertiary structure modelling of B regions of PspA1-5 was done using Swiss-Model based on the template 2PMS with sequence identity more than 30% (Additional file 1: Fig. S5).

B cell epitopes prediction

The B cell epitopes were predicted by LBTope, ABCpred, Emini surface accessibility prediction tool of IEDB, Ellipro and DiscoTope servers. The epitopes of A region of PspA2 are listed in Additional file 1: Table S2 and the epitopes of B regions of PspA1-5 are

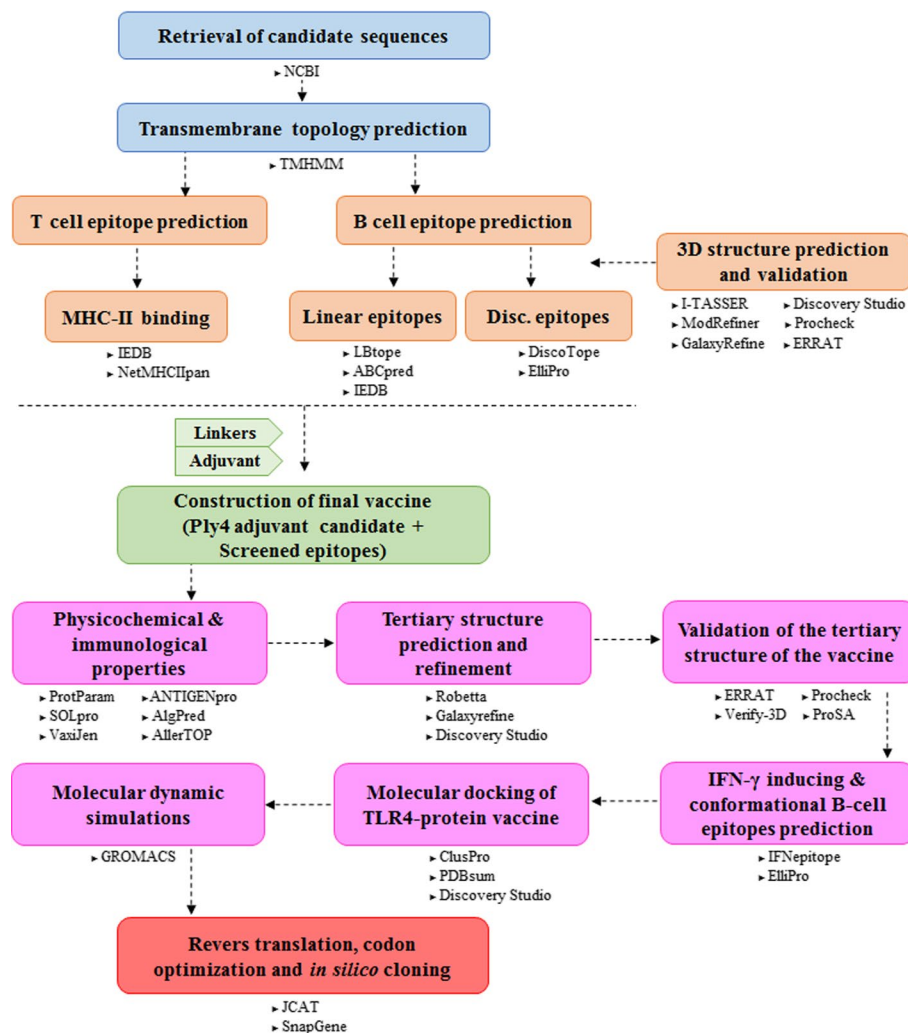


Fig. 1 Schematic flowchart of this study for developing the multi-epitope vaccine against *S. pneumoniae*. The entire approach employed in this study are represented in several phases and the used servers/softwares are mentioned below of each phase

represented in Additional file 1: Table S3–S7, respectively. The B cell epitopes of C region of PspAs were obtained considering only experimentally verified epitopes in recent studies (Additional file 1: Table S8) [23, 24, 49]. The predicted linear and conformational B cell epitopes of PhtD-C are shown in Additional file 1: Table S9 and S10, respectively.

Prediction of helper T cell (MHC-II) epitopes

The IEDB and NetMHCIIpan 4.0 servers were used to predict the epitopes of MHC class II (8 common human DRB1 alleles: 01:01, 03:01, 04:01, 07:01, 08:01, 11:01, 13:01, and 15:01, as well as 3 mouse alleles: H2-IAb, IAd, and IEd). The epitopes of A region of PspA2, B region of PspA1-5, and PhtD-C with higher binding affinity, based on percentile rank < 10.0 for IEDB and %Rank < 1.0 for NetMHCIIpan SBs, were considered for further analysis. The MHC-II-Binding epitopes predicted by IEDB and NetMHCIIpan for

PspA (PspA2-A and PspA1-5-B) and PhtD-C are shown in Additional file 1: Table S11 and S12, respectively.

Final vaccine sequence construction

Altogether, the high-scored B-cell and MHC-II epitopes shared between different servers were considered, and the 3D structures were used to choose the final suitable domains. The selected peptides of PspA and PhtD for the final construct are listed in Table 1. These peptides were merged together with the help of GPGPG linkers and the overall length of the main vaccine sequence was found to be 461 amino acids. To enhance the immunogenicity of the epitope vaccine, the domain 4 of pneumolysin was considered as an adjuvant candidate. The amino acid sequence of Ply4 containing three mutations D385N, C428G and W433F (112 amino acids long, Table 1) was attached to the N-terminal of the above peptides using an EAAAK linker. A 6xHis-tag was then added by GPGPG linker to the C-terminal of the designed vaccine construct, which could be helpful for the effective identification and purification of the protein. The final vaccine construct with a total of 590 amino acids consisting of the considered peptides of Ply, PhtD and PspA fused together with the suitable linkers is shown in Fig. 2A–B.

Evaluation of the various features of the designed vaccine

The ExPasy ProtParam results revealed that the molecular weight and theoretical pI value of the designed construct were 62.96 kDa and 4.60, respectively. The half-life was computed to be > 10 h in *E. coli*, > 20 h in yeast, and 30 h in mammalian reticulocytes. The instability index, aliphatic index and grand average of hydropathicity were predicted as 33.71, 59.63 and – 1.028, respectively. The protein solubility upon overexpression in

Table 1 The final selected peptides of PspA and PhtD, and the mutated sequence of domain 4 of Ply included in the vaccine construct

Protein	Region	Peptide sequence	Position
PspA	A Region	SPVASQSKAEKDYDAAVKSEAAKKAYE	3–30 (PspA2)
		VQQAYLAYQRASNKAEAAKMI	74–94 (PspA2)
	B Region	LKEIDESSEDIYVKEGFRAPLQSELD	182–207 (PspA1)
		KDVEDFKNSDGEQAGQYLAAEE	266–288 (PspA1)
		LKEIDSESEDIYAKEGFRAPLQSKLD	190–215 (PspA2)
		DQLKAVEENNNVEDYSTGLEK	243–264 (PspA2)
		KLLDNLDPGKKTQDELKAEAAEA	325–347 (PspA3)
		SNLEILLGGADPEDDTAALQNK	368–389 (PspA3)
		KVLATLDPGKKTQDELKAEAAEA	254–276 (PspA4)
		SKLEDNLKDAETNNVEDIYKEG	297–318 (PspA4)
		PEGKTQDELKAEAAEDANIEALQNKVA	258–284 (PspA5)
		RLQSDLKDAEENNNVEDIYKEGL	299–320 (PspA5)
	C Region	APKPEQPA	354–361 (PspA1) 426–433 (PspA3) 365–372 (PspA5)
PhtD	C-terminal of PhtD	AAQAYAKEKGLTPPSTDHQDSGNTAEKGAEIYNRVKAAK-KVPLDRMPYNLQYTVEVKNGS	559–619
		KPQTEKPEEDKEHDEVSEPTHPESDEKENHVGLNPSA	700–736
		NGDLLLDHSGAYVAQYYITWDELSYNHQGKEVLTTPKAW-DRNGQDLTAHFTTSIPLKGNVRNLSVKIREGTGLAFEW-WRTVYEKTDLPLVRKRTISIWGTTLYPQVEDKVEND	360–471

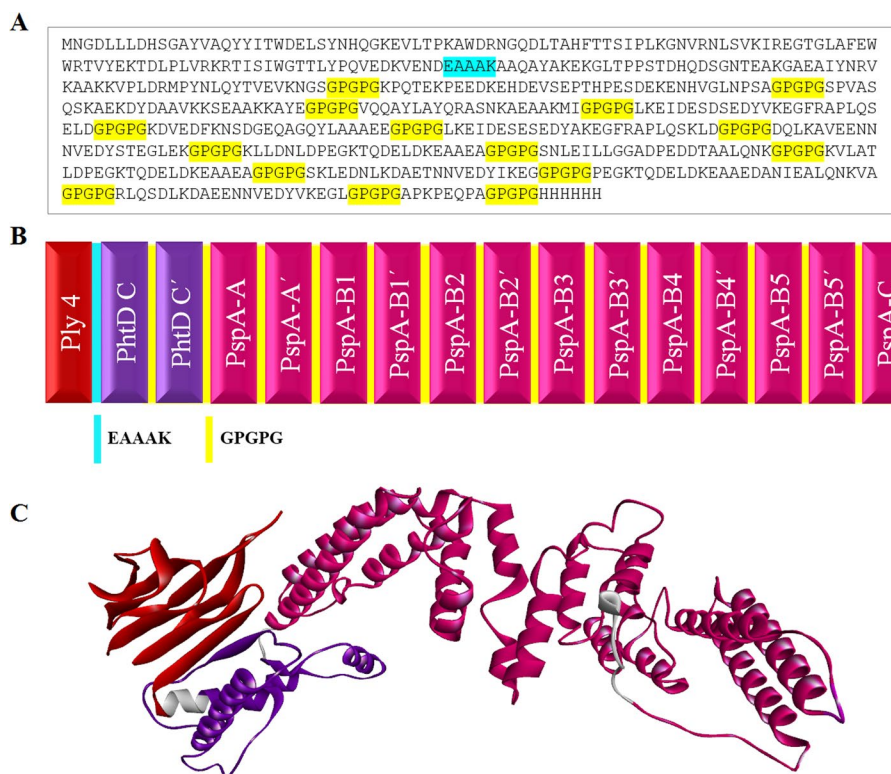


Fig. 2 The amino acid sequence, schematic diagram, and 3D refined structure of the vaccine candidate. **A** The final protein sequence of the multi-epitope construct consisted of 590 residues. The linkers EAAAK and GPGPG are shown in cyan and yellow, respectively. **B** The schematic image of vaccine construct was obtained from the selected regions of Ply4, PhtD and PspA, as well as linkers EAAAK and GPGPG, which are shown in red, purple, magenta, cyan and yellow, respectively. **C** The refined model was provided by the GalaxyRefine server and visualized by the Discovery Studio Visualizer. The regions of Ply4, PhtD and PspA are shown in red, purple, and magenta, respectively

E. coli was estimated 0.958393 by the SOLpro server, which indicated the protein construct was soluble. The antigenicity of the final construct was checked using the VaxiJen and ANTIGENpro servers to be 0.8833 and 0.850188, respectively. The prediction of allergenicity by AlgPred and AllerTOP servers demonstrated that the multi-epitope vaccine was a non-allergen.

Prediction, refinement, and validation of the 3D structure of the candidate vaccine

The tertiary structure of the designed multi-epitope construct was predicted by the Robetta server using a comparative modeling approach. The generated model number one was refined by the GalaxyRefine server. Among the five refined models, model number 1 (Fig. 2C) was selected as the best structure for further analysis. This model had a GDT-HA score of 0.9712, RMSD score of 0.380, MolProbity score of 1.591, clash score of 8.4, Poor rotamers of 0.2 and Rama favored score of 97.3. The crude and refined models were evaluated using the programs such as ERRAT, VERIFY-3D and PROCHECK from SAVES server, and also ProSA web server (Table 3). The ERRAT quality factor score of the crude model was predicted as 96.30, while the score of the refined model was calculated as 96.00 (Fig. 3A). VERIFY-3D value of the primary model was estimated as 80.68, which was increased to 83.39 in the refined model (Fig. 3B). The Ramachandran Plot

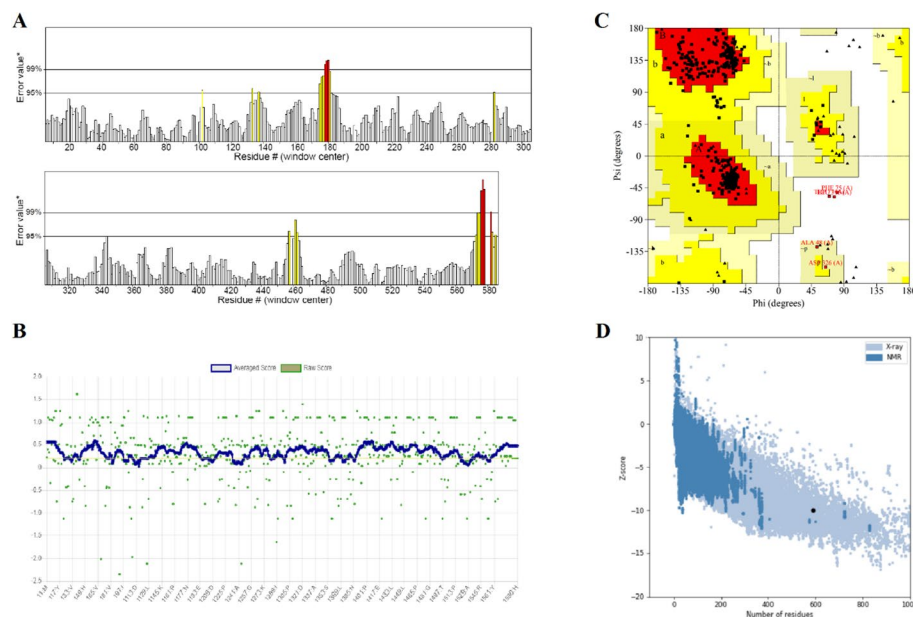


Fig. 3 Validation of the refined model of the multi-epitope vaccine. **A** The ERRAT quality factor of the refined model was estimated as 96.00. **B** The VERIFY-3D value of the refined model was calculated at 83.39. **C** Ramachandran plot of the refined model of vaccine showed 94.2%, 5.1% and 0.6% of residues in the favored, allowed, and disallowed regions, respectively. **(D)** Z-score of the 3D-structure of the refined model of vaccine was estimated as -10.02 , which lay in the range of scores reported for native protein structures

produced by PROCHECK demonstrated that in the crude model, 91.4%, 8.3% and 0.2% of residues were in the favored, allowed and disallowed regions, respectively, while in the refined model, 94.2%, 5.1% and 0.6% of residues were in the favored, allowed and disallowed regions, respectively (Fig. 3C). The Z-score obtained by the ProSA was found to be -9.83 in the crude model compared to -10.02 in the refined model (Fig. 3D). The results indicate that the final model has definitely a good quality.

IFN-inducing peptide prediction in the final proposed construct

The IFNepitope server was used to identify the epitopes that can activate T-helper cells for Interferon gamma production. A total number of 105 and 464 potential IFN- γ epitopes were predicted for the Ply4 adjuvant candidate and the main vaccine sequence, respectively. The positive epitopes with scores greater than or equal to one, a total of 29 and 45 epitopes from the Ply4 and the main vaccine sequence, respectively, are presented in Table 4.

Prediction of conformational B cell epitopes of the final designed protein

Since greater than 90% of B-cell epitopes are conformational/discontinuous, the 3D structure of the proposed multi-epitope construct was analyzed to identify these epitopes via the ElliPro server. Eleven new epitopes, comprising 3–114 residues were found with a score value of 0.518–0.776. The details of the identified conformational epitopes are given in Table 5. The 3D representation of the epitopes in the final construct and the 2D score chart are shown in Fig. 4 A and B, respectively.

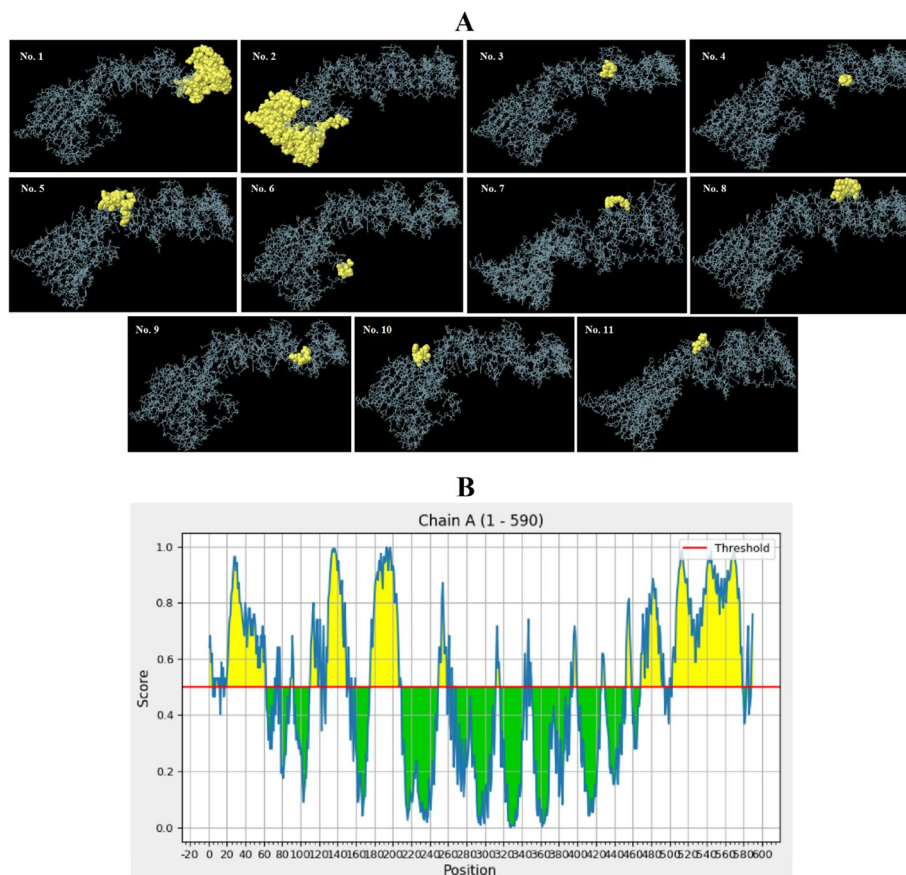


Fig. 4 3D images of the conformational B cell epitopes of the designed vaccine and 2D score chart. **A** The yellow and grey regions represent the conformational epitopes and the bulk of the polyprotein, respectively. **B** The yellow parts with a score above the threshold 0.5 display potential B cell epitopes

Vaccine-TLR-4 docking

Molecular docking of the refined model of vaccine with the TLR4 protein was done, and a total of 30 models were ranked based on the amount of energy produced using the ClusPro 2.0 server. The model M02 as the best-docked complex displayed the center weighted score of -740.8 kcal/mol and lowest energy score of -809.3 . As shown in Fig. 5, the TLR4 receptor and the vaccine protein interacted significantly through the domain 4 of Ply as the TLR ligand. PDBsum revealed that 15 residues of Ply4 were paired with 14 residues of the extracellular domain (ECD) of TLR4. The Ply4 established one salt bridge, five hydrogen bonds and 100 non-bonded contacts with the ECD of TLR4 (Fig. 5B). The salt-bridge between Ply4 and TLR4 ECD was formed among the residues Asp45 and Arg87, respectively, as well as the hydrogen bonds were formed between Ala74-Asn44, Trp77-Phe63, Thr71-Arg67, Asp45-Arg87 and Gln44-Lys186, at a distance of 3.01, 2.71, 2.93, 2.76 and 2.59 Å, respectively. This complex was applied as the initial configuration for the process of molecular dynamics simulation.

Molecular dynamics (MD) simulation of the vaccine in complex with the immune receptor

The stability of the designed vaccine in complex with TLR4 ECD was analyzed by performing MD simulation, and the results were evaluated in terms of RMSD and RMSE.

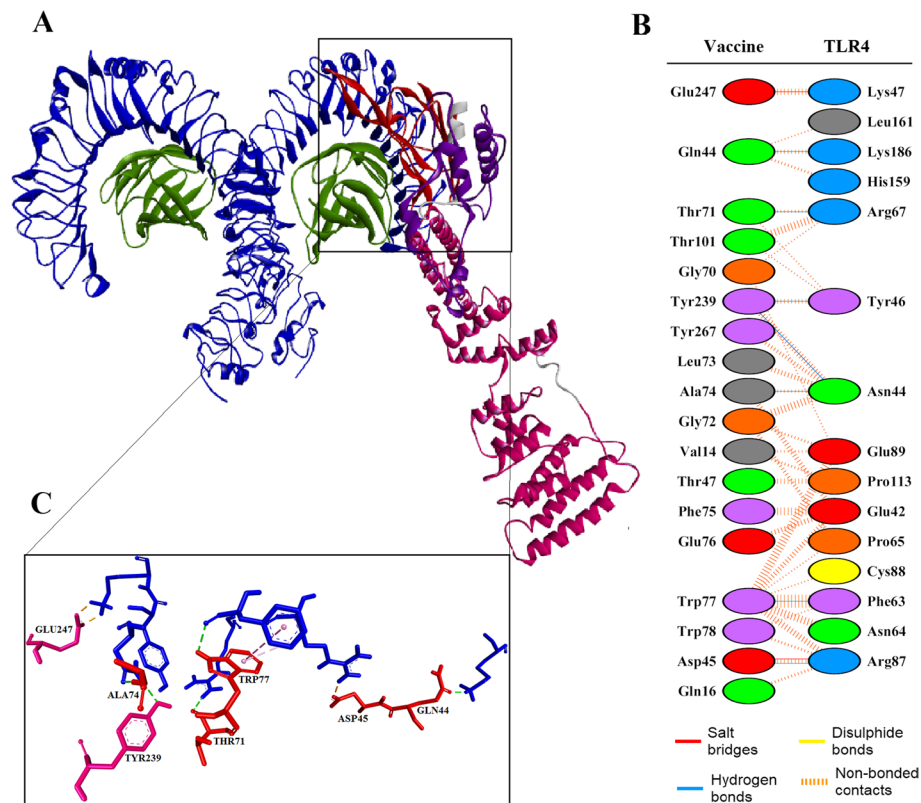


Fig. 5 Molecular docking of the designed vaccine with the TLR4 receptor. **A** The cartoon representation of the vaccine-TLR4 complex visualized by Discovery Studio Visualizer. **B** A number of 18 residues of the whole vaccine candidate paired with 15 residues of the TLR-4 ECD. The whole vaccine established 2 salt bridges, 7 hydrogen bonds and 116 non-bonded contacts with the receptor. **C** The residues involved in interactions among the vaccine and TLR4 ECD are represented as a stick model and colored in red/magenta and blue, respectively, while the vaccine residues are labeled and the interactions are shown with dashed lines

The RMSD profile displayed the changes in backbone C α atoms from the initial to final conformations of the protein complex. RMSD values of the vaccine-receptor complex indicated a sharp increase from the starting point of the simulation up to 6 ns, and after that until 20 ns remained steady with small fluctuations at an average of slightly lower than 1.5 nm (Fig. 6A). The calculations of RMSF of C α atoms determined the flexibility of all residues in the vaccine protein and TLR4 ECD. The results indicated that the most amino acids in the TLR4 ECD had small fluctuations and the minimum and maximum RMSF values were 0.1 and 0.3, respectively (Fig. 6B). Two regions of the vaccine molecule including amino acids in areas 179–181 and 551–555 showed relatively high fluctuations with RMSF of ~0.4 and 0.68 nm, respectively, indicating relatively high flexibility (Fig. 6C).

Reverse translation, codon optimization, and in silico cloning

The codon adaptation of the designed vaccine construct was done using JCat server with respect to the codon usage of *E. coli* strain K12. The total length of the target sequence was 1770 nucleotides. The Codon Adaptation Index (CAI) of the improved sequence was found to be 0.99, which is near to 1.0 indicating high level expression of the construct in this bacterial system. The GC-content was predicted to be 53.27%,

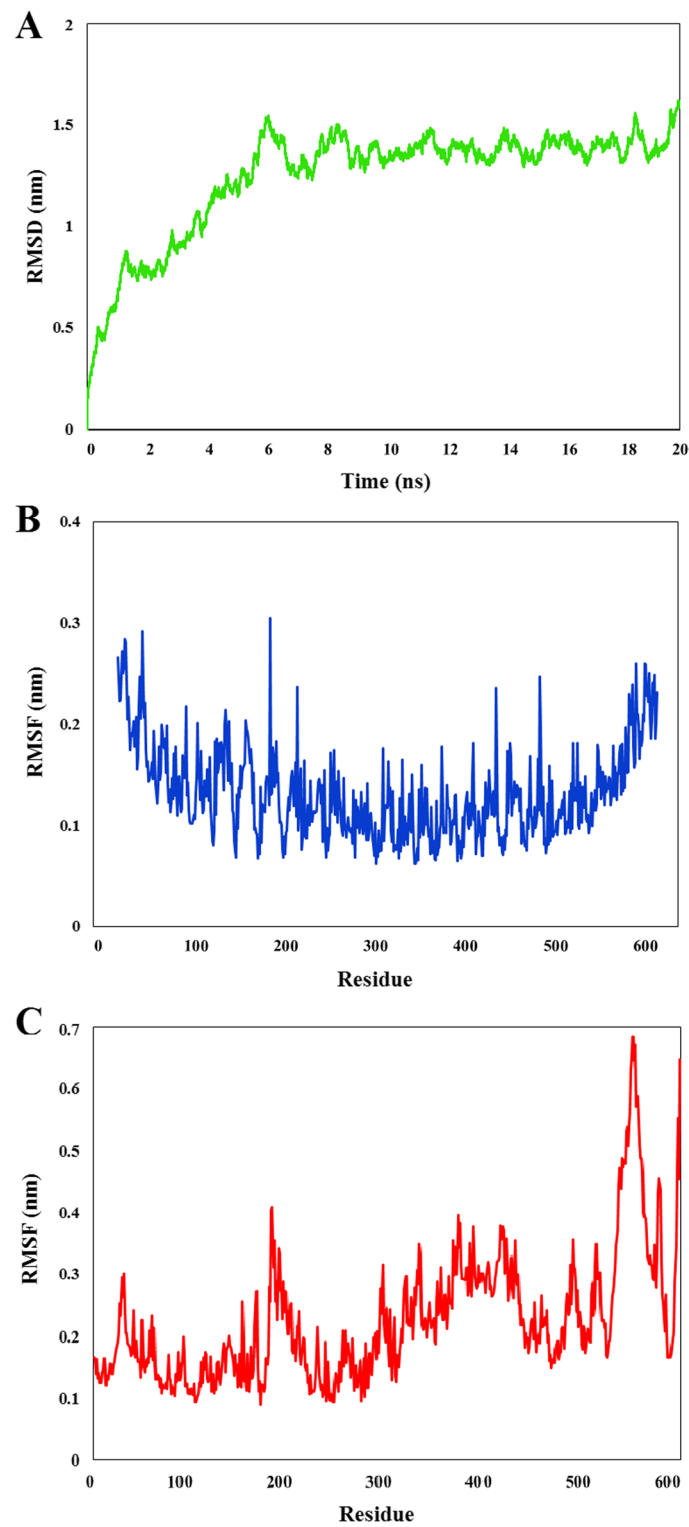


Fig. 6 The results of MD simulation of the vaccine-TLR4 docked complex. **A** RMSD values for the vaccine-receptor complex during a period of 20 ns. The complex became stable after 6 ns of MD simulation, which further confirmed the interaction between the vaccine and TLR4. **B** and **C** show the RMSF of all the residues of construct and TLR4 ECD, respectively. Most of the receptor residues underwent small fluctuations. Two regions of the vaccine indicated relatively high fluctuations, showing relatively high flexibility

which is within the optimal range (30–70) and shows a high probability of expression in the *E. coli* K12 strain (Additional file 1: Fig. S6). The adapted vaccine sequence was then inserted into the pET28a(+) expression vector between the restriction enzymes *NdeI* and *XhoI* using the SnapGene tool (Fig. 7).

Discussion

Pneumococcal infections are responsible for substantial morbidity and mortality, particularly in children, and it is believed that an effective vaccine can reduce the death rate [50, 51]. Since the existing polysaccharide-based vaccines are costly, serotype-dependent, and their immunogenicity is restricted to the serotypes included in the vaccines, the efforts of scientists have been focused on the development of serotype-independent vaccines [52]. Pneumococcal protein vaccines are currently being evaluated as interesting alternatives to available vaccines, that can induce serotype-independent immunity at a low cost [16–18]. The combination of multiple protein antigens in a vaccine seems like a very attractive strategy to block infection by targeting several important virulence factors [53]. Various studies have shown that a fusion protein comprising multiple antigens can be more effective than a mixed formulation, and also relatively simplify the purification procedure and facilitate product quality control [54–56].

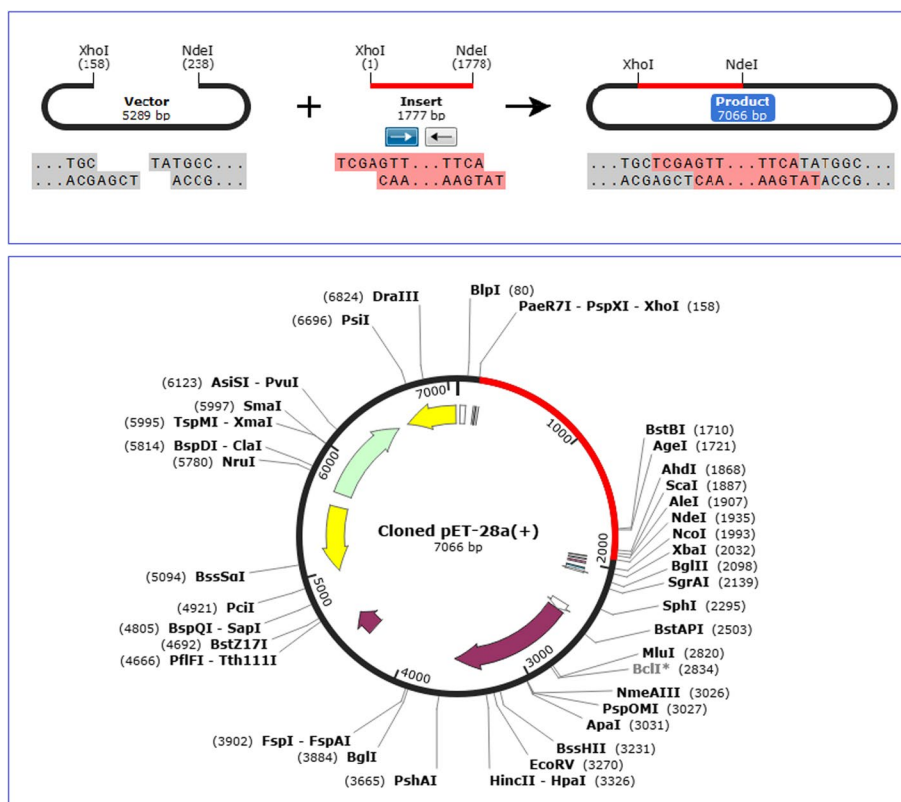


Fig. 7 In silico cloning of the target sequence in pET28a(+) vector plasmid. The vaccine sequence between the restriction sites is represented in red color while the vector backbone is shown in black. The total length of the clone is 7066 bp

Recent advances in immunoinformatics approaches could help to identify potential B- and T-cell epitopes on antigenic proteins for the construction epitope-based subunit vaccines and so to accelerate the vaccine development process [57–59]. In recent years, numerous researchers have designed novel vaccines using immunoinformatics tools against different pathogens like *Acinetobacter baumannii* [60], *Listeria monocytogenes* [61], *Leishmania donovani* [62], influenza A virus [63], and HPV16/18 [64]. Epitope-based peptide vaccines have many advantages compared to vaccines made by conventional technologies. For instance, they are a safer alternative, as they do not contain the complete pathogen and are more stable and highly specific. They have lower costs, do not require microbial culturing, and reduce the number of in vitro experiments, saving time [65]. Nevertheless, epitope vaccines have some limitations, in particular low immunogenicity due to their quick degradation by peptidases and lack of detection by immune receptors [66]. To conquer this issue several approaches can be employed such as developing adjuvanted vaccines that can improve the targeting of antigens to antigen-presenting cells and so prevent degradation of the antigen in vivo [67, 68].

Bacterial protein TLR agonists have been studied as attractive vaccine adjuvant candidates in recent decades, including several pneumococcal proteins, such as pneumolysin [40, 41]. A study from Chiu et al. has demonstrated that the truncated domain 4 of pneumolysin alone has two activities, stimulating TLR4 and causing hemolysis [43]. Besides, It has been shown that a triple mutant comprising Asp-385 to Asn, Cys-428 to Gly and Trp-433 to Phe substitutions could reduce cytolytic activity of the Ply protein [69]. Therefore, the domain 4 of Ply, along with the mutations D385N, C428G and W433F, could be considered as a potential candidate to be used as a TLR agonist adjuvant in order to interact with TLR4 and substantially increase immune responses to genetically fused peptide antigens. In this study, it was attempted to employ computational approaches for engineering a potential Ply4-adjuvanted multi-epitope vaccine comprising the immunodominant regions of PspAs from different clades and PhtD, as the most promising antigens, to trigger neutralizing antibodies against different strains of *S. pneumoniae*.

Numerous studies have demonstrated that recombinant PspAs could protect animal models from fatal pneumococcal infection [26, 70, 71] and also elicit strong antibody responses in human phase I clinical trials [72, 73]. However, the diversity of PspAs among clinical isolates could limit the coverage of PspA-based vaccines [30]. Based on analysis of amino acid sequence divergence, Hollingshead et al. classified most PspAs into two major families divided into 5 clades [72]. It has been shown that the cross-reactivity level between PspAs is dependent on the sequence identity, which is low among different PspA families and higher within each family. Moreover, some studies have suggested that the cross-reactivity/protection level differ depending on the PspA clade [30–32]. Akbari et al. developed a fusion PspA-based vaccine containing B region fragments of PspA from the five most prevalent clades (named: PspAB1–5). The vaccine could provide significant protection against different strains of pneumococci, indicating that the use of fusion proteins bearing B regions from clades 1–5 could be a promising immunization strategy. Moreover, since the A and C regions also play a significant role in the production of cross-reactive antibodies, it was suggested that including these regions in the construct could be helpful to expand the cross-protection of the vaccine [32]. In this

respect, in the current study, the regions A, B, and C of PspA were evaluated for prediction of their B- and helper T-cell epitopes. B regions from clades 1–5 were considered in order to provide broad coverage and reduce the likelihood of PspA variants escaping host immunity.

The PhtD protein has been shown to induce a broad protection against colonization and lethal infections in animal models [17, 74], and also to be safe and immunogenic in phase I clinical trials, rendering promising outcomes [18]. The studies have indicated that the N-terminal amino acids of PhtD are involved in surface attachment, while the C-terminal fragment of PhtD is a surface-exposed and protection-eliciting region [37, 75]. In a study by Plumptre et al., recombinant truncated derivatives of the C-terminal of PhtD were found to induce a very high titre of antibodies, showing that the region was highly immunogenic, while the full length PhtD was ineffective in providing significant protective immunity [37]. Therefore, herein, we analyzed the C-terminal of PhtD as a favorable immunogenic fragment to predict the B-cell and helper T-cell epitopes.

Based on the immunoinformatics results obtained in this research, the high-scoring B- and T-cell epitopes overlapping with each other shared between different servers were considered, and the final peptides from PspA (A/B region) and PhtD-C antigens were selected (Table 1). A comparative evaluation of the computationally predicted epitopes with the existing experimentally validated epitopes of PspA was done to control the accuracy of the present *in silico* prediction. The considered peptides of B region of PspA1 overlapped with the sequences ESDSEDYVK and SDGEQAGQYLAAEE reported as B-cell epitopes by Vadesilho et al. [24]. Also, one of the selected peptides of B region of clade3 was consistent with the experimentally identified epitope sequence DSEDDTAA. The chosen peptides for B regions of clades 4 and 5 overlapped with the epitope sequences NNVEDYIKEG and NNVEDYVKEG reported as B-cell epitopes by the same research group [24]. Further to these, the considered peptide with amino acids 190–215 of B region of clade 2 overlapped with the sequence EDYAKEGFRAPLQSK experimentally obtained as T-cell epitope by Singh et al. [76]. These further demonstrated the accuracy of the current computational prediction. For epitope prediction of C region of PspA, experimentally identified epitopes available on the IEDB server were used. The sequences of C region in different PspAs are mainly composed of short amino acid repeats [27]. It has been found that 46% of strains express a copy of the short repeat PKPEQP which is able to elicit protective antibodies in mice [23], therefore this significant motif was considered in this research. In the next step of this study, the screened peptides were linked to each other using appropriate linker, namely GPGPG, as flexible glycine-rich linkers can improve the solubility and allow the adjoining domains to be accessible and function freely [77, 78]. To generate whole vaccine, the sequence of Ply4, the considered potential TLR-4 agonist candidate, was added to the N-terminal end of the fused peptides using the EAAAK, as this linker provides rigidity and reduce possible interference of other regions of the protein in the interaction between the adjuvant and its receptor [79].

As shown in Table 2, physicochemical and immunological features of the construct were appraised using various computational servers. The molecular weight of the vaccine was 62.96 kDa, which makes it an acceptable construct because proteins with a molecular weight of less than 110 kDa can be more easily purified and so are thought to be more

Table 2 Physicochemical and immunological properties of the candidate vaccine

Molecular features	Prediction
No. of amino acids	590
Molecular weight (kDa)	62.96
Theoretical pI	4.60
Total no. of negatively charged residues (Asp + Glu)	115
Total no. of positively charged residues (Arg + Lys)	64
Estimated half-life	> 10 h (<i>Escherichia coli</i> , in vivo) > 20 h (yeast, in vivo) 30 h (mammalian reticulocytes, in vitro)
Instability index	33.71 (Stable)
Aliphatic index	59.63
Grand average of hydropathicity (GRAVY)	− 1.028
Solubility (SOLpro)	0.958393 (Soluble)
Antigenicity (VaxiJen)	0.8833 (Probable ANTIGEN)
Antigenicity (ANTIGENpro)	0.850188 (Probable ANTIGEN)
Allergenicity (AlgPred/AllerTOP)	Non-allergen

suitable targets for the vaccine development [80]. Also, this indicator is useful in SDS-PAGE electrophoresis and Western blot assays [81]. The theoretical pI of the vaccine was found to be rather acidic in nature (4.60), which could be beneficial for isoelectric focusing and purification of protein by ion exchange chromatography [81]. The half-life of the vaccine was found to be > 10, > 20, and 30 h in *E. coli*, yeast and mammalian, respectively, which shows the time taken by the protein to reach half of its amount after being synthesized in the cell [82]. The recombinant vaccine with instability index of 33.71 was classified as stable, since proteins with instability index less than 40 are considered to be stable and vice versa [82]. The calculated aliphatic index of multi-epitope protein was 59.63, indicating that the construct was thermostable [83]. The GRAVY index of the vaccine was a negative value (− 1.028), which indicated hydrophilic nature of the protein and its strong interaction with water molecules, suggesting high solubility [84]. For an antigen, greater interactions with water molecules reflect greater interactions with elements of the immune system, particularly with B cell receptors/antibodies [85]. Additionally, the analysis of solubility showed that the protein has a high percentage of solubility upon overexpression in *E. coli*. The antigenicity of the designed vaccine was evaluated to be 0.8833 and 0.850188 using VaxiJen and ANTIGENpro servers, respectively, showing the protein is antigenic in nature. The vaccine candidate was predicted to be a non-allergen molecule by two web servers AlgPred and AllerTOP.

The initial 3D structure of the vaccine candidate modeled by Robetta server was subjected for refinement, by GalaxyRefine software, to obtain a more high-quality 3D model. The evaluation of the initial and refined structures was performed using Ramachandran plot, ProSA Z-score, Errat score and Verify3D score. The obtained results showed that the quality of tertiary structure of vaccine has been improved by refinement (Table 3). The Ramachandran plot after refinement indicated that most of the residues were found in the favoured regions (94.2%), showing that the quality of the model was satisfactory. In addition, after refinement, the Z-score of − 10.02, ERRAT score of 96.00 and Verify3D score of 83.39 further validated the overall quality of the multi-epitope vaccine.

Table 3 Evaluation results of the modeled vaccine before and after refinement. Comparisons were performed by various servers ERRAT, Verify3D, PROCHECK, and ProSA

Server	Before refinement	After refinement
ERRAT	96.30	96.00
Verify3D	80.68	83.39
PROCHECK	Most favoured regions, 91.4% Additional & generously allowed regions, 8.3% Disallowed regions, 0.2%	Most favoured regions, 94.2% Additional & generously allowed regions, 5.1% Disallowed regions, 0.6%
ProSA	− 9.83	− 10.02

Identification of interferon- γ inducing peptides and also conformational B-cell epitopes was performed after the evaluation and confirmation of the construct. IFNepitope server predicted 74 epitopes (Table 4) in the construct's sequence that could induce IFN- γ which has been reported as one of essential factors involved in combating fatal pneumococcal infection [86]. The ElliPro server predicted eleven potential non-linear B-cell epitopes in the final 3D model of vaccine with 3–114 residues and qualifying scores of 0.518–0.776 (Table 5). The results confirm that the vaccine is capable to stimulate the humoral immune response which is essential for protection against pneumococcus. Since the TLR4 agonist Ply4 was used as a potential immune-adjuvant in the vaccine, a molecular docking analysis was performed to assess potential immune interaction between the multi-epitope vaccine and TLR4 (Fig. 5). The results confirmed that the designed construct via Ply domain 4 had affinity for TLR4, with the lowest energy scores of − 809.3. The docking analysis indicated that 2 salt bridges and 7 hydrogen bonds were formed during the interaction. To check the stability of vaccine-TLR4 complex, a molecular dynamics simulation was done using the best docked model. According to Fig. 6A, there were very mild fluctuations in the RMSD graph from 6 until 20 ns, suggesting the complex's stability. The RMSF graphs showed the most residues underwent small fluctuations, except two regions of the vaccine (Fig. 6B–C). Based on the obtained results, our developed vaccine can maintain a sustained interaction with TLR4 and thus has the chance to induce potentially protective immune responses. To assure a high-level expression in *E. coli* (strain K12), codon optimization of the proposed construct was performed using Jcat server. The expression of protein correlates with the total GC content and CAI value of the optimized reverse translated sequence [87]. Both results of the GC content (53.27%) and CAI value (0.99) were favorable for high-level expression of the multi-epitope protein in the bacteria. Finally, the vaccine sequence was successfully ligated into the pET28a(+) expression vector for in silico cloning. The results of immunoinformatics analysis of the vaccine candidate demonstrated that this construct may have a high potency against pneumococcus, but further wet laboratory studies are needed in vitro and in vivo to confirm these results.

Table 4 The predicted IFN- γ epitopes from the proposed construct. The hybrid approach (MERC and SVM) were used to predict IFN-gamma epitopes by the IFNepitope server. The results revealed that there were 29 and 45 positive epitopes with scores greater or equal to 1 for the Ply4 adjuvant candidate and the main sequence of vaccine, respectively. The sequences were sorted from higher to lower scores

No.	Sequence	Method	Score	No.	Sequence	Method	Score
<i>The candidate adjuvant</i>							
1	AYVAQYYITWDELSY	MERC	5	16	GTGLAFEWWRVYEK	MERC	2
2	YVAQYYITWDELSYN	MERC	5	17	TGLAFEWWRVYEK	MERC	2
3	VAQYYITWDELSYNH	MERC	5	18	NGQDLTAHFTTSIPL	MERC	1
4	AQYYITWDELSYNHQ	MERC	5	19	GQDLTAHFTTSIPLK	MERC	1
5	QYYITWDELSYNHQG	MERC	5	20	QDLTAHFTTSIPLKG	MERC	1
6	YYITWDELSYNHQGK	MERC	5	21	DLTAHFTTSIPLKGN	MERC	1
7	YITWDELSYNHQGKE	MERC	5	22	LTAHFTTSIPLKGNV	MERC	1
8	IREGTGLAFEWWRV	MERC	3	23	TAHFTTSIPLKGNVR	MERC	1
9	REGTGLAFEWWRVY	MERC	3	24	AHFTTSIPLKGNVRN	MERC	1
10	GAYVAQYYITWDELS	MERC	2	25	RNLSVKIREGTGLAF	MERC	1
11	LSVKIREGTGLAFEW	MERC	2	26	NLSVKIREGTGLAFE	MERC	1
12	SVKIREGTGLAFEW	MERC	2	27	GLAFEWWRVYEKTD	MERC	1
13	VKIREGTGLAFEWWR	MERC	2	28	LAFEWWRVYEKTDL	MERC	1
14	KIREGTGLAFEWWR	MERC	2	29	AFEWWWRVYEKTDLP	MERC	1
15	EGTGLAFEWWRVYE	MERC	2				
<i>The main vaccine</i>							
1	TPPSTDHQDSGNTA	MERC	3	24	ESEDYAKEGFRAPLQ	MERC	2
2	PPSTDHQDSGNTA	MERC	3	25	SEDYAKEGFRAPLQS	MERC	2
3	PSTDHQDSGNTA	MERC	3	26	EDYAKEGFRAPLQSK	MERC	2
4	STDHQDSGNTA	MERC	3	27	EQPAGPGPG	SVM	1.08
5	TDHQDSGNTA	MERC	3	28	PGPGVQQAYLAYQRA	SVM	1.03
6	DHQDSGNTA	MERC	3	29	GLTPSTDHQDSGNT	MERC	1
7	LTPSTDHQDSGNT	MERC	2	30	QDSGNTA	MERC	1
8	HQDSGNTA	MERC	2	31	LDRMPYNLQYTV	MERC	1
9	KKSEAAKKAYEGPGP	MERC	2	32	DRMPYNLQYTV	MERC	1
10	KSEAAKKAYEGPGP	MERC	2	33	RMPYNLQYTV	MERC	1
11	SEAAKKAYEGPGPV	MERC	2	34	MPYNLQYTV	MERC	1
12	EAAKKAYEGPGVQ	MERC	2	35	NPSAGPGPGSPVASQ	MERC	1
13	AAKKAYEGPGVQQ	MERC	2	36	EDYVKEGFRAPLQSE	MERC	1
14	AKKAYEGPGVQQA	MERC	2	37	DYVKEGFRAPLQSEL	MERC	1
15	PGLKEIDSESEDYA	MERC	2	38	YVKEGFRAPLQSELD	MERC	1
16	GLKEIDSESEDYAK	MERC	2	39	VKEGFRAPLQSELDG	MERC	1
17	LKEIDSESEDYAKE	MERC	2	40	KEGFRAPLQSELDGP	MERC	1
18	KEIDSESEDYAKEG	MERC	2	41	EGFRAPLQSELDGPG	MERC	1
19	EIDSESEDYAKEGF	MERC	2	42	DPEDDTAALQNKGP	MERC	1
20	IDSESEDYAKEGFR	MERC	2	43	PEDDTAALQNKGP	MERC	1
21	DESESEDYAKEGFRA	MERC	2	44	EDDTAALQNKGP	MERC	1
22	ESESEDYAKEGFRAP	MERC	2	45	DDTAALQNKGP	MERC	1
23	SESEDYAKEGFRAPL	MERC	2				

Conclusion

The purpose of this research was to design a protective, antibody-inducing, multi-epitope vaccine against *S. pneumoniae*. To achieve this aim, the immunodominant

Table 5 Conformational/discontinuous B-cell epitopes in the final construct predicted by Ellipro server

No.	Residues	No. of residues	Score
1	G453, P454, G455, P456, G457, K458, V459, L460, A461, G467, K468, T469, Q470, D471, E472, L473, D474, K475, E476, A477, A478, E479, A480, G481, P482, G483, P484, G485, S486, K487, L488, E489, D490, N491, L492, K493, D494, T497, V500, E501, Y503, I504, K505, E506, G507, G508, P509, G510, P511, G512, P513, E514, G515, K516, T517, Q518, D519, E520, L521, D522, K523, E524, A525, A526, E527, D528, A529, N530, I531, E532, A533, L534, Q535, N536, K537, V538, A539, G540, P541, G542, P543, G544, R545, L546, Q547, S548, D549, L550, K551, D552, A553, E554, E555, N556, N557, V558, E559, D560, Y561, V562, K563, E564, G565, L566, G567, P568, G569, P570, G571, A572, P573, K574, P575, E576	114	0.776
2	M1, N2, G3, D4, L6, D8, H9, S10, G11, A12, Y13, V14, A15, Q16, Y18, I19, T20, W21, D22, E23, L24, S25, Y26, N27, H28, Q29, G30, K31, E32, V33, L34, T35, P36, K37, A38, W39, D40, R41, N42, G43, Q44, D45, L46, T47, A48, H49, F50, T51, T52, S53, I54, P55, L56, K57, G58, N59, V60, R61, N62, P88, L89, V90, R91, K92, T100, V110, E111, N112, D113, E114, A115, A116, A117, K118, A119, A120, Q121, A122, A124, K125, G128, L129, T130, P131, P132, S133, T134, D135, H136, Q137, D138, S139, G140, N141, T142, E143, A144, K145, G146, A147, E148, A149, Y151, N152, K155, V173, E174, V175, K176, N177, G178, S179, G180, P181, G182, P183, G184, K185, P186, Q187, T188, E189, K190, P191, E192, E193, D194, K195, E196, H197, D198, E199, V200, S201, E202, P203, T204	137	0.734
3	H586, H589, H590	3	0.602
4	P583, G584, H585	3	0.601
5	A252, Y253, E254, G255, P256, G257, P258, V260, Q261, Y264, G312, P313, G314, P315, G316, K317, G340, G342, P343, G344, K346	21	0.589
6	H205, P206, E207, S208, D209	5	0.587
7	G426, P427, G428, P429, G430	5	0.559
8	Y390, T392, E393, G394, L395, E396, K397, G398, P399, G400, P401	11	0.534
9	E466, Q577, P578, A579, G580	5	0.532
10	T71, G72, L73, A74, F75, W77	6	0.522
11	I348, D349, E350	3	0.518

B- and helper T-cell epitopes from highly conserved and protective antigens of pneumococcus, i.e., PspA (clades 1–5) and PhtD, were predicted computationally and the final considered regions were linked together using suitable linkers. The domain 4 of Ply as a potential adjuvant candidate was incorporated into the multi-epitope construct to improve the immunogenicity of epitope vaccine and to induce and increase the humoral immune responses against pneumococcus. After evaluation of the physicochemical properties, antigenicity and allergenicity of the final construct, its 3D structure was predicted by online tool. Structural quality, binding affinity to TLR4, and stability of the vaccine-receptor complex were analyzed. Finally, codon optimization was performed for gene expression in *E. coli* and in silico cloning was performed in the pET28a(+) vector. In spite of the significant results of this study obtained using in silico analysis, further experimental evaluation of the suggested multi-epitope peptide vaccine candidate is needed to confirm its efficacy.

Materials and methods

Amino acid sequence retrieval

The NCBI protein database (www.ncbi.nlm.nih.gov/protein) was used for the retrieval of the protein sequences of PspA from strains DBL6A (Clade 1) [AC AAF27701],

WU2 (Clade 2) [AC AAF27710], BG8090 (Clade 3) [AC AAF27713], EF5668 (Clade 4) [AC AAC62252] and ATCC6303 (Clade 5) [AC AAF27715], as well as PhtD from strain R6 [AC AAK99711] and Ply from strain D39 [AC ABJ53672]. The proteins PspA and PhtD were used for prediction of B-cell and helper T-cell epitopes and pneumolysin was applied as a TLR4 agonist.

Prediction of transmembrane regions

The presence of the transmembrane regions of the proteins PspA and PhtD were evaluated using TMHMM 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0) [88]. TMHMM, which is based on a hidden Markov model, can discriminate between membrane and soluble proteins with a high degree of accuracy.

Modeling and model validation of candidate proteins

In order to predict the conformational epitopes, the protein tertiary structures were constructed by I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) [89] or SWISS-MODEL server (<http://swissmodel.expasy.org>) [90] with respect to the presence of homologous templates with sequence identity less or more than 30%, respectively. The refinement of low-quality structures was done using the servers ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) [91] and GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) [92]. The 3D structure validation was conducted using the programs PROCHECK [93] and ERRAT [94] at SAVES server version 6 (<https://saves.mbi.ucla.edu/>). PROCHECK checks the stereochemical quality of 3D models by plotting the Ramachandran plot [93], and ERRAT validates the structures through statistical analysis of non-bonded interactions among different types of atoms [94]. BIOVIA Discovery Studio Visualizer was utilized for viewing the generated molecular structures.

Identification of linear/conformational B cell epitopes

Linear B-cell epitopes were identified by the servers LBtope (<http://crdd.osdd.net/raghava/lbtope/>) [95], ABCpred (<http://crdd.osdd.net/raghava/abcpred/>) [96], Emini surface accessibility prediction tool of IEDB (<http://tools.iedb.org/bcell/result/>) [97] and Ellipro at IEDB sever (<http://tools.iedb.org/ellipro/>) [98]. LBtope server, which is based on SVM, has a predictive accuracy of ~ 81%. This server assigns scores between 0 and 100 percent to each of the predicted epitopes and higher score signifies a higher probability of being an epitope. ABCpred, based on neural networks, predicts linear epitopes with an overall accuracy of 65.93%. A score closer to one corresponds a higher likelihood of the sequence being an epitope. Emini method was utilized to predict surface-accessible epitopes holding the default threshold value 1.0.

Furthermore, conformational epitopes from the modeled structures were predicted by the softwares ElliPro and DiscoTope 2.0 (<http://tools.iedb.org/discotope/>) [99]. ElliPro server identifies linear and structural epitopes based on solvent-accessibility and flexibility. This server predicts the antibody epitopes based on a three-step process: (1) approximation of the surface of the protein as an ellipsoid, (2) computation

of the protrusion index (PI) for each protein residue and (3) clustering of neighboring protein residues on the basis of the PI values. The higher scores correspond to the larger solvent accessibility. DiscoTope 2.0 server predicts discontinuous epitopes in 3D protein structures based on half-sphere exposure and propensity scores, with the default threshold – 3.7 (corresponding to sensitivity 0.47 and specificity 0.75).

Helper T cell (MHC-II) epitope prediction

Identification of binding peptides to MHC-II molecules was performed by the consensus method of IEDB MHC-II prediction tool (<http://tools.immuneepitope.org/mhcii>) [100] and NetMHCIIpan 4.0 server (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) [101]. The MHC class II T cell epitopes were predicted for the eight common human alleles HLA-DRB1*01:01, -DRB1*03:01, -DRB1*04:01, -DRB1*07:01, -DRB1*08:01, -DRB1*11:01, -DRB1*13:01 and -DRB1*15:01 [102], and also the mouse alleles H-2-IAb, H-2-IAd and H-2-IEd. In IEDB, the prediction of MHC-II epitopes is performed by default in a set of 15-mer peptides overlapping by 10 amino acids and the peptides with lower percentile rank indicate higher affinity. NetMHCIIpan 4.0 informs if a sequence is a strong binding peptide (SB) or weak binding peptide (WB) based on %Rank < 1.0 or < 5, respectively, using artificial neural network algorithm.

Construction of final vaccine sequence

The high-scored B-cell epitopes identified by various servers were analyzed and scanned for an overlapping sequence with the highly immunogenic MHC class II epitopes. The selected peptides were linked to each other by GPGPG linkers which conserve the independent immunological activities of the subunits and facilitate epitope presentation [103]. In addition, the domain 4 of Ply which was selected as a potential adjuvant was linked at the N-terminal of above construct via EAAAK linker to increase the immunogenicity of the epitope vaccine. EAAAK linker is a helical linker that can separate this domain from the other regions of the vaccine, providing proper structure for binding to the immune receptor [85]. A His-tag was added to the C-terminal of the final construct for the easy identification and purification of the vaccine.

Evaluation of properties of the vaccine sequence

After designing and construction of the multi-epitope vaccine by positioning the selected amino acid sequences linked with suitable linkers, the most important features of the construct were evaluated using different bioinformatics tools. The physical and chemical properties like amino acid composition, molecular weight, theoretical isoelectric point, in vitro/in vivo half-life, instability index, aliphatic index and grand average of hydropathicity of the designed sequence were computed using the ExPASy ProtParam tool (<http://web.expasy.org/protparam/>) [82]. Solubility, antigenicity and allergenicity were evaluated for the final construct. SOLpro tool (<https://scratch.proteomics.ics.uci.edu/>) [104] implemented in the SCRATCH server, predicted the protein solubility on overexpression in *E. coli*. VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) [105] and ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) [106] were used to calculate the protein antigenicity. The VaxiJen is an alignment independent tool for antigenicity prediction with the default threshold of 0.4 and the accuracy of

70–89% according to target organism. The ANTIGENpro tool of SCRATCH server is an alignment-free and pathogen independent predictor for checking the protein antigenicity with the accuracy of 76% and threshold of 0.5. AlgPred (<https://www.imtech.res.in/raghava/algpred/>) [107] and AllerTOP (<http://www.ddg-pharmfac.net/AllerTOP/>) [108] were employed to predict allergenicity of the vaccine construct. AlgPred uses various approaches consisting of SVM, motif-based and BLAST-search algorithms, and also hybrid approaches in order to allergenicity predictions and mapping of IgE peptides. AllerTOP, which is an alignment-independent server, predicts the allergens based on the physicochemical features of proteins.

3D structure modeling, refinement and validation of vaccine

The prediction of tertiary structure of the designed protein vaccine was done by the Robetta server (<http://rosetta.bakerlab.org/>) [109]. This server can predict 3D models of proteins based either on comparative modeling or ab initio approaches. The GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) [92] was applied for the refinement of the best-modeled structure. The model validation was done through ERRAT [94], PROCHECK [93] and VERIFY-3D [110] at SAVES server (<https://saves.mbi.ucla.edu/>), and also ProSA-web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) [111]. The VERIFY-3D is used to evaluate the compatibility of the 3D protein models with their own 1D amino acid sequences as measured by three-dimensional profiles. ProSA (Protein Structure Analysis) is a powerful tool for validation of the overall model quality which indicates whether the model has features characteristic for native structures.

Interferon-gamma inducing peptide prediction

The peptides having the potential to activate IFN- γ inducing T-helper cells were predicted using online server IFNepitope (<http://crdd.osdd.net/raghava/ifnepitope/>) [112]. This server utilizes various approaches like motif-based method, machine learning algorithm, and hybrid approach. The prediction is based on a main dataset composed of 3705 IFN-gamma inducing and 6728 non-inducing MHC-II binders.

Conformational B-cell epitope prediction in the constructed vaccine

Conformational epitopes are comprised of residues that form the 3D structures recognized by the antibodies and play a key role in the humoral immunity. Therefore, the designed construct should have effective structural epitopes within its protein domains to provide more potent immunity. ElliPro (<http://tools.iedb.org/ellipro/>) [98] was used to identify the conformational/discontinuous B-cell epitopes in the refined final model of the multi-epitope vaccine by keeping the default parameters. Among the current tools for conformational epitope prediction, ElliPro is one of the best, with the remarkable AUC value of 0.732.

Molecular docking between vaccine candidate and TLR-4

Protein–protein docking was performed through ClusPro web server (<https://clustpro.org/login.php>) [113] to determine the binding affinity of the multi-epitope vaccine and the toll-like receptor 4. The refined model of the vaccine as the ligand and the

crystal structure of TLR4 (PDB ID: 3FXI) retrieved from RCSB (www.rcsb.org) [114] as the receptor were submitted to the server. The ClusPro applies three consecutive stages for docking process: (1) rigid-body docking using the fast Fourier transform correlation method, (2) clustering of the lowest-energy conformations, and (3) refinement of chosen conformations by energy minimization. The docked molecule was visualized by Discovery Studio Visualizer. The PDBsum tool available at <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html> [115] was used to provide a graphical view of residual interaction between the vaccine and the receptor.

Molecular dynamics simulation

The MD simulation was performed to confirm the structural stability of developed vaccine-TLR4 complex by the GROMACS 4.6.5 version [116] using the GROMOSE 54A7 force field [117], such that the applied conditions were similar to our previous studies [118, 119]. Briefly, the complex was positioned in a cubic box and solvated with TIP3P water model. The system neutralization was performed by the Na⁺ and Cl⁻ ions, followed by system energy minimization using the steepest descent algorithm. The system equilibration was done under 100 ps NVT at 300 K using Berendsen thermostat algorithm, followed by 100 ps NPT at 300 K and 1 bar using Parrinello Rahman barostat. The root mean square deviation (RMSD) versus time of the vaccine-TLR4 complex was plotted for 20 ns simulation to evaluate the system stability. The root mean square fluctuation (RMSF) per residue plots of the developed vaccine and the receptor were generated to verify the flexibility of the backbone atoms.

Codon optimization and in silico cloning of the designed construct

The codon adaptation approach is applied to improve the expression of recombinant proteins. The vaccine protein sequence was submitted to the Java Codon Adaptation Tool (JCat) (<http://www.jcat.de/>) [120] for reverse translation and codon optimization according to the codon usage of *E. coli* (strain K12) as the expression host. The additional criteria of JCat were selected such as the avoidance of unwanted sites for Rho-independent transcription termination, prokaryotic ribosome binding and restriction enzymes. Finally, the optimized vaccine sequence, with restriction sites *NdeI* and *XhoI* added to the N- and C-terminus, respectively, was cloned in pET-28a(+) plasmid vector using the SnapGene tool (<https://www.snapgene.com/try-snapgene/>) to confirm the expression of the vaccine.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05175-6>.

Additional file 1. Table S1. Protein sequences of PspA1-5 (clades 1 to 5), PhtD and Ply. A region of PspA clade 2, B and C region of PspA clades 1 to 5 are underlined and shown in blue, red and green color, respectively. The C-terminal of PhtD (amino acid 383 to 853) and domain 4 of Ply (amino acid 360 to 471) are shown in color and underlined. In Ply4, the amino acids D385, C428 and W433, which must be replaced with N, G and F, respectively, are presented in blue color. Signal peptide is represented in lowercase italics. **Table S2.** The predicted B cell epitopes of A region of PspA2. **Table S3.** The predicted B cell epitopes of B region of PspA1. **Table S4.** The predicted B cell epitopes of B region of PspA2. **Table S5.** The predicted B cell epitopes of B region of PspA3. **Table S6.** The predicted B cell epitopes of B region of PspA4. **Table S7.** The predicted B cell epitopes of B region of PspA5. **Table S8.** The experimentally verified B cell epitopes from C region of PspAs. **Table S9.** The predicted linear B cell epitopes of PhtD-C. **Table S10.**

The predicted discontinuous B cell epitopes of PhtD-C. **Table S11.** Prediction of MHC-II epitopes of PspA2-A and PspA1-5-B. The peptides with IEDB percentile rank <10.0 and NetMHCIIpan rank value <1.0 were considered for the next analysis. **Table S12.** Prediction of MHC-II epitopes of PhtD-C. The epitopes with IEDB percentile rank <10.0 and NetMHCIIpan rank value <1.0 were considered for the further analysis. **Figure S1.** Domain structure of PspA protein. Major domains of PspA are α -helical charged domain (amino acids 1-288) consisting of A, A' and B regions, proline-rich domain (amino acids 289-370, C region), and choline-binding domain (amino acids 371-571). Within the α -HD, region B is a clade-defining region of the PspA molecule, which is represented by the stippled box. **Figure S2.** The results of transmembrane helices prediction. A and B show the prediction results of transmembrane helices in PspA2 and PhtD-C, respectively. The pink lines indicate the domains facing outside. **Figure S3.** The 3D models of PspA2 and PhtD-C. The predicted and refined 3D structures were visualized by the Discovery Studio Visualizer. **Figure S4.** The validation of refined model of PspA2 (A) and PhtD-C (B) with PROCHECK and ERRAT. Ramachandran plot of the structure of PspA2 or PhtD-C represents 82.1% or 87.2% residues in favored regions, respectively. In the ERRAT plot, the overall quality factor of structure of PspA2 or PhtD-C is 88.69% or 85.23%, respectively. **Figure S5.** The 3D models of B regions of PspA1-5. The homology modeled structures of B regions of PspA clades 1 to 5 (B1 to 5, respectively) visualized by Discovery Studio Visualizer. **Figure S6.** Codon optimization results of the designed construct. Codon optimization was performed using JCAT and codons were adapted for efficient expression in *E.coli* K12. The CAI index and GC content of the optimized sequence were 0.99 and 53.27%, respectively.

Acknowledgements

We are thankful to Dr. Mohammad Reza Pourshafie, Professor of Microbiology, Pasteur Institute of Iran, for his invaluable help in completing this study. This study was supported by a grant number of 1790 from the Semnan University of Medical Sciences, Semnan, Iran.

Author contributions

MSH and ZB: Data collection and analysis, and writing the manuscript. HM and MMR: Consulting in immunoinformatics. AASH and SFM: Supervision and monitoring of implementation of the study. Authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

All the data supporting the findings are contained within the manuscript.

Declarations

Ethics approval and consent to participate

Not applicable for this work as no ethical clearance was needed.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 December 2022 Accepted: 6 February 2023

Published online: 24 February 2023

References

- Henriques-Normark B, Tuomanen EI. The pneumococcus: epidemiology, microbiology, and pathogenesis. Cold Spring Harb Perspect Med. 2013;3: a010215.
- Denoël P, Philipp MT, Doyle L, Martin D, Carletti G, Poolman JT. A protein-based pneumococcal vaccine protects rhesus macaques from pneumonia after experimental infection with *Streptococcus pneumoniae*. Vaccine. 2011;29:5495–501.
- Berical AC, Harris D, Dela Cruz CS, Possick JD. Pneumococcal vaccination strategies. An update and perspective. Ann Am Thorac Soc. 2016;13:933–44.
- Abdollahi S, Siadat SD, Shapouri R, Mirzaei B, Mousavi SF, Nikbin VS, Moosavi SH. Antibiotic susceptibility and prevalence of adhesion genes in *Streptococcus pneumoniae* isolates detected in carrier children in Tehran. Jundishapur J Microbiol. 2018;11:6.
- Sender V, Hentrich K, Henriques-Normark B. Virus-induced changes of the respiratory tract environment promote secondary infections with *Streptococcus pneumoniae*. Front Cell Infect Microbiol. 2021;11:199.
- Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic review and meta-analysis. J Infect. 2020;81:266–75.
- Daniels CC, Rogers PD, Shelton CM. A review of pneumococcal vaccines: current polysaccharide vaccine recommendations and future protein antigens. J Pediatr Pharmacol Ther. 2016;21:27–35.
- Pichichero ME, Khan MN, Xu Q. Next generation protein based *Streptococcus pneumoniae* vaccines. Hum Vaccin Immunother. 2016;12:194–205.

9. Mousavi SF, Nobari S, Ghezalgeh FR, Lyriai H, Jalali P, Shahcheraghi F, Oskoui M. Serotyping of *Streptococcus pneumoniae* isolated from Tehran by multiplex PCR: are serotypes of clinical and carrier isolates identical? Iran J Microbiol. 2013;5:220.
10. Converso TR, Goulart C, Rodriguez D, Darrieux M, Leite L. Systemic immunization with rPotD reduces *Streptococcus pneumoniae* nasopharyngeal colonization in mice. Vaccine. 2017;35:149–55.
11. Dorosti H, Eslami M, Negahdaripour M, Ghoshoon MB, Gholami A, Heidari R, Dehshahri A, Erfani N, Nezafat N, Ghasemi Y. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. J Biomol Struct Dyn. 2019;37:3524.
12. Dagan R. Serotype replacement in perspective. Vaccine. 2009;27:C22–4.
13. Dorosti H, Eslami M, Nezafat N, Fadaei F, Ghasemi Y. Designing self-assembled peptide nanovaccine against *Streptococcus pneumoniae*: an in silico strategy. Mol Cell Probes. 2019;48: 101446.
14. Entwisle C, Hill S, Pang Y, Joachim M, McIlgorm A, Colaco C, Goldblatt D, De Gorguetta PDA, Bailey C. Safety and immunogenicity of a novel multiple antigen pneumococcal vaccine in adults: a phase 1 randomised clinical trial. Vaccine. 2017;35:7181–6.
15. Scott NR, Mann B, Tuomanen EI, Orihuela CJ. Multi-valent protein hybrid pneumococcal vaccines: a strategy for the next generation of vaccines. Vaccines. 2021;9:209.
16. Lagousi T, Basdeki P, Routsias J, Spoulou V. Novel protein-based pneumococcal vaccines: assessing the use of distinct protein fragments instead of full-length proteins as vaccine antigens. Vaccines. 2019;7:9.
17. Converso T, Assoni L, André G, Darrieux M, Leite LCC. The long search for a serotype independent pneumococcal vaccine. Expert Rev Vaccines. 2020;19:57–70.
18. Masomian M, Ahmad Z, Ti Gew L, Poh CL. Development of next generation *Streptococcus pneumoniae* vaccines conferring broad protection. Vaccines. 2020;8:132.
19. Sempere J, Llamós M, del Río MI, López Ruiz B, Domenech M, González-Camacho F. Pneumococcal choline-binding proteins involved in virulence as vaccine candidates. Vaccines. 2021;9:181.
20. Mukerji R, Mirza S, Roche AM, Widener RW, Cronney CM, Rhee D-K, Weiser JN, Szalai AJ, Briles DE. Pneumococcal surface protein A inhibits complement deposition on the pneumococcal surface by competing with the binding of C-reactive protein to cell-surface phosphocholine. J Immunol. 2012;189:5327–35.
21. Shaper M, Hollingshead SK, Benjamin WH Jr, Briles DE. PspA protects *Streptococcus pneumoniae* from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin. Infect Immun. 2004;72:5031–40.
22. Hollingshead SK, Becker R, Briles DE. Diversity of PspA: mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. Infect Immun. 2000;68:5889–900.
23. Daniels CC, Coan P, King J, Hale J, Benton KA, Briles DE, Hollingshead SK. The proline-rich region of pneumococcal surface proteins A and C contains surface-accessible epitopes common to all pneumococci and elicits antibody-mediated protection against sepsis. Infect Immun. 2010;78:2163–72.
24. Vadesilho CF, Ferreira DM, Gordon SB, Briles DE, Moreno AT, Oliveira MLS, Ho PL, Miyaji EN. Mapping of epitopes recognized by antibodies induced by immunization of mice with PspA and PspC. Clin Vaccine Immunol. 2014;21:940–8.
25. McDaniel LS, Ralph BA, McDaniel DO, Briles DE. Localization of protection-eliciting epitopes on PspA of *Streptococcus pneumoniae* between amino acid residues 192 and 260. Microb Pathog. 1994;17:323–37.
26. Roche H, Håkansson A, Hollingshead SK, Briles DE. Regions of PspA/EF3296 best able to elicit protection against *Streptococcus pneumoniae* in a murine infection model. Infect Immun. 2003;71:1033–41.
27. Mukerji R, Hendrickson C, Genschmer KR, Park S-S, Bouchet V, Goldstein R, Lefkowitz EJ, Briles DE. The diversity of the proline-rich domain of pneumococcal surface protein A (PspA): potential relevance to a broad-spectrum vaccine. Vaccine. 2018;36:6834–43.
28. Converso TR, Goulart C, Rodriguez D, Darrieux M, Leite L. Rational selection of broadly cross-reactive family 2 PspA molecules for inclusion in chimeric pneumococcal vaccines. Microb Pathog. 2017;109:233–8.
29. Miyaji EN, Ferreira DM, Lopes AP, Brandileone MCC, Dias WO, Leite LC. Analysis of serum cross-reactivity and cross-protection elicited by immunization with DNA vaccines against *Streptococcus pneumoniae* expressing PspA fragments from different clades. Infect Immun. 2002;70:5086–90.
30. Goulart C, Darrieux M, Rodriguez D, Pimenta FC, Brandileone MCC, de Andrade ALS, Leite LC. Selection of family 1 PspA molecules capable of inducing broad-ranging cross-reactivity by complement deposition and opsonophagocytosis by murine peritoneal cells. Vaccine. 2011;29:1634–42.
31. Darrieux M, Moreno AT, Ferreira DM, Pimenta FC, de Andrade ALS, Lopes AP, Leite LC, Miyaji EN. Recognition of pneumococcal isolates by antisera raised against PspA fragments from different clades. J Med Microbiol. 2008;57:273–8.
32. Akbari E, Negahdari B, Faraji F, Behdani M, Kazemi-Lomedasht F, Habibi-Anbouhi M. Protective responses of an engineered PspA recombinant antigen against *Streptococcus pneumoniae*. Biotechnol Rep. 2019;24: e00385.
33. Kristian SA, Ota T, Bubeck SS, Cho R, Groff BC, Kubota T, Destito G, Laudenslager J, Koriazova L, Tahara T. Generation and improvement of effector function of a novel broadly reactive and protective monoclonal antibody against pneumococcal surface protein A of *Streptococcus pneumoniae*. PLoS ONE. 2016;11: e0154616.
34. Plumtre CD, Ogunniyi AD, Paton JC. Polyhistidine triad proteins of pathogenic streptococci. Trends Microbiol. 2012;20:485–93.
35. Malekan M, Siadat SD, Aghasadeghi M, Shahrokhi N, Eybpoosh S, Afshari E. Assessment of PhtD C-terminal immunogenicity by opsonophagocytosis assay (OPA) with OMVs as adjuvants. Vaccine Res. 2019;6:37–41.
36. Bahadori Z, Shafaghi M, Madanchi H, Ranjbar MM, Shabani AA, Mousavi SF. In silico designing of a novel epitope-based candidate vaccine against *Streptococcus pneumoniae* with introduction of a new domain of PepO as adjuvant. J Transl Med. 2022;20:389.
37. Plumtre CD, Ogunniyi AD, Paton JC. Vaccination against *Streptococcus pneumoniae* using truncated derivatives of polyhistidine triad protein D. PLoS ONE. 2013;8: e78916.

38. Zubeldia J, Ferrer M, Dávila I, Justicia J. Adjuvants in allergen-specific immunotherapy: modulating and enhancing the immune response. *J Investig Allergol Clin Immunol*. 2018;29:103–11.
39. Bendelac A, Medzhitov R. Adjuvants of immunity: harnessing innate immunity to promote adaptive immunity. *J Exp Med*. 2002;195:F19–23.
40. Kumar S, Sunagar R, Gosselin E. Bacterial protein toll-like-receptor agonists: a novel perspective on vaccine adjuvants. *Front Immunol*. 2019;10:1144.
41. Douce G, Ross K, Cowan G, Ma J, Mitchell TJ. Novel mucosal vaccines generated by genetic conjugation of heterologous proteins to pneumolysin (PLY) from *Streptococcus pneumoniae*. *Vaccine*. 2010;28:3231–7.
42. Zhang H, Kang L, Yao H, He Y, Wang X, Xu W, Song Z, Yin Y, Zhang X. *Streptococcus pneumoniae* endopeptidase O (PepO) elicits a strong innate immune response in mice via TLR2 and TLR4 signaling pathways. *Front Cell Infect Microbiol*. 2016;6:23.
43. Chiu FF, Leng CH, Ding YJ, Chang JC, Chang LS, Lien SP, Chen HW, Siu LK, Liu SJ. Domain 4 of pneumolysin from *Streptococcus pneumoniae* is a multifunctional domain contributing TLR4 activating and hemolytic activity. *Biochem Biophys Res Commun*. 2019;517:596–602.
44. Shafaghi M, Shabani AA, Minucheir Z. Rational design of hyper-glycosylated human luteinizing hormone analogs (a bioinformatics approach). *Comput Biol Chem*. 2019;79:16–23.
45. Chen L, Wu D, Ji L, Wu X, Xu D, Cao Z, Han J. Bioinformatics analysis of the epitope regions for norovirus capsid protein. *BMC Bioinform*. 2013;14:1–6.
46. Fereshteh S, Goodarzi NN, Sepehr A, Shafiei M, Ajdary S, Badmasti F. In silico analyses of extracellular proteins of *Acinetobacter baumannii* as immunogenic candidates. *Iran J Pharm Res*. 2022. <https://doi.org/10.5812/ijpr-126559>.
47. Fathollahi M, Fathollahi A, Motamedi H, Moradi J, Alvandi A, Abiri R. In silico vaccine design and epitope mapping of New Delhi metallo-beta-lactamase (NDM): an immunoinformatics approach. *BMC Bioinform*. 2021;22:1–24.
48. Afshari E, Cohan RA, Sotoodehnejadnematlahi F, Mousavi SF. In-silico design and evaluation of an epitope-based serotype-independent promising vaccine candidate for highly cross-reactive regions of pneumococcal surface protein A. *J Transl Med*. 2023;21:13.
49. Tamborrini M, Geib N, Marrero-Nodarse A, Jud M, Hauser J, Aho C, Lamelas A, Zuniga A, Pluschke G, Gasparian A. A synthetic virus-like particle streptococcal vaccine candidate using B-cell epitopes from the proline-rich region of pneumococcal surface protein A. *Vaccines*. 2015;3:850–74.
50. Gharailoo Z, Mousavi SF, Halvani N, Feizabadi MM. Antimicrobial resistant pattern and capsular typing of *Streptococcus pneumoniae* isolated from children in Sistan–Baluchestan. *Mædica*. 2016;11:203.
51. Afshari E, Ahangari Cohan R, Sotoodehnejadnematlahi F. In-silico analysis of pneumococcal heat-shock protein (DnaJ) to predict novel multi-epitope vaccine candidates. *Vaccine Res*. 2021;8:65–87.
52. Norolahi F, Siadat SD, Malekan M, Mousavi SH, Janani A, Mousavi SF. Relationship between prevalence of pneumococcal serotypes and their neuraminidases in carriers, predictive facts? *Arch Pediatr Infect Dis*. 2020. <https://doi.org/10.5812/pedinfect.14100>.
53. Lu J, Sun T, Wang D, Dong Y, Xu M, Hou H, Kong FT, Liang C, Gu T, Chen P. Protective immune responses elicited by fusion protein containing PsaA and PspA fragments. *Immunol Invest*. 2015;44:482–96.
54. Goulart C, Silva TRD, Rodriguez D, Politano WR, Leite LC, Darrieux M. Characterization of protective immune responses induced by pneumococcal surface protein A in fusion with pneumolysin derivatives. *PLoS ONE*. 2013;8:e59605.
55. Nguyen CT, Kim SY, Kim MS, Lee SE, Rhee JH. Intranasal immunization with recombinant PspA fused with a flagellin enhances cross-protective immunity against *Streptococcus pneumoniae* infection in mice. *Vaccine*. 2011;29:5731–9.
56. Lu Y-J, Forte S, Thompson CM, Anderson PW, Malley R. Protection against Pneumococcal colonization and fatal pneumonia by a trivalent conjugate of a fusion protein with the cell wall polysaccharide. *Infect Immun*. 2009;77:2076–83.
57. Oli AN, Obialor WO, Ifeanyichukwu MO, Odimegwu DC, Okoyeh JN, Emechebe GO, Adejumo SA, Ibeanu GC. Immunoinformatics and vaccine development: an overview. *ImmunoTargets Ther*. 2020;9:13.
58. Safavi A, Kefayat A, Abiri A, Mahdevar E, Behnia AH, Ghahremani F. In silico analysis of transmembrane protein 31 (TMEM31) antigen to design novel multi-epitope peptide and DNA cancer vaccines against melanoma. *Mol Immunol*. 2019;112:93–102.
59. Li Z, Zhang F, Zhang C, Wang C, Lu P, Zhao X, Hao L, Ding J. Immunoinformatics prediction of OMP2b and BCSP31 for designing multi-epitope vaccine against Brucella. *Mol Immunol*. 2019;114:651–60.
60. Fereshteh S, Abdoli S, Shahcheraghi F, Ajdary S, Nazari M, Badmasti F. New putative vaccine candidates against *Acinetobacter baumannii* using the reverse vaccinology method. *Microb Pathog*. 2020;143: 104114.
61. Asadollahi P, Pakzad I, Sadeghifard N, Ghafourian S, Kazemian H, Kaviar VH, Fattahi R, Kalani BS. Immunoinformatics insights into the internalin A and B proteins to design a multi-epitope subunit vaccine for *L. monocytogenes*. *Int J Pept Res Ther*. 2022;28:1–10.
62. Saha S, Vashishtha S, Kundu B, Ghosh M. In-silico design of an immunoinformatics based multi-epitope vaccine against *Leishmania donovani*. *BMC Bioinform*. 2022;23:1–28.
63. Maleki A, Russo G, Parasiliti Palumbo GA, Pappalardo F. In silico design of recombinant multi-epitope vaccine against influenza A virus. *BMC Bioinform*. 2021;22:1–19.
64. Sanami S, Rafeian-Kopaei M, Dehkordi KA, Pazoki-Toroudi H, Azadegan-Dehkordi F, Mobini G-R, Alizadeh M, Nezhad MS, Ghasemi-Dehnoo M, Bagheri N. In silico design of a multi-epitope vaccine against HPV16/18. *BMC Bioinform*. 2022;23:1–24.
65. Naz A, Shahid F, Butt TT, Awan FM, Ali A, Malik A. Designing multi-epitope vaccines to combat emerging coronavirus disease 2019 (COVID-19) by employing immuno-informatics approach. *Front Immunol*. 2020;11:1663.
66. Sanami S, Azadegan-Dehkordi F, Rafeian-Kopaei M, Salehi M, Ghasemi-Dehnoo M, Mahooti M, Alizadeh M, Bagheri N. Design of a multi-epitope vaccine against cervical cancer using immunoinformatics approaches. *Sci Rep*. 2021;11:1–15.

67. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS, Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol.* 2017;48:83–94.
68. Perrie Y, Mohammed AR, Kirby DJ, McNeil SE, Bramwell VW. Vaccine adjuvant systems: enhancing the efficacy of sub-unit protein antigens. *Int J Pharm.* 2008;364:272–80.
69. Berry AM, Alexander JE, Mitchell TJ, Andrew PW, Hansman D, Paton JC. Effect of defined point mutations in the pneumolysin gene on the virulence of *Streptococcus pneumoniae*. *Infect Immun.* 1995;63:1969–74.
70. Kono M, Hotomi M, Hollingshead SK, Briles DE, Yamanaka N. Maternal immunization with pneumococcal surface protein A protects against pneumococcal infections among derived offspring. *PLoS ONE.* 2011;6: e27102.
71. Nagano H, Kawabata M, Sugita G, Tsuruhara A, Ohori J, Jimura T, Miyashita K, Kurono Y, Tomonaga K, Briles DE. Transcutaneous immunization with pneumococcal surface protein A in mice. *Laryngoscope.* 2018;128:E91–6.
72. Briles DE, Hollingshead SK, King J, Swift A, Braun PA, Park MK, Ferguson LM, Nahm MH, Nabors GS. Immunization of humans with recombinant pneumococcal surface protein A (rPspA) elicits antibodies that passively protect mice from fatal infection with *Streptococcus pneumoniae* bearing heterologous PspA. *J Infect Dis.* 2000;182:1694–701.
73. Nabors GS, Braun PA, Herrmann DJ, Heise ML, Pyle DJ, Gravenstein S, Schilling M, Ferguson LM, Hollingshead SK, Briles DE. Immunization of healthy adults with a single recombinant pneumococcal surface protein A (PspA) variant stimulates broadly cross-reactive antibodies to heterologous PspA molecules. *Vaccine.* 2000;18:1743–54.
74. Kallio A, Sepponen K, Hermand P, Denoël P, Godfroid F, Melin M. Role of Pht proteins in attachment of *Streptococcus pneumoniae* to respiratory epithelial cells. *Infect Immun.* 2014;82:1683–91.
75. Plumpre CD, Ogunniyi AD, Paton JC. Surface association of Pht proteins of *Streptococcus pneumoniae*. *Infect Immun.* 2013;81:3644–51.
76. Singh R, Singh S, Sharma PK, Singh UP, Briles DE, Hollingshead SK, Lillard JW Jr. Helper T cell epitope-mapping reveals MHC-peptide binding affinities that correlate with T helper cell responses to pneumococcal surface protein A. *PLoS ONE.* 2010;5: e9432.
77. Kavooosi M, Creagh AL, Kilburn DG, Haynes CA. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. *Biotechnol Bioeng.* 2007;98:599–610.
78. Kar T, Narsaria U, Basak S, Deb D, Castiglione F, Mueller DM, Srivastava AP. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci Rep.* 2020;10:1–24.
79. Sanches RC, Tiwari S, Ferreira LC, Oliveira FM, Lopes MD, Passos MJ, Maia EH, Taranto AG, Kato R, Azevedo VA. Immunoinformatics design of multi-epitope peptide-based vaccine against *Schistosoma mansoni* using transmembrane proteins as a target. *Front Immunol.* 2021;12:490.
80. Sanami S, Zandi M, Pourhossein B, Mobini G-R, Safaei M, Abed A, Arveje PM, Chermahini FA, Alizadeh M. Design of a multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach. *Int J Biol Macromol.* 2020;164:871–83.
81. Zhao X, Zhang F, Li Z, Wang H, An M, Li Y, Pang N, Ding J. Bioinformatics analysis of EgA31 and EgG1Y162 proteins for designing a multi-epitope vaccine against *Echinococcus granulosus*. *Infect Genet Evol.* 2019;73:98–108.
82. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *Proteom Protoc Hand.* 2005. <https://doi.org/10.1385/1-59259-890-0:571>.
83. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem.* 1980;88:1895–8.
84. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157:105–32.
85. Bemani P, Amirghofran Z, Mohammadi M. Designing a multi-epitope vaccine against blood-stage of *Plasmodium falciparum* by in silico approaches. *J Mol Graph Model.* 2020;99: 107645.
86. Brooks LR, Mias GI. *Streptococcus pneumoniae*'s virulence and host immunity: aging, diagnostics, and prevention. *Front Immunol.* 2018;9:1366.
87. Khan S, Ali SS, Zaheer I, Saleem S, Zaman N, Iqbal A, Suleman M, Wadood A, Rehman AU. Proteome-wide mapping and reverse vaccinology-based B and T cell multi-epitope subunit vaccine designing for immune response reinforcement against *Porphyromonas gingivalis*. *J Biomol Struct Dyn.* 2020. <https://doi.org/10.1080/07391102.2020.1819423>.
88. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
89. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods.* 2015;12:7–8.
90. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46:W296–303.
91. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J.* 2011;101:2525–34.
92. Shin W-H, Lee GR, Heo L, Lee H, Seok C. Prediction of protein structure and interaction by GALAXY protein modeling programs. *Bio Design.* 2014;2:1–11.
93. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26:283–91.
94. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993;2:1511–9.
95. Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS ONE.* 2013;8: e62216.
96. Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins.* 2006;65:40–8.
97. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol.* 1985;55:836–9.
98. Ponomarenko J, Bui H-H, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* 2008;9:1–8.

99. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput Biol*. 2012;8: e1002829.
100. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol*. 2008;4: e1000048.
101. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*. 2018;154:394–406.
102. van de Garde MD, van Westen E, Poelen MC, Rots NY, van Els CA. Prediction and validation of immunogenic domains of pneumococcal proteins recognized by human CD4+ T cells. *Infect Immun*. 2019;87:e00098-e119.
103. Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. *Infect Genet Evol*. 2017;51:227–34.
104. Magnan CN, Randall A, Baldi P. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics*. 2009;25:2200–7.
105. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform*. 2007;8:1–7.
106. Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*. 2010;26:2936–43.
107. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34:W202–9.
108. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. *J Mol Model*. 2014;20:1–6.
109. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004;32:W526–31.
110. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992;356:83–5.
111. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35:W407–10.
112. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct*. 2013;8:1–15.
113. Kozakov D, Hall DR, Xia B, Porter KA, Padhorna D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein–protein docking. *Nat Protoc*. 2017;12:255–78.
114. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* 2009;458:1191–5.
115. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Sci*. 2018;27:129–34.
116. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008;4:435–47.
117. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, van Gunsteren WF. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J*. 2011;40:843–56.
118. Bahadori Z, Shabani AA, Minucheer Z. Rational design of hyper-glycosylated human follicle-stimulating hormone analogs (a bioinformatics approach). *J Biomol Struct Dyn*. 2021. <https://doi.org/10.1080/07391102.2021.1924268>.
119. Nabizadeh Z, Minucheer Z, Shabani AA. Rational design of hyper-glycosylated human chorionic gonadotropin analogs (a bioinformatics approach). *Lett Drug Des Discov*. 2020;17:1001–14.
120. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005;33:W526–31.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

