American Journal of Medical and Biological Research, 2016, Vol. 4, No. 5, 90-94
Available online at http://pubs.sciepub.com/ajmbr/4/5/2

DOI:10.12691/ajmbr-4-5-2

Computational Design of Serotype Independent Vaccine against *Streptococcus pneumoniae* Based on B-Cell Epitopes of Pneumococcal Plasmid Stabilization Protein

Shirin Tarahomjoo*

Division of Genomics and Genetic Engineering, Department of Biotechnology and Central Laboratory, Razi Vaccine and Serum Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj 31975/148, Iran
*Corresponding author: starahomjoo@gmail.com

Abstract Pneumococcal conjugate vaccines (PCVs) were constructed through chemical conjugation of pneumococcal capsules to immunogenic carrier proteins. The PCVs implementation in developing countries was prevented by their high manufacturing costs. This issue can be overcome by development of protein based vaccines against pneumococci. Antibody responses are necessary for protection against *S. pneumoniae*. The plasmid stabilization protein (PSP) was already identified as a pneumococcal surface protein able to elicit protection against *S. pneumoniae* serotype 19F and its protective B-cell epitope regions were determined. Whole antigens are not as potent as epitope based vaccines and every epitope in a multi epitope based vaccine can individually induce a protective immune response against the pathogen. Thus better immunoprotection can be achieved by multi epitope based vaccines. In the present study, therefore, we aim to design a multi epitope vaccine against pneumococci based on the identified B-cell epitope regions of PSP using immunoinformatic tools. These regions were joined together using the $(EAAAK)_4$ linker. The resulting antigen (HPBE) showed much higher immunoprotective ability compared to PSP regarding the VaxiJen scores. The codon optimization was done for HPBE using OPTIMIZER. Analysis of the mRNA secondary structure using Mfold tool revealed no stable hairpin at the 5' end and thus the antigen can be expressed appropriately. The 3D model of the antigen resulted from I-TASSER indicated the presence of alpha helix, beta sheet, turn, coil, and 3_{10} helix as the protein structural elements. Analyzing physicochemical properties of the antigen using ProtParam showed that it was stable and its half life in *Escherichia coli* was more than 10 h. Considering the GRAVY score, HPBE possessed a hydrophilic nature and it can be expressed in the soluble form in *E. coli* at 79.6% probability. Our results demonstrated that HPBE is a suitable vaccine candidate, which can elicit protection against common *S. pneumoniae* serotypes causing invasive pneumococcal disease in children less than 5 years of age.

Keywords: *computational design, pneumococcal conjugate vaccines, protective epitope, protein based vaccines, Streptococcus pneumoniae*

Cite This Article: Shirin Tarahomjoo, "Computational Design of Serotype Independent Vaccine against *Streptococcus pneumoniae* Based on B-Cell Epitopes of Pneumococcal Plasmid Stabilization Protein." *American Journal of Medical and Biological Research*, vol. 4, no. 5 (2016): 90-94. doi: 10.12691/ajmbr-4-5-2.

1. Introduction

Streptococcus pneumoniae is a major cause of diseases such as meningitis, pneumoniae and sepsis mainly in children less than 5 years of age [1,2]. World Health Organization (WHO) reported that 476000 annual deaths among children less than 5 years of age were caused by pneumococcal infections [3].

Pneumococcal vaccines have been used to protect against pneumococcal infections. The capsular polysaccharides of pneumococci are main antigenic components of these vaccines. However, the capsules are weakly immunogenic in children and are not able to induce immune memory. Pneumococcal conjugate vaccines (PCVs) are then constructed through chemical conjugation of pneumococcal capsules to immunogenic carrier proteins. The conjugation process increases the antibody response and induces the immune memory. WHO recommends the inclusion of PCVs in national immunization programs for children. However, the PCVs production requires a multistep process and the elicited protection is serotype dependent. These vaccines include conjugated capsules of multiple pneumococcal serotypes. Therefore, the vaccine manufacturing costs are high and limit PCVs implementation in national immunization programs of developing countries [1,2,3].

The development of protein based vaccines against pneumococci offers a more affordable protective strategy against pneumococcal infections. However, the pathway to licensure of these vaccines is yet to be defined. The regulatory pathways should enable the vaccine developers to demonstrate the protein based vaccine efficiency in direct comparison to licensed conjugate vaccines [4]. Cell

surface proteins are key factors in infection processes of pathogens and have extensively been evaluated as vaccine candidates [5,6]. Several pneumococcal surface proteins such as pneumococcal surface protein A, pneumococcal surface protein C, pneumococcal surface adhesin A, pilus proteins and histidine triad proteins were able to induce protection against *S. pneumoniae* [7]. We already identified plasmid stabilization protein (PSP), a pneumococcal surface protein as a suitable candidate for eliciting protection against *S. pneumoniae* serotype 19F. The candidate selection was based on the protein ability to elicit immunoprotection, the absence of autoimmunity induction and the amino acid sequence conservancy in serotype 19F pneumococcal strains [8]. Antibody responses are essential for protection against pneumococcal infections [7]. Antibodies bind specifically to a continuous amino acid sequence known as the linear B-cell epitope or to a folded structure composed of discontinuous amino acids known as the conformational B-cell epitopes. Therefore, the analysis of protective B-cell epitopes in vaccine candidates is necessary to develop effective vaccines against pneumococci [9]. The majority of B-cell epitopes are conformational. Nevertheless, the identification of linear B-cell epitopes has shown promising results for selection of the vaccine constituents [10,11]. Whole antigens are not as potent as epitope based vaccines and usually a few epitopes are sufficient to elicit protective immune responses. As a result, B-cell epitope based vaccines are more effective than whole antigen based vaccines in the prevention of infections. In addition, every epitope in a multi epitope based vaccine can individually induce a protective immune response against the pathogen. Hence, more efficient immunoprotection can be achieved against infections by multi epitope based vaccines [12,13].

Compared with conventional laboratory methods, computational approaches offer the ability to undertake rapid and comprehensive epitope assessments at much lower costs [14]. Simultaneous applications of B-cell epitope prediction tools enhance the epitope prediction accuracy. Bepipred and BCPreds are widely used as bioinformatics tools for identification of linear B-cell epitopes. The CBTope program is used for the prediction of conformational B-cell epitopes using amino acid sequences [15]. We already determined B-cell epitope regions of PSP by simultaneous applications of results of these three programs [16]. In the present study, we used the identified B-cell epitope regions of PSP to design a multi epitope vaccine against pneumococci. The immunological, physicochemical, and structural properties of the designed antigen were evaluated using immunoinformatic tools to elucidate its suitability as a protein based vaccine against pneumococci.

2. Methods

The immunoprotective abilities of antigens were evaluated using VaxiJen [17]. PSP sequences were retrieved from NCBI. Accession numbers of PSPs were CIW52344 (serotype 14), CEY90865 (serotype 6B), CIS79081 (serotype 1), CIV22281 (serotype 23F), CIU16151 (serotype 5), CGF29401 (serotype 6A), and WP_000834749 (serotype 19F). Amino acid sequences were aligned using Clustal Omega to analyze the B-cell epitope regions conservancy. The codon optimization and calculations of the codon adaptation index (CAI) were performed using the OPTIMIZER online tool. *Escherichia coli* K-12 is the basic strain for the construction of *E. coli* based expression hosts. Therefore, the codon usage table of *E. coli* K-12 obtained from Kazusa codon usage database was used for the codon optimization [18]. Prediction of the mRNA secondary structure was carried out using Mfold web server [19]. The protein physicochemical properties were evaluated using PROTPARAM based on the amino acid composition [20]. Moreover, the probability of the protein over-expression in *E. coli* in the soluble form was calculated using the PROSOII tool [21]. The YASARA program was used to determine the protein secondary structure constituents. I-TASSER web server was used for de novo prediction of the protein tertiary structure from the amino acid sequence [22]. The stereochemical quality of the tertiary model was examined using the Ramachandran plot on the Cambridge RAMPAGE server. YASARA was used to visualize the protein tertiary structure.

3. Results and Discussion

3.1. Conservation of PSP B-cell Epitope Regions in Common Pneumococcal Serotypes

The incidence of invasive pneumococcal disease (IPD) is frequently used as a measure of the incidence of severe pneumococcal disease. Pneumococcal serotypes 1, 5, 6A, 6B, 14, 19F and 23F are common causes of IPD globally in children less than 5 years of age [3]. We already identified eight B-cell epitope regions in PSP from serotype 19F *S. pneumoniae* [16]. The amino acid sequences of PSPs from other common pneumococcal serotypes including 1, 5, 6A, 6B, 14, 23F were retrieved from NCBI. Amino acid sequence alignments using Clustal Omega demonstrated that the first identified B-cell epitope region was not present in these serotypes (data not shown). In addition, the second B-cell epitope region was not present in serotype 14 *S. pneumonia*e. However, the last six B-cell epitope regions were highly conserved in the pneumococcal serotypes. Therefore, these regions can elicit protection against the common IPD causing pneumococcal serotypes.

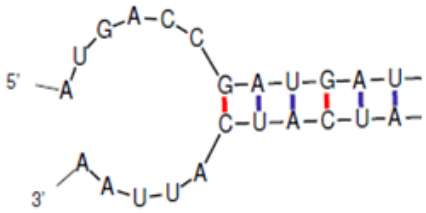
3.2. Design of Multi Epitope Vaccine

A multi epitope vaccine was designed by fusing the last six B-cell epitope regions of PSP through the $(EAAAK)_4$ linker. A histidine tag was introduced at the C-terminus of the construct (HPBE). The VaxiJen score of HPBE (1.2451) was much higher than that of PSP (0.6568). VaxiJen was developed to classify antigens solely based on their physicochemical properties without the need for sequence alignments. A higher VaxiJen score refers to a higher probability for the immunoprotective ability [16]. Therefore, our result indicated the higher efficiency of the multi epitope antigen compared with PSP in eliciting immunoprotection.

Adapting the gene codon usage to the host codon usage can enhance the expression level of the encoded protein. The codon optimization includes altering the target gene codons so that they reflect the codon usage of the host more closely without changing the amino acid sequence of the encoded protein [18]. Results of OPTIMIZER showed that CAI of native *hpbe* gene was 0.623. However, the codon optimized *hpbe* gene sequence showed a CAI of 1. CAI measures the similarity between the codon usage of a gene and the codon usage of the reference set of genes. Its value is in the range of 0-1 [18]. Therefore, the codon usage of the optimized *hpbe* gene sequence was the same as the codon usage of *E. coli* reference group of genes and the optimized gene can be expressed at suitable levels in *E. coli*. The GC content of the native *hpbe* was increased from 32.3% to 51.3% upon codon optimization. The GC content of the gene affects its codon usage [23]. The gene expression level correlates inversely with the stability of mRNA hairpin structures close to the 5' end of mRNA [24]. The 5' end of the optimized *hpbe* mRNA did not constrain any stable hairpins as predicted by Mfold (Figure 1). Therefore the designed vaccine is expected to be expressed at appropriate levels.

3.3. Structures of Muti Epitope Vaccine

The tertiary structure of the designed vaccine was predicted using I-TASSER. I-TASSER predicts 5 models for the designed vaccine (data not shown). The most confident model given for HPBE (Figure 2) had a confidence score (C-score) of -3.4. C-score is in the range of -5 to 2 and a higher C-score indicates a model with a higher confidence [22]. Alpha helix, beta sheet, turn, coil and 3_{10} helix were the structural elements of HPBE. The Ramachandran plot assessment indicated that 80.3% of the amino acid residues (favorable region: 47.2% and allowed region: 33.1%) occupied the desired space. This result indicated the structural stability of the designed vaccine. The percentages of the structural elements in the antigen secondary structure were predicted using YASARA (Table 1).



Structural element	δG	Information
Extemalloop	-2.00	11 ss bases & 1 closing helices.
Stack	-2.40	External closing pair is G_7-C_{901}
Stack	-1.10	External closing pair is A_8-U_{900}
Stack	-2.10	External closing pair is U_9-A_{899}
Stack	-2.40	External closing pair is G_{10}-C_{898}
Stack	-1.10	External closing pair is A_{11}-U_{897}
Stack	-2.10	External closing pair is U_{12}-A_{896}

Figure 1. The mRNA secondary structure and free energy details at the 5' end of multi epitope vaccine predicted by Mfold

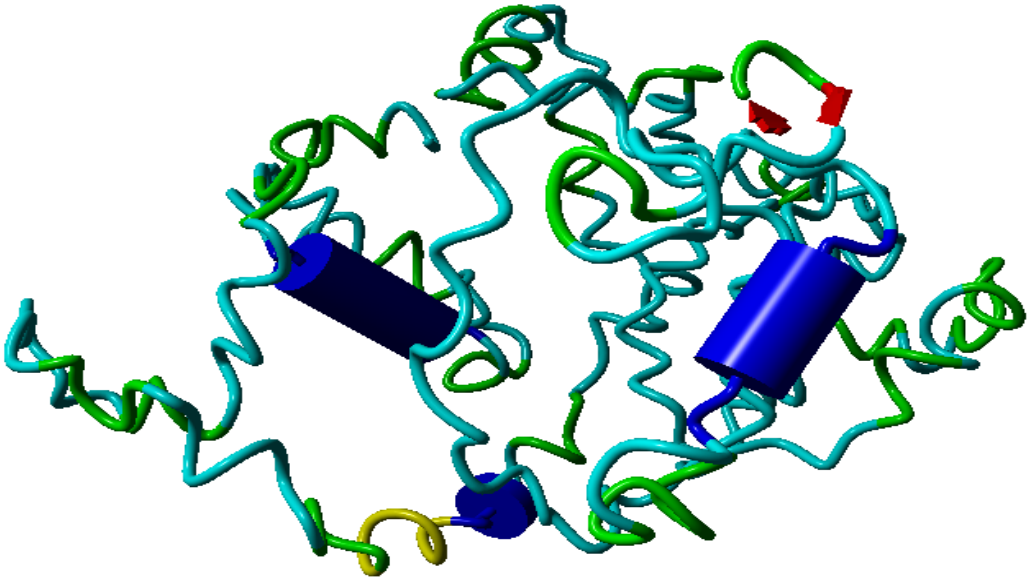


Figure 2. Tertiary structure of HPBE obtained using I-TASSER. Alpha helix was shown as cylinder. Beta sheets are in red. Coils and turns were shown in light blue and green respectively. The 3_{10} helix is shown in dark yellow

Table 1. Percentages of HPBE secondary structure elements estimated using YASARA

Secondary structure element	HPBE
Alpha helix	10.3%
Beta sheet	0.7%
Turn	28.9%
Coil	58.5%
3_{10} helix	1.7%

3.4. Physicochemical Properties of Multi Epitope Vaccine

PROTPARAM analysis revealed that the average molecular weight (MW) of HPBE was 32.13 kDa and the isoelectric points (pI) of the chimerical antigen was 5.49 (Table 2). The estimated half life of the antigen in *E. coli* was more than 10 h. The instability index (II) provides an estimate of the protein stability in a test tube. Proteins with II values lower than 40 are considered as stable [20]. The II value of HPBE was less than 40 that indicated the antigen stability. The grand average of hydropathicity (GRAVY) score is used to evaluate the hydropathical character of proteins. A negative GRAVY score indicates a hydrophilic protein [20]. The negative GRAVY score of HPBE therefore indicated that it can interact with water molecules.

E. coli is a popular host for expression of recombinant proteins owing to its superior properties including the ease of genetic manipulation and the high expression level. However, *E. coli* is often not able to fold foreign proteins properly and inclusion bodies are then formed. The protein recovery from inclusion bodies requires a multistep complex process, which can lead to a significant loss in the final protein yield [25,26]. PROSOII was used to assess the chance of a protein to be soluble upon heterologous expression in *E. coli* based on its amino acid sequence composition. The solubility probability of HPBE was 0.796. Therefore, the majority of the protein is expressed in the soluble form in *E. coli* and can then be recovered from the soluble cellular fraction instead of inclusion bodies.

Table 2. Physicochemical properties of HPBE calculated by Protparam

Protein	Mwt	pI	Half life in *E. coli*	Half life in yeast	Half life in mammalian reticulocyte	II	AI	GRAVY score
HPBE	32.13 kDa	5.49	> 10 h	> 20 h	30 h	28.95	52.26	- 1.15

4. Conclusions

The multi epitope vaccine designed in this study can elicit protection against *S. pneumoniae* independent of the serotype. It can be expressed at appropriate levels in *E. coli* in the soluble form and is stable. Therefore, HPBE is a suitable proteinaceous vaccine candidate against *S. pneumoniae*.

References

[1] Mook-Kanamori, B.B., et al., *Pathogenesis and pathophysiology of pneumococcal meningitis.* Clinical microbiology reviews, 2011. 24(3): p. 557-591.

[2] Johnson, H.L., et al., *Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project.* PLoS Med, 2010. 7(10): p. e1000348.

[3] Publication, W., *Pneumococcal vaccines WHO position paper–2012–recommendations.* Vaccine, 2012. 30(32): p. 4717-4718.

[4] Ginsburg, A.S., et al., *Issues and challenges in the development of pneumococcal protein vaccines.* Expert review of vaccines, 2012. 11(3): p. 279-285.

[5] Foster, T.J., et al., *Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus.* Nature Reviews Microbiology, 2014. 12(1): p. 49-62.

[6] Bergmann, S. and S. Hammerschmidt, *Versatility of pneumococcal surface proteins.* Microbiology, 2006. 152(2): p. 295-303.

[7] Tarahomjoo, S., *Recent approaches in vaccine development against Streptococcus pneumoniae.* Journal of molecular microbiology and biotechnology, 2014. 24(4): p. 215-227.

[8] Tarahomjoo, S., *Bioinformatic analysis of surface proteins of Streptococcus pneumoniae serotype 19F for identification of vaccine candidates.* American Journal of Microbiological Research, 2014. 2(6): p. 174-177.

[9] Novotný, J.í., et al., *Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains).* Proceedings of the National Academy of Sciences, 1986. 83(2): p. 226-230.

[10] Quijada, L., et al., *Mapping of the linear antigenic determinants of the Leishmania infantum hsp70 recognized by leishmaniasis sera.* Immunology letters, 1996. 52(2): p. 73-79.

[11] Faria, A.R., et al., *High-throughput analysis of synthetic peptides for the immunodiagnosis of canine visceral leishmaniasis.* PLoS Negl Trop Dis, 2011. 5(9): p. e1310.

[12] Zhao, Z., et al., *Multiple B-cell epitope vaccine induces a Staphylococcus enterotoxin B-specific IgG1 protective response against MRSA infection.* Scientific reports, 2015. 5.

[13] Lu, Y., et al., *A candidate vaccine against influenza virus intensively improved the immunogenicity of a neutralizing epitope.* International archives of allergy and immunology, 2002. 127(3): p. 245-250.

[14] Kelly, D.F. and R. Rappuoli, *Reverse vaccinology and vaccines for serogroup B Neisseria meningitidis*, in *Hot Topics in Infection and Immunity in Children II*. 2005, Springer. p. 217-223.

[15] Assis, L., et al., *B‐cell epitopes of antigenic proteins in Leishmania infantum: an in silico analysis.* Parasite immunology, 2014. 36(7): p. 313-323.

[16] Tarahomjoo, S., *Identification of B-cell Epitope Regions in Cell Surface Proteins of Streptococcus pneumoniae Serotype 19F Using Bioinformatic Tools.* American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 2015. 14(3): p. 107-117.

[17] Doytchinova, I.A. and D.R. Flower, *Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines.* Open Vaccine J, 2008. 1(1): p. 4.

[18] Puigbo, P., et al., *OPTIMIZER: a web server for optimizing the codon usage of DNA sequences.* Nucleic acids research, 2007. 35(suppl 2): p. W126-W131.

[19] Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic acids research, 2003. 31(13): p. 3406-3415.

[20] Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*. 2005: Springer.

[21] Smialowski, P., et al., *PROSO II–a new method for protein solubility prediction.* Febs Journal, 2012. 279(12): p. 2192-2200.

[22] Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nature protocols, 2010. 5(4): p. 725-738.

[23] Li, J., et al., *GC-content of synonymous codons profoundly influences amino acid usage.* G3: Genes| Genomes| Genetics, 2015. 5(10): p. 2027-2036.

[24] Seo, S.W., J. Yang, and G.Y. Jung, *Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli.* Biotechnology and bioengineering, 2009. 104(3): p. 611-616.

[25] Singh, S.M. and A.K. Panda, *Solubilization and refolding of bacterial inclusion body proteins.* Journal of bioscience and bioengineering, 2005. 99(4): p. 303-310.

[26] Yin, J., et al., *Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes.* Journal of Biotechnology, 2007. 127(3): p. 335-347.