

Bioinformatic Analysis of Surface Proteins of *Streptococcus pneumoniae* Serotype 19F for Identification of Vaccine Candidates

Shirin Tarahomjoo*

Department of Biotechnology, Razi Vaccine and Serum Research Institute, Karaj, Iran

*Corresponding author: starahomjoo@hotmail.com

Received October 15, 2014; Revised October 23, 2014; Accepted October 26, 2014

Abstract *Streptococcus pneumoniae* serotype 19F is one of major pneumococcal serotypes responsible for pneumococcal invasive disease in children less than 5 years worldwide. Pneumococcal conjugate vaccines (PCVs) were developed through chemical coupling of capsular polysaccharides of pneumococci to immunogenic carrier proteins and World Health Organization recommends the inclusion of these vaccines in national immunization programs for children. However, costly manufacture of PCVs can prevent their implementation in developing countries. This issue can be addressed by construction of protein based vaccines against pneumococci. Cell surface proteins are key factors in infectious processes of pathogens and are attractive as vaccine candidates. LPxTG motif containing proteins, lipoproteins, and choline binding proteins are among main groups of pneumococcal surface proteins. In this study, therefore, we aim to identify suitable candidates among these proteins for development of proteinaceous vaccines against *S. pneumoniae* serotype 19F infection using bioinformatic tools. These proteins were then identified in proteome data of *S. pneumoniae* Taiwan 19F-14 through BLAST with LPxTG, lipobox consensus motifs, and the choline binding protein consensus sequence. PRED-LIPO online tool was used to confirm the presence of lipoprotein specific signal peptide. The cellular location of the proteins was analyzed with PSORTb v.3.0. Vaxijen v.2.0 was used to evaluate the protein antigenicity. BLAST against human proteome was done to remove the possibility for autoimmunity induction by the proteins. Moreover, the presence of homolog proteins in other pneumococcal serotype 19F strains including *S. pneumoniae* A026, *S. pneumoniae* G54, and *S. pneumoniae* ST556 were investigated. Our analysis revealed that cell wall surface anchor family protein (YP_002741626.1), D-alanyl-D-alanyl-carboxy peptidase, surface protein PspC, and choline binding protein D are promising candidates for development of protein based vaccines against *S. pneumoniae* serotype 19F infection.

Keywords: choline binding proteins, lipoproteins, LPxTG motifs, protein, streptococcus pneumoniae serotype 19F, vaccine

Cite This Article: Shirin Tarahomjoo, "Bioinformatic Analysis of Surface Proteins of *Streptococcus pneumoniae* Serotype 19F for Identification of Vaccine Candidates." *American Journal of Microbiological Research*, vol. 2, no. 6 (2014): 174-177. doi: 10.12691/ajmr-2-6-2.

1. Introduction

Streptococcus pneumoniae is a major pathogen that causes diseases such as pneumoniae, meningitis and sepsis [1]. World Health Organization (WHO) estimated that 476000 global annual deaths among children less than 5 years of age were caused by pneumococcal infections [2]. *S. pneumoniae* serotype 19F is among main pneumococcal serotypes that cause invasive pneumococcal disease in children less than 5 years of age in the world. Pneumococcal vaccines have been used for protection against pneumococcal infections. Capsular polysaccharides of pneumococci are main antigenic components of these vaccines. However, the capsule elicits weak immunogenicity in children and is not able to induce immune memory. In order to overcome these shortcomings, the capsule is chemically conjugated to a proteinaceous carrier. WHO recommends the inclusion

of pneumococcal conjugate vaccines (PCVs) in national immunization programs for children [3]. However, high manufacturing costs of PCVs limit their implementation in developing countries [1,2]. Development of protein based vaccines against pneumococci offers a more affordable protective strategy against pneumococcal infections. Cell surface proteins are key factors in infectious processes of pathogens and can be considered as vaccine candidates [4,5]. LPxTG motif containing proteins are a group of cell surface proteins in gram positive bacteria and are first synthesized as a preprotein containing an N-terminal signal peptide and a C-terminal cell wall anchor domain. The anchor domain is attached to the cell wall by a sortase enzyme [5,6]. Lipoproteins constitute another group of cell surface proteins in gram positive bacteria. These proteins are lipid modified proteins produced as secretory precursors with a signal peptide linked to their amino termini. The C-terminal region of their signal peptide

contains a well conserve lipobox motif of four amino acid residues and invariably, the last residue is cysteine. The covalent binding of lipoproteins to the cell membrane is achieved by a diacylglycerol transferase [7,8]. Choline binding proteins (CBPs) are a family of surface proteins noncovalently bound to the phosphorylcholine moiety of the cell wall via a conserved choline binding domain. LPxTG motif containing proteins, lipoproteins and CBPs are among main groups of pneumococcal cell surface proteins [5].

Conventional experimental approaches for vaccine discovery cultivate pathogens under laboratory conditions, dissecting them into their main component proteins. Antigens offering protective immunity are then identified by testing individual components. The antigens also need to be produced on a large scale, making the process extremely time consuming [9]. However, bioinformatic approaches for identification of vaccine candidates offer the ability to undertake a rapid and comprehensive assessment of a microorganism's surface proteins [10]. To our knowledge, there is no report on bioinformatic analysis of surface proteins of a serotype 19F *S. pneumoniae*. In this study therefore, we aim to determine appropriate vaccine candidates among the above mentioned three groups of surface proteins in *S. pneumoniae* serotype 19F to develop protein based vaccines against this pneumococcal serotype. The pneumococcal surface proteins were therefore determined using bioinformatic tools. Then suitability of the proteins as vaccine candidates was evaluated in terms of their antigenicity and the absence of autoimmunity induction.

2. Materials and Methods

Protein BLAST was used to search for LPxTG consensus sequence in the proteome data of *S. pneumoniae* Taiwan 19F-14. CW-PRED tool was used to determine the cell surface anchoring domain in these proteins. Cellular localization of the proteins and the signal peptide presence were specified using PSORTb v.3.0 program and Signal P 4.1 tool respectively.

The predicted lipoproteins of *S. pneumoniae* D39, *S. pneumoniae* R6 and *S. pneumoniae* TIGR4 available in PRED-LIPO were used to specify conserved motifs of lipobox. Protein BLAST was used to search for these motifs in proteome data of *S. pneumoniae* Taiwan 19F-14. The PRED-LIPO tool was used to verify the presence of lipoprotein specific signal peptide in these proteins.

Protein BLAST was done to find proteins of *S. pneumoniae* Taiwan 19F-14 containing the consensus sequence of the choline binding domain (CBD) (TGWVKINGNWYYFDSNGVM) available in Pfam. The cellular localization of these proteins was determined by PSORTb v.3.0.

VaxiJen v.2.0 was used to evaluate the antigenicity of the proteins. Blast against human proteome was carried out to exclude the possibility of inducing autoimmunity by the proteins.

3. Results and Discussion

3.1. LPxTG Motif Containing Proteins

Search for LPxTG motif in proteome data of *S. pneumoniae* Taiwan 19F-14 demonstrated the presence of 25 proteins containing this motif. LPxTG motif in 20 of these proteins resides in the C-terminal region and it is in the N-terminal region of the rest of the proteins. The LPxTG motif in cell surface proteins has been reported to reside in the C-terminal region. Therefore, we continued our analyses with the 20 proteins possessing this status. Among these proteins, the presence of cell surface anchoring domain was confirmed in 12 proteins. PSORTb v.3.0 results demonstrated the cell wall localization for these proteins. The Signal P 4.1 analysis indicated the presence of signal peptides belonging to YSIRK family of signal peptides in 7 proteins (Table 1). Signal peptides direct proteins towards secretory pathways and are located in amino terminal regions of these proteins. Evaluation of antigenicity of these proteins using VaxiJen v.2.0 demonstrated their capabilities as protective antigens. However, cell wall surface anchor family protein under accession number YP_002741626.1 showed the highest antigenic probability.

Table 1. Pneumococcal cell surface proteins containing LPxTG motifs

| Protein | Accession No. | Signal peptide | Protective antigen probability (%) | Homology with human proteom (max %) |
|---|----------------|----------------|------------------------------------|--------------------------------------|
| MucBP domain family | YP_002742835.1 | No | ND * | ND |
| Cell wall surface anchor family protein | YP_002743319.1 | No | ND | ND |
| Cell wall surface anchor family protein | YP_002743140.1 | Yes | 0.6150 | 0 |
| Neuraminidase | YP_002743027.1 | No | ND | ND |
| Cell wall surface anchor family protein | YP_002741626.1 | Yes | 0.6742 | 0 |
| Hyaluronate lyase | YP_002741863.1 | Yes | 0.4148 | 32 |
| Pullulanase | YP_002741814.1 | Yes | 0.555 | 41 |
| Beta-N-acetyl-hexoseaminidase | YP_002741607.1 | Yes | 0.4494 | 35 |
| Endo-beta-N-acetylglucosaminidase D | YP_002742025.1 | No | ND | ND |
| Endo-alpha-N-acetylgalactosaminidase | YP_002741907.1 | Yes | 0.6058 | 44 |
| Serine protease | YP_002742145.1 | Yes | 0.6338 | 33 |
| Beta-galactosidase | YP_002742149.1 | No | ND | ND |

* Not determined.

Antigenic similarity of either amino acid sequences or the conformational structure between antigens of pathogenic agents and host tissues might induce an immune response against the shared determinant. Therefore, the host tolerance to these antigens

(autoantigens) breaks down and the elicited autoantigen specific immune response cross reacts with host structures to cause tissue damage and disease [11]. As a result, we excluded the possibility for induction of autoimmunity by analysis of homology between pneumococcal antigens and

human proteome via BLAST. The cell wall surface anchor family protein under accession number YP_002741626.1 showed no homology to human proteome. The homolog proteins of this protein found in other pneumococcal strains belonging to serotype 19F including *S. pneumoniae* A026, *S. pneumoniae* G54, and *S. pneumoniae* ST556 with homology percentages in the range of 98-100%. The predicted function for the cell wall surface anchor family protein obtained via BLAST was fibronectin binding. Our results therefore demonstrated that this protein can be considered as an appropriate vaccine candidate for development of protein based vaccines against serotype 19F pneumococci.

3.2. Lipoproteins

Conserved motifs of pneumococcal lipobox were found in 61 proteins of *S. pneumoniae* Taiwan 19F-14. Results of PRED-LIPO analysis indicated that 34 proteins among them contain lipoprotein signal peptide and are thus

considered as lipoproteins. PSORTb v.3.0 demonstrated the extracytoplasmic localization for 31 proteins among the predicted lipoproteins and analysis with VaxiJen v.2.0 showed that only 20 proteins were probable protective antigens (Table 2). The two most probable antigens were foldase protein PrsA and D-alanyl-D-alanyl-carboxy peptidase. BLAST results showed that homology percentages of foldase protein PrsA and D-alanyl-D-alanyl-carboxy peptidase with the human proteome were 28% and 0% respectively. The homolog proteins of D-alanyl-D-alanyl-carboxy peptidase were available in other pneumococcal serotype 19F strains including *S. pneumoniae* A026, *S. pneumoniae* G54, and *S. pneumoniae* ST556 with homology percentages in the range of 88-100%. These results demonstrated that D-alanyl-D-alanyl-carboxy peptidase is an appropriate protein based vaccine candidate. This protein is involved in the control of peptidoglycan crosslinking extents [12].

Table 2. Pneumococcal antigenic lipoproteins

| Lipoproteins | Accession No. | Lipobox motif | Cellular location | Protective antigen probability | Human proteome homology (max%) |
|---|----------------|---------------|----------------------|--------------------------------|--------------------------------|
| D-alanyl-D-alanine carboxypeptidase | YP_002742135.1 | LAAC | Extracytoplasmic | 0.6625 | 0 |
| Amino acid ABC transporter substrate binding protein | YP_002741699.1 | LAAC | Extracytoplasmic | 0.5334 | 29 |
| Lipoprotein | YP_002741700.1 | LAAC | Cytoplasmic membrane | 0.5158 | 29 |
| Foldase protein PrsA | YP_002742641.1 | LAAC | Cytoplasmic membrane | 0.7132 | 28 |
| Lipoprotein | YP_002742768.1 | LAAC | Extracytoplasmic | 0.6122 | 27 |
| ABC transporter substrate binding protein | YP_002742232.1 | LAAC | Extracytoplasmic | 0.6442 | 26 |
| Sugar ABC transporter substrate binding protein | YP_002743018.1 | LAAC | Extracytoplasmic | 0.5216 | 30 |
| AmiA protein | YP_002743192.1 | LAAC | Cell wall | 0.5679 | 29 |
| Lipoprotein | YP_002741745.1 | LVGC | Extracytoplasmic | 0.5626 | 40 |
| ABC transporter substrate binding protein-glutamine transport | YP_002742126.1 | LVAC | Extracytoplasmic | 0.5218 | 47 |
| ABC transporter substrate binding protein-phosphate | YP_002743411.1 | LVAC | Extracytoplasmic | 0.6127 | 28 |
| Manganese ABC transporter substrate binding lipoprotein | YP_002742988.1 | LVAC | Cytoplasmic membrane | 0.5038 | 27 |
| Zinc binding lipoprotein AdcA | YP_002743496 | LVAC | Cytoplasmic membrane | 0.5742 | 0 |
| Thioredoxin family protein | YP_002742159.1 | LTAC | Cell wall | 0.6399 | 49 |
| Peptidyl-prolyl cis trans isomerase | YP_002742253.1 | LAGC | Extracytoplasmic | 0.6314 | 45 |
| Thioredoxin family protein | YP_002742624.1 | LMAC | Cell wall | 0.5867 | 58 |
| Zinc binding lipoprotein AdcA | YP_002742622.1 | LGAC | Cytoplasmic membrane | 0.5797 | 31 |
| Iron compound ABC transporter | YP_002743174.1 | LGAC | Cytoplasmic membrane | 0.5961 | 29 |
| Oligopeptide binding protein AliB | YP_002742868.1 | LSAC | Cell wall | 0.5664 | 34 |
| Oxa-like protein precursor | YP_002743302.1 | LTGC | Cytoplasmic membrane | 0.5023 | 27 |

3.3. Choline Binding Proteins

The consensus sequence of CBD was found in 100 proteins of *S. pneumoniae* Taiwan 19F-14. Extracellular, cell wall and extracytoplasmic localizations were demonstrated by PSORTb v.3.0 for 7, 5, and 8 proteins respectively (Table 3). Thirteen proteins among them were antigenic. However, only 5 proteins did not show similarity to the human proteome and homologs of just three proteins among them including autolysin, surface protein PspC, and choline binding protein D (CBPD) were available in other pneumococcal serotype 19F strains including *S. pneumoniae* A026, *S. pneumoniae* G54, and *S. pneumoniae* ST556 with homology percentages of more than 70%. The antigenicity of PspC was more than that of CBPD. However, its homology percentage to the similar protein in the above mentioned three pneumococcal strains (71-90%) was less than that of CBPD (99-100%).

PspC promotes uptake of pneumococci into nasopharyngeal epithelial cells and interferes with components of the innate immune systems [5]. Vaccination with surface protein PspC derived from *S. pneumoniae* serotype 2 has shown to be protective against sepsis in mice caused by the same pneumococcal serotype [13]. Our results also indicated the suitability of surface protein PspC and CBPD as antigenic candidates for development of proteinaceous vaccines against *S. pneumoniae* serotype 19F infection. CBPD is a key component of competence induced DNA release process of pneumococci [14].

4. Conclusions

Among 186 proteins entered our bioinformatic based analysis for selection of vaccine candidates against serotype 19F pneumococcal infection, four proteins

including cell wall surface anchor family protein under accession number YP_002741626.1, D-alanyl-D-alanyl-carboxy peptidase, surface protein PspC, and CBPD were appropriate candidates. For future studies, these proteins

will be expressed in suitable hosts and their protective abilities against serotype 19F pneumococci will be evaluated experimentally.

Table 3. Extracytoplasmic pneumococcal choline binding proteins

| Protein | Accession No. | Cellular location | Protective antigen probability | Human proteom homology (max%) |
|---|----------------|-------------------|--------------------------------|-------------------------------|
| Pneum surface protein A | YP_002741673.1 | Extracellular | 0.6613 | 27 |
| Choline binding protein F | YP_002741916.1 | Extracellular | 0 | ND * |
| Autolysin | YP_002743245.1 | Extracellular | 0.5216 | 0 |
| Surface protein PspC | YP_002743515.1 | Extracellular | 0.7933 | 0 |
| Choline binding protein J | YP_002741917.1 | Extracellular | 0 | ND |
| Lysozyme | YP_002742915.1 | Extracellular | 0.5068 | 29 |
| Endo-beta-N-acetylglucosaminidase | YP_002742657.1 | Extracellular | 0 | ND |
| Beta-N-acetyl-hexosaminidase | YP_002741607.1 | Cell Wall | 0 | ND |
| G5 domain family | YP_002742888.1 | Cell Wall | 1 | 61 |
| Endo-alpha-N-acetyl-galactosaminidase | YP_002741907.1 | Cell Wall | 0.6053 | 44 |
| Serine protease | YP_002742145.1 | Cell Wall | 0.6338 | 35 |
| Beta-galactosidase | YP_002742149.1 | Cell Wall | 0.5874 | 26 |
| Hypothetical protein SPT_0691 | YP_002742167.1 | Extracytoplasmic | 0 | ND |
| Choline binding protein D | YP_002743526.1 | Extracytoplasmic | 0.583 | 0 |
| Choline binding protein PcpA | YP_002743464.1 | Extracytoplasmic | 0 | ND |
| ABC transporter substrate binding protein | YP_002742343.1 | Extracytoplasmic | 0 | ND |
| Cell wall surface anchor family protein | YP_002741994.1 | Extracytoplasmic | 0.5419 | 49 |
| Pneumococcal histidine triad protein B | YP_002742621.1 | Extracytoplasmic | 0.6095 | 0 |
| Pneumococcal histidine triad protein B | YP_002742492.1 | Extracytoplasmic | 0.6033 | 53 |
| Hypothetical protein SPT_1059 | YP_002742500.1 | Extracytoplasmic | 0.632 | 0 |

* Not determined.

References

- [1] Mook-Kanamori, B. B., Geldhoff, M., van der Poll, T. and van de Beek, D., "Pathogenesis and pathophysiology of pneumococcal meningitis," *Clin Microbiol Rev*, 24 (3), 557-591, Jul. 2011.
- [2] WHO, "Pneumococcal vaccines, WHO position paper," *Wkly Epidemiol Rec*, 87 (14), 129-144, Apr. 2012.
- [3] Johnson, H. L., Deloria-Knoll, M., Levin, R. S. et al., "Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project," *PLoS Med*, 7 (10), e1000348, Oct. 2010.
- [4] Foster, T. J., Geoghegan J. A., Ganesh, V. K. and Hook, M., "Adhesion, invasion and evasion: the many functions of surface proteins of *Staphylococcus aureus*," *Nature Rev Microbiol*, 12 (1), 49-62, Dec. 2014.
- [5] Bergmann, S. and Hammerschmidt, S., "Versatility of pneumococcal surface proteins," *Microbiol*, 152 (2), 295-303, Feb. 2006.
- [6] Leenhouts, K., Buist, G. and Kok, J., "Anchoring of proteins to lactic acid bacteria," *Antonie van Leeuwen*, 76 (1-4), 367-376, Jul.-Nov. 1999.
- [7] Venema, R., Tjalsma, H., van Dijk, J. M. et al., "Active lipoprotein precursors in the gram positive eubacterium *Lactococcus lactis*," *J Biol Chem*, 278 (17), 14739-14746, Apr. 2003.
- [8] Tjalsma, H., Kontinen, V. P., Pragai, Z. et al., "The role of lipoprotein processing by signal peptidase II in the gram-positive eubacterium *Bacillus subtilis*," *J Biol Chem*, 274 (3), 1698-1707, Jan. 1998.
- [9] Davis, M. N. and Flower, D. R., "Harnessing bioinformatics to discover new vaccines," *Drug Discov Today*, 12 (9-10), 389-395, May 2007.
- [10] Kelly, D. F. and Rappuoli, R., *Hot topics in infection and immunity in children II*, Springer, New York, 2005, 217-223.
- [11] Blank, M., Krause, I., Fridkin, M. et al., "Bacterial induction of autoantibodies to β 2-glycoprotein-I accounts for the infectious etiology of antiphospholipid syndrome," *J Clin Invest*, 109 (6), 797-804, March 2002.
- [12] Scheffers, D.-J. and Pinho, M. G., "Bacterial cell wall synthesis: new insights from localization studies," *Microb Mol Biol Rev*, 69 (4), 585-607, Dec. 2005.
- [13] Ogunniyi, A. D., Woodrow, M. C., Poolman, J. T., Paton, J.C., "Protection against *Streptococcus pneumoniae* elicited by immunization with pneumolysin and CbpA," *Infect Immun*, 69 (10), 5997-6003, Oct. 2001.
- [14] Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E. and Havarstein, L. S., "Choline-binding protein D (CbpD) in *Streptococcus pneumoniae* is essential for competence induced cell lysis," *J Bacteriol*, 187 (13), 4338-4345, July 2005.