American Journal of Microbiological Research, 2015, Vol. 3, No. 6, 190-196
Available online at http://pubs.sciepub.com/ajmr/3/6/3

DOI:10.12691/ajmr-3-6-3

In *silico* Analysis of Surface Proteins of *Streptococcus pneumoniae* Serotype 19F for Identification of Immunoprotective Epitopes

Shirin Tarahomjoo*

Division of Genomics and Genetic Engineering, Department of Biotechnology and Central Laboratory, Razi Vaccine and Serum Research Institute, Karaj, Iran
*Corresponding author: starahomjoo@hotmail.com

Abstract Pneumococcal conjugate vaccines (PCVs) were developed through chemical coupling of polysaccharide capsules of pneumococci to immunogenic carrier proteins and World Health Organization recommends inclusion of these vaccines in national immunization programs for children. However, the PCVs implementation in developing countries can be prevented by the high manufacturing costs. This issue can be overcome by construction of protein based vaccines against pneumococci. We already identified three pneumococcal surface proteins including D-alanyl-D-alanine-carboxy peptidase (DDCP), choline binding protein D (CBPD), and cell wall surface anchor family protein (CWSAP) as appropriate protein candidates for eliciting protection against *S. pneumoniae* serotype 19F. The protein protective antigenicity, the absence of autoimmunity induction, and the amino acid sequence conservancy in serotype 19F pneumococcal strains were used as selection criteria. However, regarding the requirement of both antibody and cellular immune responses for protection against pneumococci, analysis of protective B and T-cell epitopes of these proteins is necessary to examine their usefulness in new vaccine formulations. In the present study, therefore, we aim to identify protective epitopes of these proteins via widely used bioinformatic tools. The Bepipred program was used for identification of linear B-cell epitopes. The conformational B-cell epitopes were predicted using the CBTope program. T-cell epitopes were identified using the Immune Epitope Database tool. The immunoprotective abilities of epitopes were evaluated using VaxiJen. Our results showed that all of the three studied proteins included protective epitopes. However, the greatest number of epitopes was identified in a truncated form of CWSAP. Moreover, the most probable immunoprotective epitopes reside in this protein and these epitopes were highly conserved in CWSAPs of the most common pneumococcal serotypes in the world. Therefore, the truncated CWSAP was an appropriate candidate for development of protein based vaccines against the most common pneumococcal serotypes.

Keywords: *cell surface proteins, epitopes, pneumococcal serotypes, pneumococcal vaccines, Streptococcus pneumoniae*

Cite This Article: Shirin Tarahomjoo, "In *silico* Analysis of Surface Proteins of *Streptococcus pneumoniae* Serotype 19F for Identification of Immunoprotective Epitopes." *American Journal of Microbiological Research*, vol. 3, no. 6 (2015): 190-196 doi: 10.12691/ajmr-3-6-3.

1. Introduction

Streptococcus pneumoniae is a major pathogen that causes diseases such as pneumoniae, meningitis and sepsis most commonly in children under 5 years of age, especially those under 2 years of age [1,2]. World Health Organization (WHO) estimated that 476000 annual deaths among children less than 5 years of age were caused by pneumococcal infections [3]. *S. pneumoniae* serotype 19F is among the main pneumococcal serotypes that cause most invasive pneumococcal disease in children less than 5 years of age globally [2]. Pneumococcal vaccines have been used for protection against pneumococcal infections. The polysaccharide capsules of pneumococci are main antigenic components of these vaccines. However, the capsules are poorly immunogenic in children less than 2 years and are not able to induce anamnestic antibody responses upon revaccination. In order to overcome these shortcomings, the capsules are chemically conjugated to immunogenic carrier proteins. Therefore, the capsules are converted from T-cell independent antigens to T-cell dependent antigens, which enhance antibody responses and induce the immune memory. WHO recommends inclusion of pneumococcal conjugate vaccines (PCVs) in national immunization programs for children [2]. 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 have extensively been evaluated as vaccine candidates [4,5]. We already identified three pneumococcal surface proteins including D-alanyl-D-alanine-carboxy peptidase (DDCP), choline

binding protein D (CBPD), and cell wall surface anchor family protein (CWSAP) as appropriate protein candidates for eliciting protection against *S. pneumoniae* serotype 19F. The protein selection was done based on the protective antigenicity, the absence of autoimmunity induction, and the amino acid sequence conservancy in serotype 19F pneumococcal strains [6]. Both antibody and cell mediated immune responses are required to protect against pneumococcal infections [7]. Thus, analysis of immunoprotective epitopes of these proteins is necessary to evaluate their effectiveness as new vaccine constituents. Compared with conventional laboratory methods, computational approaches offer the ability to undertake rapid and comprehensive epitope assessments for vaccine candidates at much lower costs [8]. In the present study, therefore, we analyzed DDCP, CBPD, and CWSAP via widely used bioinformatic tools for identification of immunoprotective epitopes.

2. Materials and Methods

Accession numbers of DDCP, CBPD, and CWSAP are WP_000747819, ACO22665, and WP_000671416 respectively. Conserved domains of proteins were specified using Pfam. Protein sequences were submitted to the Bepipred program for identification of linear B-cell epitopes. The immunoprotective abilities of selected B-cell epitopes were evaluated using VaxiJen. A higher VaxiJen score refers to a higher probability for protective ability [9]. The CBTope program was used to predict conformational B-cell epitopes using amino acid sequences of the proteins.

T-cell epitopes binding to DRB1*0101 and DRB1*0401 were identified using the MHC class II T-cell epitope prediction tool available at the Immune Epitope Database (IEDB). The prediction was performed using artificial neural network (ANN) method. T-cell epitopes were classified based on the binding affinity for MHC alleles using the half maximum inhibitory concentration (IC_{50}) value. IC_{50}s of < 50 nM indicate high affinity binder epitopes and IC_{50}s of < 500 nM indicate intermediate affinity binder epitopes. A lower IC_{50} indicates a higher affinity of binding to MHC allels. The immunoprotective T-cell epitopes were identified using VaxiJen.

SIM alignment tool was used for the epitope conservancy analysis in CWSAPs of different pneumococcal serotypes. Accession numbers of the proteins are ADM90238 (serotype 6B), COP01275 (serotype 23F), ACO18463 (serotype 14), EDT51092 (serotype 6A), ACO17525 (serotype 5), ACO21750 (serotype 1), and ACB89332 (serotype 14), ACA35889 (serotype 19A), EDT92624 (serotype 9V), EDK67540 (serotype 18C), ABJ53732 (serotype 2), AAK74270 (serotype 4), EDT90844 (serotype 7F), CFV12342 (serotype 12F), COA29153 (serotype 3), CIT03281 (serotype 8), CNZ75298 (serotype 46), CIT23802 (serotype 15B), and CON04543 (serotype 45). The default settings were applied to all the tools used.

3. Results and Discussion

3.1. Identification of B- and T-cell Epitopes of DDCP

Antibodies bind specifically to a continuous amino acid sequence of a protein known as the linear B-cell epitope or to a folded structure formed by discontinuous amino acids known as the conformational B-cell epitopes. The majority of B-cell epitopes are conformational [10]. Nevertheless, the identification of linear B-cell epitopes has demonstrated promising results for selection of the vaccine antigens [11,12]. Therefore, in this study, we determined both types of the B-cell epitopes in the proteins.

DDCP is a low molecular weight (24 kDa) penicillin binding protein, which is involved in the control of peptidoglycan crosslinking extent [13]. DDCP has a lipoprotein type signal peptide and the mature protein obtained following cleavage of the signal peptide between amino acids 18 and 19 was used in our epitope analysis [6]. Seven linear B-cell epitopes were identified in the protein using the Bepipred program. VaxiJen results indicated that 3 epitopes among them were immunoprotective (Table 1). Regarding the VaxiJen scores, the linear B-cell epitope beginning at the first amino acid of the protein shows the highest probability for the protective ability.

Table 1. Immunoprotective linear B-cell epitopes of mature DDCP

Position[1]	Sequence[2]	Vaxijen score
1	**CSQEKTKVEENTQKTEQSSQPEGTVGSKSQASSQKKAEVSNKGSYYSIQG**	1.4456
62	YPLSKDYNPGENPTAK	0.5208
114	YVNQDGKEAADRYSARPGYSEHQ	1.347

1: The amino acid residue number at the epitope beginning is indicated.
2: The most probable protective epitope is shown in bold.

Table 2. Conformational B-cell epitopes of mature DDCP

Position[1]	No. of epitope residues	Epitope region sequence[2]
1	6	*CSQEKT*
15	14	*TEQSSQPEGTVGSKSQAS*
60	28	KR*YPL SKDYNPGENPTA*KAELLKLI AAM
106	30	TQAKLYQD*YVNQDGKEAADRYSARPGYSEH*
148	11	GDLVTEEKAAQWLLDH

1: Amino acid residue numbers at the beginning of epitope regions are indicated.
2: Amino acid residues of conformational B-cell epitopes in each region are shown underlined. Amino acids shown in italics belong to linear B-cell epitopes at the same region.

The CBTope program uses the protein amino acid sequence for the prediction of conformational B-cell epitopes. Using this program, 5 conformational B-cell epitope regions were identified in the protein (Table 2). The epitope region beginning at the amino acid residue 148 of the mature DDCP is exclusively conformational. However, the conformational epitope regions beginning at the amino acids 1 and 15 are parts of the most probable protective linear B-cell epitope. The conformational epitope beginning at the amino acid 60 contains a complete linear B-cell epitope. The conformational epitope beginning at the amino acid 106 contains a linear B-cell epitope lacking the last amino acid residue. These results indicated that the majority of conformational B-cell epitopes overlapped with the linear B-cell epitopes.

Mucosal colonization by *S. pneumoniae* is a perquisite for otitis media and lung infections. $CD4^+$ T-cell responses mediate resistance to the mucosal colonization by *S. pneumoniae* [7]. DRB1*0101 and DRB1*0401 are the most common MHC class II alleles in the human population [14]. Therefore, in this study, we determined the presence of T-cell epitopes binding to these MHC alleles in the proteins.

The mature DDCP contains 3 protective DRB1*0101 T-cell epitopes with high affinity binding (Table 3). Considering the VaxiJen score, YYSIQGKYD is the most probable immunoprotective epitope. This epitope has the lowest IC50 value, which indicates it has the highest affinity of binding to DRB1*0101. Moreover, the protein contains 12 T-cell epitopes with intermediate affinity of binding to DRB1*0101 (data not shown). FRSYETQAK is the only T-cell epitope in the protein with high affinity of binding to DRB1*0401. The protein also contains 3 DRB1*0401 T-cell epitopes with intermediate affinity binding (data not shown).

Table 3. Immunoprotective T-cell epitopes of mature DDCP with high affinity binding

MHC allele	Sequence[1]	IC50 (nM)	VaxiJen score	Position[2]
DRB1*0101	**YYSIQGKYD**	7.2	1.3186	45
	IAAMQAEGY	8.1	0.5903	84
	YVGKEAKEI	19.1	0.5773	192
DRB1*0401	FRSYETQAK	14.7	1.3839	98

1: The most probable immunoprotective epitope is shown in bold.
2: The amino acid residue number at the epitope beginning is indicated.

3.2. Identification of B- and T-cell Epitopes of CBPD

CBPD is involved in promoting pneumococcal colonization of the nasopharynx [15]. Pfam analysis showed that CBPD consists of an N-terminal CHAP domain followed by two bacterial SH3 domains, and a C-terminal choline binding domain (CBD) comprising four repeats (Figure 1). The CHAP domain corresponds to an amidase function. CBD mediates the protein noncovalent attachment to choline moieties decorating the pneumococcal cell surface. Bacterial SH3 domains recognize the bacterial peptidoglycan [15].

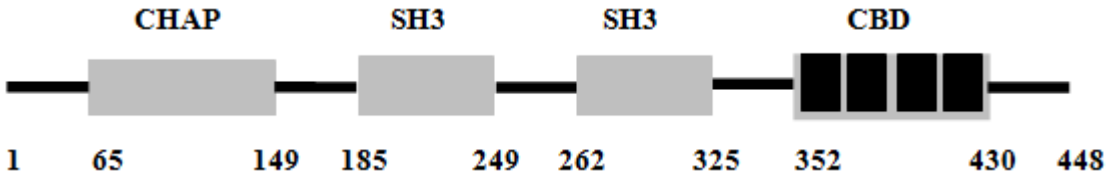

Figure 1. Schematic representation of CBPD. Four repeats of CBD are shown as black bars in the domain. Amino acid residue numbers are indicated

Thirteen linear B-cell epitopes were identified in CBPD. Nine epitopes among them are immunoprotective as indicated by VaxiJen results (Table 4). The linear B-cell epitope beginning at the amino acid 175 has the highest probability for the protective ability. This epitope resides partially in the first bacterial SH3 domain.

Table 4. Immunoprotective linear B-cell epitopes of CBPD

Position[1]	Sequence[2]	Vaxijen score
38	VYAYSRGNGSIARGDDYPAYYKNGSQEID	0.6125
91	IPAAYGNANEWGHRARREGYRVDNTPTIGSI	0.5929
123	WSTAGTY	0.9387
148	YNYGYTESY	0.6308
175	**LDGGSVGNSQSSTSTGGTHYF**	1.8539
215	YYPGEKVH	0.5095
261	VLSSTGGTHYF	1.0932
291	YYPGEKVH	0.5095
328	VTSSQNYQNQSGNISSYGSNNSSTVG	1.161

1: The amino acid residue number at the epitope start is indicated.
2: The most probable protective epitope is shown in bold.

In the CBTope program, seven conformational B-cell epitope regions were recognized in the protein (Table 5). The epitope region beginning at the amino acid residue 176 resides entirely in the most probable protective linear

B-cell epitope of CBPD. The epitope regions beginning at the amino acids 356 and 435 are exclusively conformational. The rest of the epitope regions are entirely or partially located in the identified linear B-cell epitopes. These results showed that the majority of conformational B-cell epitopes overlapped with the linear B-cell epitopes.

Table 5. Conformational B-cell epitopes of CBPD

Position[1]	No. of epitope residues	Epitope region sequence[2]
50	22	*RGDDYPAYYKNGSQEIDQ*WRMYSRQCTSFVAFR
107	9	*REGYRVDNT*
176	14	*DGGSVGNSQSSTSTGGTH*
229	34	KDGYKWLSYTAYNGSYRYVQLEAVNKNPLGNS*VLSSTGGTH*
305	27	KDGYKWLSYTAYNGSRRYIQLEG*VTSSQ*
356	39	KINGS WYHFKSNGSKSTGWLKDGSSWYYLKLSGEMQTGWLKENGSWYYLGSSGAMKTGWYQVS
435	14	TVDGYRVNSDGERV

1: Amino acid residue numbers at the beginning of epitope regions are indicated.
2: Amino acid residues of conformational B-cell epitopes in each region are shown underlined. Amino acids shown in italics belong to linear B-cell epitopes at the same region.

CBPD contains 16 DRB1*0101 T-cell epitopes with high affinity binding (Table 6). Two epitopes beginning at the amino acids 246 and 244, two epitopes beginning at the amino acids 383 and 381, and four epitopes beginning at the amino acids 30, 32, 29, and 25 are overlapped. Vaxijen scores showed that YQNQSGNIS was the most probable protective epitope among the identified epitopes. This epitope resides between the second bacterial SH3 domain and CBD. YIQLEGVTS has the lowest IC50 value, which indicates it has the highest affinity of binding to DRB1*0101. However, its VaxiJen score is lower than that of YQNQSGNIS. In addition, the protein contains 25 DRB1*0101 T-cell epitopes with intermediate affinity binding (data not shown). The CBPD analysis for identification of DRB1*0401 T-cell epitopes indicated the presence of 6 and 21 epitopes with high and intermediate affinity binding respectively. The most probable protective DRB1*0401 T-cell epitope (WYHFKSNGS) is located in the first repeat of CBD at the protein C-terminus. Three epitopes including YIQLEGVTS, YVQLEAVNK, and FVAFRLSNV are able to bind to both DRB1*0101 and DRB1*0401.

Table 6. Immunoprotective T-cell epitopes of CBPD with high affinity binding

MHC allele	Sequence[1]	IC50 (nM)	VaxiJen score	Position[2]
DRB1*0101	*YIQLEGVTS*	4.5	1.0994	322
	WYYLGSSGA	5.1	0.6779	401
	YVQLEAVNK	5.9	0.496	246
	YLKLSGEMQ	6	0.7775	383
	YRYVQLEAV	6.4	0.6505	244
	LMLAAGDSV	9.1	0.4188	30
	IARGTSYYL	11	0.9228	7
	LAAGDSVYA	11.5	1.0481	32
	FKTKSAIKT	11.8	0.753	195
	FVAFRLSNV	11.9	0.841	78
	GLMLAAGDS	12.7	0.4753	29
	FLVVGLMLA	29.7	0.6527	25
	YTAYNGSRR	32.1	0.4462	313
	YYKNGSQEI	36.7	0.6354	57
	YQNQSGNIS	39	1.1567	334
	WYYLKLSGE	46.9	1.0789	381
DRB1*0401	*YIQLEGVTS*	12.5	1.0994	322
	YVQLEAVNK	24.7	0.496	246
	FVAFRLSNV	29.9	0.841	78
	YRVDNTPTI	31.8	0.758	110
	FRLSNVNGF	39.7	0.4186	80
	WYHFKSNGS	48.6	1.3108	361

1: The common epitope of MHC alleles is shown in italics. The most probable immunoprotective epitopes are shown in bold.
2: The amino acid residue number at the epitope beginning is indicated.

3.3. Identification of B- and T-cell Epitopes of CWSAP

The search for conserved domains within CWSAP using Pfam showed the presence of five fibronectin binding repeats (FBRs). These repeats bind to fibronectin, which is an extracellular matrix glycoprotein [16]. CWSAP contains a YSIRK type family signal peptide, which is cleaved between the amino acids 42 and 43. Moreover, it contains an LPxTG motif beginning at the amino acid residue 973 [6]. FBRs reside between the

amino acids 213 to 901 of the protein (Figure 2). The first FBR (FBR1) contains 81 amino acid residues, whereas the other repeats contain 83 amino acid residues. The amino acid sequences of FBR2-FBR5 are identical except for the fourth amino acid residue of FBR3. However, FBR1 differs from the four last repeats in 23 amino acid residues.

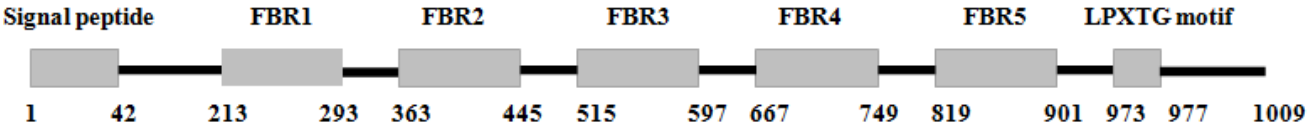

Figure 2. Schematic representation of CWSAP. Amino acid residue numbers are indicated

The truncated CWSAP containing the amino acids 43-972 was used for the epitope analysis. The linear B-cell epitope analysis using Bepipred demonstrated the presence of 32 linear B- cell epitopes in the protein. VaxiJen results indicated that twenty four epitopes among them were protective (Table 7). FBR1 contains three B-cell epitopes. Considering VaxiJen scores, the linear B-cell epitope of the protein with the highest probability for the protective ability (YGNKDGKAD) resides in FBR1. Each of FBR2-FBR5 contains one partial linear B-cell epitope and two complete linear B-cell epitopes.

Table 7. Protective linear B-cell epitopes of truncated CWSAP

Position[1]	Sequence[2]	Vaxijen score
1	DVVNPTPGQVLPEETSGTKEGDLSEKPGDTVLTQAKPEGVTGNTNSLPTPTERTEVSEETNSSS	1.169
70	EKDEEAQENPELTDA	1.442
86	KETVDTADVDGTQASPAETTPEQVKGGVKENTKDSIDV	1.3233
128	LEKAEGKGPFTAG	0.9492
197	LNGNTVGKQ	2.4913
215	ANGTQTYK	1.823
228	**YGNKDGKAD**	3.0921
278, 430, 582, 734	LEKAKGEGPFTA	0.4999
315, 619, 771	DKAPWSDNGEAKNPALSPLGENVKTK	0.4802
367, 519, 671, 823	ANGTQTYSA	1.228
378, 530, 682, 834	NVYGNKDGKPDLD	0.8914
467	DKAPWSDNGDAKNPALSPLGENVKTK	0.5127
861	KETSDTANGSLSPSNSGSGVTPMNHNHATGTTDSMPADTMTSSTNTMAGENMAASANKM	1.0833

1: The amino acid residue number at the epitope beginning is indicated.
2: The most probable protective epitope is shown in bold.

CBTope analysis revealed the presence of 15 conformational B-cell epitope regions in the protein (Table 8). Thirteen epitopes among them contain entirely or partially amino acid residues of the linear B-cell epitopes. However, two of these epitopes (LLLKASDNAPWSDNGTA and TVEKAVK) are exclusively conformational. The most probable protective linear B-cell epitope (YGNKDGKAD) resides in the conformational B-cell epitope region beginning at the amino acid residue 219.

Table 8. Conformational B-cell epitopes of truncated CWSAP

Position[1]	No. of epitope residues	Epitope region sequence[2]
3	5	*VN*PT*PGQ*
79	20	*ELTDAL*K*ETVDTADVDGTQASPAETTPEQVKGGV*
159	14	LLLKASDNAPWSDNGTA
182	22	LEGLTKGKYFYEVD*LNGNTVGKQ*GQALID
219	22	*QTYK*ATVKV*YGNKDGKAD*LTNL
258	7	TVEKAVK
307, 611, 763	42	TRLLLKAS*DKAPWSDNGEAKNPALSPLGENVKTK* GQYFYQVALDGN
369, 521, 672, 825	21	*GTQTYSA*TV*NVYGNKDGKPDLDNI*
459	41	TRLLLKAS*DKAPWSDNGDAKNPALSPLGENVK* *TK*GQYFYQVALDGN
859	6	NV*KETSDT*

1: Amino acid residue numbers at the beginning of epitope regions are indicated.
2: The amino acid residues of conformational B-cell epitopes in each region are shown underlined. Amino acids shown in italics belong to linear B-cell epitopes at the same region.

Our analysis using IEDB MHC class II T-cell epitope prediction tool and VaxiJen indicated the presence of 22 protective T-cell epitopes, which showed high affinity binding to DRB1*0101 (Table 9). Some of the identified epitopes are overlapped. The most probable immunoprotective DRB1*0101 T-cell epitope (LVATKNVDI) resides in FBR1. IDQLRANGT shows the lowest IC50 value, which indicates it has the highest

affinity of binding to DRB1*0101. However, its VaxiJen score is lower than that of LVATKNVDI. In addition, the protein contains 50 protective T-cell epitopes with intermediate affinity binding for DRB1*0101 (data not shown). Analysis of the protein for immunoprotective DRB1*0401 T-cell epitopes revealed the presence of 9 epitopes with high affinity binding and 34 epitopes with intermediate affinity binding. The most probable protective T-cell epitope with high affinity binding to DRB1*0401 (VALDGNVAG) is present in each of FBR2-FBR5. YFYQVALDG is the epitope with high affinity of binding to both DRB1*0101 and DRB1*0401.

The truncated CWSAP analyzed in this study for identification of protective epitopes belong to *S. pneumoniae* serotype 19F (Taiwan 19F-14 strain). Twenty one *S. pneumoniae* serotypes including 19F, 14, 6B, 1, 23F, 5, 6A, 19A, 9V, 18C, 2, 4, 7F, 12F, 3, 12A, 8, 46, 15B, and 45 are the most common pneumococcal serotypes causing invasive pneumococcal disease in children less than 5 years globally [2].

The CWSAP sequences were available for all of these pneumococcal serotypes in NCBI except for *S. pneumoniae* serotype 12A. We analyzed the conservancy of the most probable protective epitopes identified in the serotype 19F truncated CWSAP including YGNKDGKAD, LVATKNVDI and VALDGNVAG in these 19 CWSAPs using SIM alignment tool. All of the three epitopes had a 100% amino acid sequence match in the CWSAPs. However, VALDGNVAG had an 89% sequence match in the serotype 15B CWSAP and this epitope was converted to LALDGNVAG in these CWSAP. These results indicated that the identified epitopes are highly conserved in the most common occurring pneumococcal serotypes in the world and the immunological responses elicited against them can be protective against these pneumococcal serotypes.

In our country, the pneumococcal vaccination has not been started yet and considering the available studies in our country, the pneumococcal serotype 19F showed a high occurring frequency among *S. pneumoniae* serotypes isolated from the nasopharynx of children less than 2 years [17]. Nasopharyngeal colonization is a perquisite for the invasive pneumococcal disease [18]. Therefore, our previous study [6] and the present study were designed based on the analysis of the surface proteins of this pneumococcal serotype for identification of appropriate proteinaceous vaccine candidates. However, the pneumococcal serotype 19F may not occur at high frequencies in other regions of the world e.g. due to the vaccine pressure. The results of the present study demonstrated the applicability of the serotype 19F truncated CWSAP for eliciting protection against several pneumococcal serotypes. In addition, the systematic methodology developed in our previous paper and this work for identification of the proteinaceous vaccine candidates against pneumococci can be applied to *S. pneumoniae* serotypes other than 19F and it may lead to the identification of other protein candidates.

Table 9. Immunoprotective T-cell epitopes of truncated CWSAP with high affinity binding

MHC allele	Sequence[1]	IC50 (nM)	VaxiJen score	Position[2]
DRB1*0101	IDQLRANGT	5.4	0.9481	210
	YFYQVALDG	7.4	1.0491	343, 495, 647, 799
	INGLVAKET	32.8	0.4828	250
	FTAGVNQVI	28.8	0.8712	137
	IDQFRANGT	11.6	0.5436	362, 514, 666, 818
	LTNLVATKN	38.7	0.4030	237
	LVATKNVDI	24.4	1.6487	240
	NMAASANKM	35	0.5837	911
	FTAGVNHVI	29.1	0.8124	287,439,591,743
	QVALDGNVA	18.3	1.1309	346, 498, 650, 802
DRB1*0401	*YFYQVALDG*	29.6	1.0491	343, 495, 647, 799
	YFYEVDLNG	26	1.7635	191
	VALDGNVAG	43.6	2.0973	347, 499, 651, 803

1: The common epitope of MHC alleles is shown in italics. The most immunoprotective epitopes are shown in bold.
2: The amino acid residue number at the epitope beginning is indicated.

4. Conclusions

Among three proteins including the mature DDCP, CBPD, and the truncated CWSAP of *S. pneumoniae* serotype 19F, the truncated CWSAP possesses the greatest number of protective epitopes. In addition, the most probable immunoprotective B- and T-cell epitopes reside in this protein and these epitopes are highly conserved in CWSAPs of the most common pneumococcal serotypes in the world. Therefore, the truncated CWSAP is an appropriate candidate for development of protein based vaccines against the most common pneumococcal serotypes. Currently, the recombinant production of this protein is under investigation in our lab. In future, the protein ability in eliciting protection against pneumococci will be assessed experimentally.

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] Johnson, H. L., Deloria-Knoll, M., Levine, O. 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.

[3] WHO, "Pneumococcal vaccines, WHO position paper." *Wkly Epidemiol Rec*, 87 (14), 129-144, Apr. 2012.

[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] Tarahomjoo, S. "Bioinformatic analysis of surface proteins of *Streptococcus pneumoniae* serotype 19F for identification of vaccine candidates," *Am J Microbiol Research*, 2 (6), 174-177, Oct. 2014.

[7] Tarahomjoo, S. "Recent approaches in vaccine development against *Streptococcus pneumonia*e," *J Mol Microbiol Biotechnol*, 24 (4), 215-227, August 2014.

[8] Kelly, D. F. and Rappuoli, R. *Hot topics in infection and immunity in children II*, Springer, New York, 2005, 217-223.

[9] Doytchinova, I. A. and Flower, D. R. "Bioinformatic approach for identifying parasite and fungal candidate subunit vaccines," *Open Vaccine J*, 1 (1), 22-26, Sep. 2008.

[10] Novotny, J., Handschumacher, M., Haber, E. et al. "Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains), *Proc Natl Acad Sci USA*, 83 (2), 226-230, Jan. 1986.

[11] Quijada, L., Requema M. J., Soto, M. et al. "Mapping of the linear antigenic determinants of the *Leishmania infantum* Hsp 70 recognized by leishmaniasis sera," *Immunol Lett*, 52 (2-3), 73-79, Sep. 1996.

[12] Faria, A. R., Costa, M. M., Giusta, M. S. et al. "High throughput analysis of synthetic peptides for the immunodiagnosis of canine visceral leishmaniasis," *PLos Negl Trop dis*, 5 (9): e1310, Sep. 2011.

[13] Abdullah, M. R., Gutierrez-Fernandez, J., Pribyl, T. et al. "Structure of the pneumococcal L, D carboxypeptidase DacB and pathophysiological effects of disabled cell wall hydrolases DacA and DacB," *Mol Microbiol*, 93 (6), 1183-1206, Sep. 2014.

[14] Barh, D., Barve, N., Gupta, K. et al. "Exoproteome and secretome derived broad spectrum novel drug and vaccine candidates in *Vibrio cholerae* targeted by *Piper betel* derived compounds," *PLos ONE*, 8(1): e52773, Jan. 2013.

[15] Gosink, K. K., Mann, E. R., Guglielmo, C. et al. "Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*," *Infect Immun*, 68 (10), 5690-5695, Oct. 2000.

[16] Henderson, B., Nair, S., Pallas, J. and Williams, M. A. "Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin binding proteins," *FEMS Microbiol Rev*, 35 (1), 147-200, Jan. 2011.

[17] Tabatabaei, S. R., Fallah, F., Shiva, F. et al, "Multiplex PCR assay for detection of pneumococcal serotypes in nasopharyngeal samples of healthy children; Tehran, 2009-2010," *Ann Res Rev Biol*, 4(24), 3780-3790, Jul. 2014.

[18] Gray, B. M., Converse, G. M. 3rd, Dillon, H. C. Jr. "Epidemiological studies of *Streptococcus pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life," J Infect Dis, 142 (6), 923-933, Dec. 1980.