

(19)



(11) Publication number:

**SG 177402 A1**

(43) Publication date:

**28.02.2012**

(51) Int. Cl:

**A61K 38/02, A61K 38/16, A61K**

**39/09, A61K 31/04;**

(12)

## Patent Application

(21) Application number: **2011096898**

(71) Applicant:

**GENOCEA BIOSCIENCES, INC. 161  
FIRST STREET, SUITE 2C CAMBRIDGE,  
MA 02142 MC US  
CHILDREN'S MEDICAL CENTER  
CORPORATION 55 SHATTUCK STREET  
BOSTON, MA 02115 MC US**

(22) Date of filing: **29.06.2010**

**US 61/221,541 29.06.2009**

(30) Priority: **US 61/240,598 08.09.2009**

**US 61/240,616 08.09.2009**

**US 61/316,267 22.03.2010**

(72) Inventor:

**GIERAHL, TODD 80 FRANCIS STREET,  
APARTMENT 1 BROOKLINE, MA 02446  
US  
MALLEY, RICHARD 15 BAYVIEW  
AVENUE BEVERLY, MA 01915-4719 US**

(54) **Title:**

**VACCINES AND COMPOSITIONS AGAINST  
STREPTOCOCCUS PNEUMONIAE**

(57) **Abstract:**

Streptococcus pneumoniae is a major health concern, especially in very young, elderly, or immunocompromised patients. The present disclosure provides, *inter alia*, certain highly effective vaccines and pharmaceutical compositions in Streptococcus pneumoniae. The antigens may be used therapeutically or prophylactically.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 January 2011 (20.01.2011)

(10) International Publication Number  
WO 2011/008548 A1

(51) International Patent Classification:  
A61K 38/02 (2006.01) A61K 39/09 (2006.01)  
A61K 38/16 (2006.01) A61K 31/04 (2006.01)

(74) Agents: VARMA, Anita et al.; Ropes & Gray LLP, Prudential Tower, 800 Boylston Street, Boston, MA 02199-3600 (US).

(21) International Application Number:  
PCT/US2010/040406

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) International Filing Date:  
29 June 2010 (29.06.2010)

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language:  
English

(26) Publication Language:  
English

(30) Priority Data:  
61/221,541 29 June 2009 (29.06.2009) US  
61/240,598 8 September 2009 (08.09.2009) US  
61/240,616 8 September 2009 (08.09.2009) US  
61/316,267 22 March 2010 (22.03.2010) US

(71) Applicants (for all designated States except US): GENO-CEA BIOSCIENCES, INC. [US/US]; 161 First Street, Suite 2C, Cambridge, MA 02142 (US). CHILDREN'S HOSPITAL BOSTON [US/US]; 300 Longwood Avenue, Boston, MA 02115 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GIERAIHN, Todd [US/US]; 80 Francis Street, Apartment 1, Brookline, MA 02446 (US). MALLEY, Richard [US/US]; 15 Bayview Avenue, Beverly, MA 01915-4719 (US).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: VACCINES AND COMPOSITIONS AGAINST STREPTOCOCCUS PNEUMONIAE

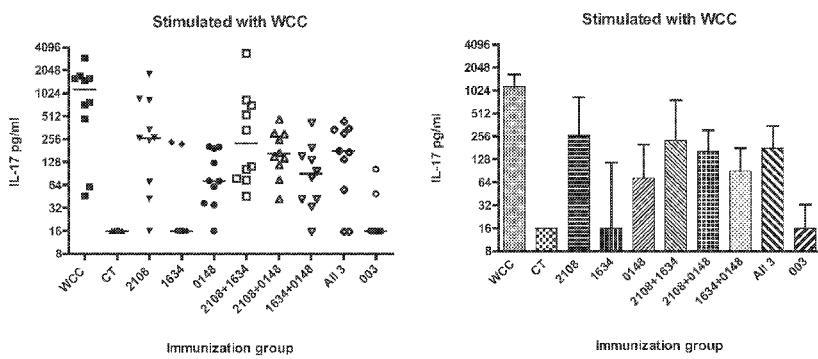


Figure 1

(57) Abstract: *Streptococcus pneumoniae* is a major health concern, especially in very young, elderly, or immunocompromised patients. The present disclosure provides, *inter alia*, certain highly effective vaccines and pharmaceutical compositions in *Streptococcus pneumoniae*. The antigens may be used therapeutically or prophylactically.

## ***Vaccines and Compositions Against *Streptococcus Pneumoniae****

### **Related Applications**

This application claims the benefit of the filing date of U.S. Provisional Application No. 61/221,541, filed June 29, 2009, U.S. Provisional Application No. 5 61/240,616, filed September 8, 2009, U.S. Provisional Application No. 61/240,598, filed September 8, 2009, and U.S. Provisional Application No. 61/316,267, filed March 22, 2010. The entire teachings of the referenced applications are expressly incorporated herein by reference.

### **I. Background**

10 Pneumococcal disease continues to be a leading cause of sickness and death in the United States and throughout the world. Each year, millions of cases of pneumonia, meningitis, bacteremia, and otitis media are attributed to infection with the pathogen *Streptococcus pneumoniae*. *S. pneumoniae* is a Gram-positive encapsulated coccus that colonizes the nasopharynx in about 5-10% of healthy 15 adults and 20-40% of healthy children. Normal colonization becomes infectious when *S. pneumoniae* is carried into the Eustachian tubes, nasal sinuses, lungs, bloodstream, meninges, joint spaces, bones and peritoneal cavity. *S. pneumoniae* has several virulence factors that enable the organism to evade the immune system. Examples include a polysaccharide capsule that prevents phagocytosis by host 20 immune cells, proteases that inhibit complement-mediated opsonization, and proteins that cause lysis of host cells. In the polysaccharide capsule, the presence of complex polysaccharides forms the basis for dividing pneumococci into different serotypes. To date, 93 serotypes of *S. pneumoniae* have been identified.

25 Various pharmaceutical compositions have been used to harness an immune response against infection by *S. pneumoniae*. A polyvalent pneumococcal vaccine, PPV-23, was developed for preventing pneumonia and other invasive diseases due to *S. pneumoniae* in the adult and aging populations. The vaccine contains capsular polysaccharides (CPs) from 23 serotypes of *S. pneumoniae*. As T cell independent antigens, these CPs induce only short-lived antibody responses, necessitating 30 repeated doses, which increases the risk of immunological tolerance. The antibodies

raised against *S. pneumoniae*, termed anticapsular antibodies, are recognized as protective in adult and immunocompetent individuals. However, children under 2 years of age and immunocompromised individuals, including the elderly, do not respond well to T-cell independent antigens and, therefore, are not afforded optimal protection by PPV-23. A second *S. pneumoniae* vaccine, Prevnar, includes bacterial polysaccharides from 7 *S. pneumoniae* strains conjugated to the diphtheria toxoid protein. This vaccine induces both B and T cell responses. However, because it only protects against 7 pneumococcal serotypes, serotype replacement can render Prevnar ineffective. PPV-23 suffers from the same limitation. Serotype replacement has already been demonstrated in several clinical trials and epidemiologic studies, and raises the possibility that different formulations of the vaccines will need to be developed, presumably at even higher cost. Furthermore, both PPV-23 and Prevnar are expensive to manufacture, greatly limiting their availability in the developing world.

15        Thus, there remains a need to design more effective pharmaceutical compositions than the current strategies offer. In particular, such compositions need to incorporate novel or specific antigens that elicit an immune response against *S. pneumoniae*.

## 20        II.      Summary

15        *Streptococcus pneumoniae* is a major health concern, especially in very young, elderly, or immunocompromised patients. While DNA and protein sequence information for *S. pneumoniae* has been known for some time, and researchers have long attempted to produce vaccines against *S. pneumoniae*, a major problem was 25 how to identify protective polypeptides from among the approximately 2100 genes in the *S. pneumoniae* genome. The instant application presents the results of whole-genome screens designed to identify the most immunogenic proteins in the *S. pneumoniae* genome. Several of the hits from the screen have been shown to protect against *S. pneumoniae* colonization in a mouse model. Accordingly, the present 30 disclosure provides, *inter alia*, certain highly effective vaccines in *Streptococcus pneumoniae*. The vaccines may be used therapeutically or prophylactically. The

present disclosure also provides specific antigens and methods for using the antigens to elicit an immune response against *S. pneumoniae*.

The present disclosure provides, for example, a vaccine formulation comprising a pharmaceutically acceptable carrier and one or more polypeptides 5 having an amino acid sequence comprising any of SEQ ID NOS: 1-11 or an immunogenic fragment thereof, and optionally further comprising a polypeptide having an amino acid sequence comprising either of SEQ ID NOS: 12 or 13 or an immunogenic fragment thereof.

The present disclosure also provides a vaccine formulation comprising a 10 pharmaceutically acceptable carrier and a polypeptide having an amino acid sequence consisting of SEQ ID NO: 11 or an immunogenic fragment thereof. In addition, the present disclosure provides a vaccine formulation comprising a pharmaceutically acceptable carrier and a polypeptide having an amino acid sequence comprising SEQ ID NO: 12.

15 Furthermore, the instant application provides a vaccine formulation comprising a pharmaceutically acceptable carrier and one or more polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 14-21 or an immunogenic fragment thereof.

This application provides, *inter alia*, a method for treating a subject suffering 20 from or susceptible to *S. pneumoniae* infection, comprising administering an effective amount of any of the vaccine formulations described herein.

The present disclosure further provides an immunogenic composition comprising a pharmaceutically acceptable carrier and two or more polypeptides having amino acid sequences comprising any of SEQ ID NOS: 1-13 or an 25 immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising one of SEQ ID NOS: 1-10 or an immunogenic fragment thereof.

### III. Brief Description of the Drawings

FIG. 1 shows the concentration of IL-17 generated by blood samples from mice that were immunized with the indicated protein(s) and cholera toxin, then stimulated with killed, unencapsulated whole cell *S. pneumoniae*, as described in Example 5. The left panel shows the data in scatter format, and the right panel shows the average and standard deviation for each sample. Immunization group “All 3” represents animals immunized with a combination of SP2108, SP0148, and SP1634.

FIG. 2 shows the concentration of IL-17 generated by blood samples from mice that were immunized with the indicated protein(s) and cholera toxin, then stimulated with a combination of three proteins (SP2108, SP0148, and SP1634), as described in Example 5.

FIG. 3 shows the number of *S. pneumoniae* colonies obtained from a nasal wash in mice that were immunized with the indicated protein(s) and cholera toxin, then challenged with intranasal administration of *S. pneumoniae*, as described in Example 5. 003 represents a control irrelevant antigen.

FIG. 4 shows the concentration of IL-17 generated by blood samples from mice that were immunized with the indicated protein(s) and cholera toxin, then stimulated with killed, unencapsulated whole cell *S. pneumoniae*, as described in Example 6.

FIG. 5 shows the concentration of IL-17 generated by blood samples from mice that were immunized with the indicated protein(s) and cholera toxin, then stimulated by the indicated proteins and combination, as described in Example 5.

FIG. 6 shows the number of *S. pneumoniae* colonies obtained from a nasal wash in mice that were immunized with the indicated protein(s) and cholera toxin, then challenged with intranasal administration of *S. pneumoniae*, as described in Example 6. The HSV-2 protein ICP47 with the gene name US12 (NP\_044543.1, NC\_001798.1; shown in the figure as 003) and ovalbumin (OVA) represent control antigens.

FIG. 7 shows the number of *S. pneumoniae* colonies obtained from a nasal wash in mice that were immunized with the indicated protein(s) and cholera toxin, then challenged with intranasal administration of *S. pneumoniae*, as described in Example 7.

5 FIG. 8 shows the number of *S. pneumoniae* colonies obtained from a nasal wash in BALB/c mice that were immunized with the indicated protein(s) and cholera toxin, then challenged with intranasal administration of *S. pneumoniae*, as described in Example 8.

10 **IV. Detailed Description**

**A. Specific polypeptides and nucleic acids for use in *S. pneumoniae* vaccines and immunogenic compositions**

This application describes *S. pneumoniae* vaccines that include one or more of the polypeptides or genes listed in Table 1, or variants or fragments thereof as 15 described below. The vaccine may include a polypeptide that comprises a sequence of Table 1 or a variant or immunogenic fragment thereof or a polypeptide that consists of a sequence of Table 1 or a variant or immunogenic fragment thereof. The DNA and protein sequence of each gene and polypeptide may be found by searching for the Locus Tag in the publicly available database, Entrez Gene (on the 20 NCBI NIH web site on the World Wide Web, at [www.ncbi.nlm.nih.gov/sites/entrez?db=gene](http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene)), in the *Streptococcus pneumoniae* TIGR4 genome, and the indicated sequences are also included in this application.

**Table 1. Immunogenic polypeptides for vaccine formulations**

<b>Locus tag name and description</b>	<b>Protein SEQ ID No.</b>	<b>DNA SEQ ID No.</b>	<b>DNA GenBank Accession No. (from March 30, 2010)</b>
SP0024	1	-	NC_003028.3 :27381-27878
SP0882	2	-	NC_003028.3 :831804-832628
SP0882N	3	24	-
SP0882 with exogenous leader	4	25	-

Locus tag name and description	Protein SEQ ID No.	DNA SEQ ID No.	DNA GenBank Accession No. (from March 30, 2010)
SP0882N with exogenous leader	5	26	-
SP0148 lacking signal sequence	6	27	-
SP0148 including signal sequence	7	28	
SP1072	8	-	NC_003028.3l:1008420-1010180
SP2108 including signal sequence	9	-	NC_003028.3l:2020750-2022021
SP2108 lacking signal sequence	10	29	-
SP0641M	11	30	-
SP0641	12	-	NC_003028.3l:2020750-2022021
SP0641N	13	31	-
SP0882 consensus	14	-	-
SP0882N consensus	15	-	-
SP0882 consensus with exogenous leader	16	-	-
SP0882N consensus with exogenous leader	17	-	-
SP0148 consensus lacking signal sequence	18	-	-
SP0148 consensus including signal sequence	19	-	-
SP2108 consensus lacking signal sequence	20	-	-
SP2108 consensus including signal sequence	21	-	-
SP1634	22	-	NC_003028.3l:1534348-1535421
SP0314	23	-	NC_003028.3l:287483-290683

Certain polypeptides of Table 1, and variants thereof, are described in greater detail below.

### 1. SP0024 (SEQ ID NO: 1) and variants thereof

5 SP0024 represents a hypothetical protein of 165 amino acids, containing a conserved carbonic anhydrase domain that extends from amino acid 27 to amino acid 163. Based on this consensus motif, SP0024 may be a zinc-binding protein.

In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP0024. The polypeptide may also be a variant of the at least 20 residue fragment. In certain embodiments, the 5 polypeptide includes no more than 150, 125, or 100 consecutive amino acids from SP0024.

## **2. SP0882 (SEQ ID NO: 2) and variants thereof**

SP0882 is a conserved hypothetical protein of 274 amino acids. Much of the protein (amino acids 2-270) forms an esterase or lipase-like region.

10 In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP0882. The polypeptide may also be a variant of the at least 20 residue fragment. In certain embodiments, the polypeptide includes no more than 250, 275, 200, 175, 150, 125, or 100 consecutive 15 amino acids from SP0882.

One particular truncation variant named SP0882N consists of the N-terminal 130 amino acids of SP0882, and is shown as SEQ ID NO: 3. SP0882N includes a region that is particularly well conserved among different serotypes. In certain 20 embodiments, a polypeptide comprising SP0882 or SP0882N, or an immunogenic fragment of either, also comprises an exogenous leader sequence. The leader sequence may be, for example, the leader sequence of SP2108. Two exemplary such polypeptides are SEQ ID NOS: 4 and 5.

25 Variants of DNA and protein sequences of SP0882 are described, *inter alia*, in US Patent Application Publication No. 2009/0215149 and International Applications WO2002/077021, WO98/18931, and WO2007/106407. A variant of SP0882N is disclosed in International Application WO2008/146164.

Sequence variation occurs at the protein level between different *S. pneumoniae* serotypes, and consensus sequences illustrating combinations of 30 SP0882 sequences from different *S. pneumoniae* serotypes are provided as SEQ ID NOS: 14-17. Accordingly, in certain embodiments, the vaccine formulation

comprises a polypeptide having an amino acid sequence comprising, or consisting of, any of SEQ ID NOS: 14-17, or an immunogenic fragment thereof (e.g., in place of a polypeptide having an amino acid sequence comprising one of SEQ ID NOS: 2-5).

5 Nucleic acid sequences encoding different variants of SP0882 are provided as SEQ ID NOS: 24-26, although due to degeneracy in the genetic code, other DNA sequences (including codon-optimized sequences) could encode these polypeptides.

### 3. SP0148 (SEQ ID NO: 7) and variants thereof

The protein SP0148 is named “ABC transporter, substrate-binding protein”.

10 Proteins of this class are typically extracellular proteins that interact transiently with a transmembrane protein complex. Such complexes use energy generated by ATP hydrolysis to translocate specific substrates across a cell membrane. SP0148 is a 276 amino acid protein that contains a conserved PBP<sub>b</sub> (periplasmic binding protein) domain, spanning amino acids 40-246, which is typical of membrane-bound

15 transport complexes. In addition, SB0148 has a bacterial extracellular solute-binding proteins, family 3 domain which is largely co-extensive with the PBP<sub>b</sub> domain and extends from amino acid 40 to 244. In some embodiments, a vaccine or other composition comprises a truncation mutant of SP0148 comprising or lacking one or more of said domains and motifs.

20 In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP0148. The polypeptide may also be a variant of the at least 20 residue fragment. In certain embodiments, the polypeptide includes no more than 250, 275, 200, 175, 150, 125, or 100 consecutive

25 amino acids from SP0148.

Endogenous SP0148 comprises a putative signal sequence that may direct its secretion. In some embodiments, a variant of SP0148 that lacks the signal sequence (SEQ ID NO: 6) is used. The polypeptide of SEQ ID NO: 6 is encoded by the nucleic acid of SEQ ID NO: 27, although other nucleic acid sequences (including

30 codon-optimized sequences) may be used. SEQ ID NO: 28 encodes the full length sequence of SP0148 used in the screens herein.

Variants of the amino acid sequence and nucleotide sequence of SP0148 may be found in U.S. Patent Application Publication No. 2005/0020813, U.S. Patent Nos. 7,378,514 and 7,504,110, and European Patent Application No. EP1572868 and EP1855717.

5 Consensus sequences illustrating combinations of SP0148 sequences from different *S. pneumoniae* serotypes are provided as SEQ ID NOS: 18 and 19. Accordingly, in certain embodiments, the vaccine formulation comprises a polypeptide having an amino acid sequence comprising, or consisting of, either of SEQ ID NOS: 18-19, or an immunogenic fragment thereof (e.g., in place of a 10 polypeptide having an amino acid sequence comprising one of SEQ ID NOS: 6 or 7).

#### **4. SP1072 (SEQ ID NO: 8) and variants thereof**

SP1072, also known as dnaG, is a DNA primase enzyme that catalyzes formation of an RNA primer which allows DNA polymerase to initiate DNA 15 replication. A protein of 586 amino acids, SP1072 contains several conserved motifs. Beginning at the N-terminus, amino acids 2 – 96 form a zinc finger domain, the DNA primase catalytic core spans amino acids 122 – 250, and a highly conserved topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain region extends from amino acid 258 to 330. In some embodiments, a 20 vaccine or other composition comprises a truncation mutant of SP1072 comprising or lacking one or more of said domains and motifs.

In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP1072. The polypeptide may also 25 be a variant of the at least 20 residue fragment. In certain embodiments, the polypeptide includes no more than 550, 500, 450, 400, 350, 300, 250, 200, 150, or 100 consecutive amino acids from SP1072.

#### **5. SP2108 (SEQ ID NO: 9) and variants thereof**

The polypeptide SP2108 is 423 amino acids in length and is alternatively 30 known as MalX, maltose/maltodextrin ABC transporter, or maltose/maltodextrin-

binding protein. Much of the protein (amino acids 3-423) is classified as a MalE (Maltose-binding periplasmic) domain. In addition, SP2108 contains a signal sequence that directs its secretion. In some embodiments, a vaccine or other composition comprises a truncation mutant of SP2108 comprising one or more of 5 said domains and motifs.

In some embodiments, the compositions and methods herein call for the use of an SP2108 variant that lacks the signal sequence. This variant is represented by polypeptide sequence SEQ ID NO: 10 and may be encoded by, for example, a nucleic acid according to SEQ ID NO: 29.

10 In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP2108. The polypeptide may also be a variant of the at least 20 residue fragment. In certain embodiments, the polypeptide includes no more than 400, 350, 300, 250, 200, 150, or 100 consecutive 15 amino acids from SP2108.

Consensus sequences illustrating combinations of SP2108 sequences from different serotypes are provided as SEQ ID NOS: 20 and 21. Thus, in certain embodiments, the vaccine formulation comprises a polypeptide having an amino acid sequence comprising, or consisting of, either of SEQ ID NOS: 20-21, or an 20 immunogenic fragment thereof (e.g., in place of a polypeptide having an amino acid sequence comprising one of SEQ ID NOS: 9 or 10).

## 6. SP0641 (SEQ ID NO: 12) and variants thereof

At 2144 amino acids in length, SP0641 is also known as PrtA, a cell wall-associated serine protease. Full-length SP0641 contains a number of conserved 25 motifs: the PA\_2 motif, extending between amino acids 485 and 597, which may form a protein binding surface; the Fn3-like domain (amino acids 800 – 939); and two predicted catalytic domains of the S8 C5a type located at amino acids 226 – 449 and 639 – 777. In some embodiments, a vaccine or other composition comprises a truncation mutant of SP0641 comprising or lacking one or more of said domains and 30 motifs.

In some embodiments, vaccines or pharmaceutical compositions comprising an *S. pneumoniae* polypeptide include a polypeptide containing at least 20 consecutive amino acid residues selected from SP0641. The polypeptide may also be a variant of the at least 20 residue fragment. In certain embodiments, the 5 polypeptide includes no more than 1000, 900, 800, 700, 600, 500, 400, 300, 200, or 100 consecutive amino acids from SP0641.

Certain other truncation mutants of SP0641 may also be used. For instance, the polypeptide designated SP0641N (SEQ ID NO: 13) consists of 661 amino acids corresponding to amino acids 24-684 near the N-terminus of SP0641. Roughly 10 adjacent to SP0641N (and corresponding to amino acids 686-1333 of SP0641) lies the 648 residue region captured by the truncation variant SP0641M (SEQ ID NO: 11).

Variants of SP0641 are disclosed in, for example, U.S. Patents No. 7,338,786, 6,573,082, and 7,132,107, as well as International Application 15 WO00/06738.

SEQ ID NOS: 30 and 31 display the DNA sequences of SP0641M and SP0641N, respectively, although due to degeneracy in the genetic code, other DNA sequences (including codon-optimized sequences) could encode SP0641.

20 Polypeptides homologous to the polypeptides of Tables 1 and 2 (for example, SP0024, 0882, 0882N, 0148 with or without a signal sequence, 1072, SP1028 with or without a signal sequence, SP0641, SP0641M, or SP0641N) may also be used in the compositions and methods disclosed herein. Individual strains of *S. pneumoniae* contain numerous mutations relative to each other, and some of these 25 result different protein sequences between the different strains. One of skill in the art may readily substitute an amino acid sequence, or a portion thereof, with the homologous amino acid sequence from a different *S. pneumoniae* strain. In certain aspects, this application provides immunogenic polypeptides with at least 90%, 95%, 97%, 98%, 99%, or 99.5% identity to the polypeptides of Tables 1 and 2 or an 30 immunogenic fragment thereof. Serotypic variation may be used to design such variants of the polypeptides of Tables 1 and 2.

In some embodiments, the vaccine compositions herein comprise a fragment of a protein of Table 1 or 2 (for example, fragments of SP0024, SP0882, SP0882N, OSP148 with or without a signal sequence, SP1072, SP1028 with or without a signal sequence, SP0641, SP0641M, or SP0641N). In some embodiments, this application 5 provides truncation mutants that are close in size to the polypeptide of Table 1 or 2 (for example, one of SEQ ID NOS: 1-13). For example, they may lack at most one, two three, four, five, ten, or twenty amino acids from one or both termini. Internal deletions, e.g., of 1-10, 11-20, 21-30, or 31-40 amino acids, are also contemplated.

In certain embodiments the vaccine formulation comprises one or more 10 polypeptides having an amino acid sequence comprising, or consisting of, any of SEQ ID NOS: 14-21. In certain embodiments, the fragment is a truncated fragment of any of SEQ ID NOS: 14-21 wherein from 1-5, 1-10, or 1-20 amino acid residues are removed from the N-terminus, C-terminus, or both. In certain embodiments, the fragment is a truncated fragment of any of SEQ ID NOS: 14-21 wherein from 1-10 15 amino acid residues are removed from the N-terminus, C-terminus, or both. For instance, 10 amino acid residues may be removed from each of the N-terminus and C-terminus resulting in a protein with 20 amino acid residues removed.

In addition to those nucleic acids and polypeptides described in Table 1 above, this application also provides immunogenic compositions that include one or 20 more of the polypeptides or genes listed in Table 1 and/or Table 2, or variants or fragments thereof as described herein. The DNA and protein sequence of each gene and protein may be found by searching for the Locus Tag in the publicly available database, Entrez Gene, as described above.

25 **Table 2. Immunogenic proteins identified in human and mouse screens**

<b>Locus tag name</b>	<b>Protein accession number</b>	<b>DNA accession number (from March 30, 2010)</b>
SP1574	AAK75660.1	NC_003028.3 c1481367-1480609
SP1655	AAK75734.1	NC_003028.3 c1557922-1557230
SP2106	AAK76165.1	NC_003028.3 c2018657-2016399
SP1473	AAK75567.1	NC_003028.3 c1386534-1386277
SP0605	AAK74757.1	NC_003028.3 571604-572485
SP1177	AAK75286.1	NC_003028.3 c1115580-1115317
SP0335	AAK74510.1	NC_003028.3 306559-306876

<b>Locus tag name</b>	<b>Protein accession number</b>	<b>DNA accession number (from March 30, 2010)</b>
SP0906	AAK75031.1	NC_003028.3 :c859160-859029
SP1828	AAK75901.1	NC_003028.3 :c1740010-1739000
SP2157	AAK76211.1	NC_003028.3 :c2072146-2070995
SP1229	AAK75335.1	NC_003028.3 :c1163388-1161718
SP1128	AAK75238.1	NC_003028.3 :1061773-1063077
SP1836	AAK75909.1	NC_003028.3 :1746104-1746280
SP1865	AAK75937.1	NC_003028.3 :c1772987-1771923
SP0904	AAK75029.1	NC_003028.3 :c858126-857311
SP0765	AAK74903.1	NC_003028.3 :724170-725207
SP1634	AAK75714.1	NC_003028.3 :1534348-1535421
SP0418	AAK74581.1	NC_003028.3 :396692-396916
SP1923	AAK75991.1	NC_003028.3 :c1833311-1831896
SP1313	AAK75991.1	NC_003028.3 :c1833311-1831896
SP0775	AAK74913.1	NC_003028.3 :731798-732070
SP0314	AAK74491.1	NC_003028.3 :287483-290683
SP0912	AAK75037.1	NC_003028.3 :864707-865465
SP0159	AAK74341.1	NC_003028.3 :c157554-156292
SP0910	AAK75035.1	NC_003028.3 :863462-863734
SP2148	AAK76205.1	NC_003028.3 :2062144-2063373
SP1412	AAK75510.1	NC_003028.3 :c1332393-1331605
SP0372	AAK74539.1	NC_003028.3 :350268-350597
SP1304	AAK75407.1	NC_003028.3 :c1232491-1232390
SP2002	AAK76069.1	NC_003028.3 :c1906183-1905446
SP0612	AAK74764.1	NC_003028.3 :579708-579806
SP1988	AAK76055.1	NC_003028.3 :c1892598-1890565
SP0484	AAK74643.1	NC_003028.3 :465572-466402
SP0847	AAK74978.1	NC_003028.3 :794144-795202
SP1527	AAK75616.1	NC_003028.3 :c1439494-1437536
SP0542	AAK74699.1	NC_003028.3 :515940-516059
SP0441	AAK74602.1	NC_003028.3 :414869-415057
SP0350	AAK74523.1	NC_003028.3 :323990-324625
SP0014	AAK74207.1	NC_003028.3 :14450-14929
SP1965	AAK76032.1	NC_003028.3 :c1873279-1873073
SP0117	AAK74303.1	NC_003028.3 :118423-120657
SP0981	AAK75102.1	NC_003028.3 :927115-928056
SP2229	AAK76277.1	NC_003028.3 :c2148627-2147602
SP2136	AAK76194.1	NC_003028.3 :c2048521-2046656
SP1179	AAK75288.1	NC_003028.3 :1116230-1118389
SP1174	AAK75283.1	NC_003028.3 :c1110717-1108258
SP2216	AAK76264.1	NC_003028.3 :c2136445-2135267
SP1393	AAK75491.1	NC_003028.3 :1316756-1318027
SP0641.1	Amino acids 28 - 1006 of AAK74791.1	Nucleotides 603976-606910 of NC_003028.3

<b>Locus tag name</b>	<b>Protein accession number</b>	<b>DNA accession number (from March 30, 2010)</b>
	(which is full-length SP0641)	
SP1384	AAK75482.1	NC_003028.3l:c1309464-1308967
SP2032	AAK76097.1	NC_003028.3l:c1939994-1938321

Typically, the polypeptides present in compounds of the invention are immunogenic, either alone or as a variant, which includes polypeptides fused to another polypeptide or mixed with or complexed to an adjuvant. Variants also 5 include sequences with less than 100% sequence identity, as described herein. In certain embodiments, an antigen of Table 1 or 2 is provided as a full length polypeptide. In addition, one may use fragments, precursors and analogs that have an appropriate immunogenicity.

These polypeptides may be immunogenic in mammals, for example mice, 10 guinea pigs, or humans. An immunogenic polypeptide is typically one capable of raising a significant immune response in an assay or in a subject. For instance, an immunogenic polypeptide may increase the amount of IL-17 produced by T cells. The IL-17 assay described in Examples 1-4 is an example of an assay that may be used to identify an immunogenic polypeptide. Alternatively, an immunogenic 15 polypeptide may (i) induce production of antibodies, e.g., neutralizing antibodies, that bind to the polypeptide (ii) induce T<sub>H</sub>1 immunity, (iii) activate the CD8+ CTL response, for example by increasing CD8+ T cells and/or increasing localization of CD8+ T cells to the site of infection or reinfection, (iv) induce T<sub>H</sub>17 immunity, and/or (v) activate innate immunity. In some embodiments, an immunogenic 20 polypeptide causes the production of a detectable amount of antibody specific to that antigen.

In certain embodiments, polypeptides have less than 20%, 30%, 40%, 50%, 60% or 70% identity to human autoantigens and/or gut commensal bacteria (e.g., 25 certain *Bacteroides*, *Clostridium*, *Fusobacterium*, *Eubacterium*, *Ruminococcus*, *Peptococcus*, *Peptostreptococcus*, *Bifidobacterium*, *Escherichia* and *Lactobacillus* species). Examples of human autoantigens include insulin, proliferating cell nuclear antigen, cytochrome P450, and myelin basic protein.

The present disclosure provides, for example, a vaccine formulation comprising a pharmaceutically acceptable carrier and one or more polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-11 or an immunogenic fragment thereof, and optionally further comprising a polypeptide 5 having an amino acid sequence comprising either of SEQ ID NOS: 12 or 13 or an immunogenic fragment thereof. In certain embodiments, the vaccine formulation comprises at least two different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13 or an immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising one of SEQ 10 ID NOS: 1-10 or an immunogenic fragment thereof. Here, the term “different” signifies that each of said two peptides originates from a different sequence selected from SEQ ID NOS: 1-13.

The vaccine formulation may also comprise one or more polypeptides having an amino acid sequence consisting of any of SEQ ID NOS: 1-11.

15 In some embodiments, the vaccine formulation comprises at least two polypeptides, each polypeptide belonging to a different group of (i)-(vi): (i) SEQ ID NO: 1 or an immunogenic fragment thereof, (ii) one of SEQ ID NOS: 2-5 or an immunogenic fragment thereof, (iii) one of SEQ ID NOS: 6-7 or an immunogenic fragment thereof, (iv) SEQ ID NO: 8 or an immunogenic fragment thereof, (v) one 20 of SEQ ID NOS: 9-10 or an immunogenic fragment thereof, and (vi) one of SEQ ID NO: 11-13 or an immunogenic fragment thereof. Examples of such combinations are listed below:

SEQ ID NO: 1 and SEQ ID NO: 2  
SEQ ID NO: 1 and SEQ ID NO: 3  
25 SEQ ID NO: 1 and SEQ ID NO: 4  
SEQ ID NO: 1 and SEQ ID NO: 5  
SEQ ID NO: 1 and SEQ ID NO: 6  
SEQ ID NO: 1 and SEQ ID NO: 7  
SEQ ID NO: 1 and SEQ ID NO: 8  
30 SEQ ID NO: 1 and SEQ ID NO: 9  
SEQ ID NO: 1 and SEQ ID NO: 10  
SEQ ID NO: 1 and SEQ ID NO: 11  
SEQ ID NO: 1 and SEQ ID NO: 12  
SEQ ID NO: 1 and SEQ ID NO: 13  
35 SEQ ID NO: 2 and SEQ ID NO: 6

SEQ ID NO: 2 and SEQ ID NO: 7  
SEQ ID NO: 2 and SEQ ID NO: 8  
SEQ ID NO: 2 and SEQ ID NO: 9  
SEQ ID NO: 2 and SEQ ID NO: 10  
5 SEQ ID NO: 2 and SEQ ID NO: 11  
SEQ ID NO: 2 and SEQ ID NO: 12  
SEQ ID NO: 2 and SEQ ID NO: 13

SEQ ID NO: 3 and SEQ ID NO: 6  
10 SEQ ID NO: 3 and SEQ ID NO: 7  
SEQ ID NO: 3 and SEQ ID NO: 8  
SEQ ID NO: 3 and SEQ ID NO: 9  
SEQ ID NO: 3 and SEQ ID NO: 10  
SEQ ID NO: 3 and SEQ ID NO: 11  
15 SEQ ID NO: 3 and SEQ ID NO: 12  
SEQ ID NO: 3 and SEQ ID NO: 13

SEQ ID NO: 4 and SEQ ID NO: 6  
20 SEQ ID NO: 4 and SEQ ID NO: 7  
SEQ ID NO: 4 and SEQ ID NO: 8  
SEQ ID NO: 4 and SEQ ID NO: 9  
SEQ ID NO: 4 and SEQ ID NO: 10  
SEQ ID NO: 4 and SEQ ID NO: 11  
SEQ ID NO: 4 and SEQ ID NO: 12  
25 SEQ ID NO: 4 and SEQ ID NO: 13

SEQ ID NO: 5 and SEQ ID NO: 6  
30 SEQ ID NO: 5 and SEQ ID NO: 7  
SEQ ID NO: 5 and SEQ ID NO: 8  
SEQ ID NO: 5 and SEQ ID NO: 9  
SEQ ID NO: 5 and SEQ ID NO: 10  
SEQ ID NO: 5 and SEQ ID NO: 11  
SEQ ID NO: 5 and SEQ ID NO: 12  
35 SEQ ID NO: 5 and SEQ ID NO: 13

SEQ ID NO: 6 and SEQ ID NO: 8  
40 SEQ ID NO: 6 and SEQ ID NO: 9  
SEQ ID NO: 6 and SEQ ID NO: 10  
SEQ ID NO: 6 and SEQ ID NO: 11  
SEQ ID NO: 6 and SEQ ID NO: 12  
SEQ ID NO: 6 and SEQ ID NO: 13

SEQ ID NO: 7 and SEQ ID NO: 8  
45 SEQ ID NO: 7 and SEQ ID NO: 9  
SEQ ID NO: 7 and SEQ ID NO: 10  
SEQ ID NO: 7 and SEQ ID NO: 11  
SEQ ID NO: 7 and SEQ ID NO: 12  
SEQ ID NO: 7 and SEQ ID NO: 13

SEQ ID NO: 8 and SEQ ID NO: 9  
SEQ ID NO: 8 and SEQ ID NO: 10  
SEQ ID NO: 8 and SEQ ID NO: 11  
5 SEQ ID NO: 8 and SEQ ID NO: 12  
SEQ ID NO: 8 and SEQ ID NO: 13  
  
SEQ ID NO: 9 and SEQ ID NO: 11  
10 SEQ ID NO: 9 and SEQ ID NO: 12  
SEQ ID NO: 9 and SEQ ID NO: 13  
  
SEQ ID NO: 10 and SEQ ID NO: 11  
SEQ ID NO: 10 and SEQ ID NO: 12  
15 SEQ ID NO: 10 and SEQ ID NO: 13

In certain embodiments, the vaccine formulation comprises at least three different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13 or an immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising one of SEQ ID NOS: 1-10. In 20 certain such embodiments, the vaccine formulation comprises at least three polypeptides, each polypeptide belonging to a different group of (i)-(vi): (i) SEQ ID NO: 1 or an immunogenic fragment thereof, (ii) one of SEQ ID NOS: 2-5 or an immunogenic fragment thereof, (iii) one of SEQ ID NOS: 6-7 or an immunogenic fragment thereof, (iv) SEQ ID NO: 8 or an immunogenic fragment thereof, (v) one 25 of SEQ ID NOS: 9-10 or an immunogenic fragment thereof, and (vi) one of SEQ ID NO: 11-13 or an immunogenic fragment thereof. Examples of such combinations are listed below:

SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 6  
SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 7  
30 SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 9  
SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 11  
35 SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 13  
  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 6  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 7  
40 SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 9

SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 3, and SEQ ID NO: 13

5                   SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 6  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 7  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 9

10                  SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 4; and SEQ ID NO: 13

15                  SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 6  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 7  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 9  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 10

20                  SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 5; and SEQ ID NO: 13

25                  SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 9  
SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 6; and SEQ ID NO: 13

30                  SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 8  
SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 9  
SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 11

35                  SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 7; and SEQ ID NO: 13

40                  SEQ ID NO: 1, SEQ ID NO: 8; and SEQ ID NO: 9  
SEQ ID NO: 1, SEQ ID NO: 8; and SEQ ID NO: 10  
SEQ ID NO: 1, SEQ ID NO: 8; and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 8; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 8; and SEQ ID NO: 13

45                  SEQ ID NO: 1, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 1, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 9; and SEQ ID NO: 13

                      SEQ ID NO: 1, SEQ ID NO: 10; and SEQ ID NO: 11

SEQ ID NO: 1, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 1, SEQ ID NO: 10; and SEQ ID NO: 13

5           SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 8  
          SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 9  
          SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 10  
          SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 11  
          SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 12  
          SEQ ID NO: 2, SEQ ID NO: 6; and SEQ ID NO: 13

10           SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 8  
          SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 9  
          SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 10  
          SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 11  
15           SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 12  
          SEQ ID NO: 2, SEQ ID NO: 7; and SEQ ID NO: 13

20           SEQ ID NO: 2, SEQ ID NO: 8; and SEQ ID NO: 9  
          SEQ ID NO: 2, SEQ ID NO: 8; and SEQ ID NO: 10  
          SEQ ID NO: 2, SEQ ID NO: 8; and SEQ ID NO: 11  
          SEQ ID NO: 2, SEQ ID NO: 8; and SEQ ID NO: 12  
          SEQ ID NO: 2, SEQ ID NO: 8; and SEQ ID NO: 13

25           SEQ ID NO: 2, SEQ ID NO: 9; and SEQ ID NO: 11  
          SEQ ID NO: 2, SEQ ID NO: 9; and SEQ ID NO: 12  
          SEQ ID NO: 2, SEQ ID NO: 9; and SEQ ID NO: 13

30           SEQ ID NO: 2, SEQ ID NO: 10; and SEQ ID NO: 11  
          SEQ ID NO: 2, SEQ ID NO: 10; and SEQ ID NO: 12  
          SEQ ID NO: 2, SEQ ID NO: 10; and SEQ ID NO: 13

35           SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 8  
          SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 9  
          SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 10  
          SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 11  
          SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 12  
          SEQ ID NO: 3, SEQ ID NO: 6; and SEQ ID NO: 13

40           SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 8  
          SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 9  
          SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 10  
          SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 11  
          SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 12  
          SEQ ID NO: 3, SEQ ID NO: 7; and SEQ ID NO: 13

45           SEQ ID NO: 3, SEQ ID NO: 8; and SEQ ID NO: 9  
          SEQ ID NO: 3, SEQ ID NO: 8; and SEQ ID NO: 10  
          SEQ ID NO: 3, SEQ ID NO: 8; and SEQ ID NO: 11

SEQ ID NO: 3, SEQ ID NO: 8; and SEQ ID NO: 12  
SEQ ID NO: 3, SEQ ID NO: 8; and SEQ ID NO: 13

5           SEQ ID NO: 3, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 3, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 3, SEQ ID NO: 9; and SEQ ID NO: 13

10           SEQ ID NO: 3, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 3, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 3, SEQ ID NO: 10; and SEQ ID NO: 13

15           SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 8  
SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 9  
SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 10  
SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 11  
SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 12  
SEQ ID NO: 4, SEQ ID NO: 6; and SEQ ID NO: 13

20           SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 8  
SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 9  
SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 10  
SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 11  
SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 12  
SEQ ID NO: 4, SEQ ID NO: 7; and SEQ ID NO: 13

25           SEQ ID NO: 4, SEQ ID NO: 8; and SEQ ID NO: 9  
SEQ ID NO: 4, SEQ ID NO: 8; and SEQ ID NO: 10  
SEQ ID NO: 4, SEQ ID NO: 8; and SEQ ID NO: 11  
SEQ ID NO: 4, SEQ ID NO: 8; and SEQ ID NO: 12  
SEQ ID NO: 4, SEQ ID NO: 8; and SEQ ID NO: 13

30           SEQ ID NO: 4, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 4, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 4, SEQ ID NO: 9; and SEQ ID NO: 13

35           SEQ ID NO: 4, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 4, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 4, SEQ ID NO: 10; and SEQ ID NO: 13

40           SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 8  
SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 9  
SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 10  
SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 11  
SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 12  
SEQ ID NO: 5, SEQ ID NO: 6; and SEQ ID NO: 13

45           SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 8  
SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 9

SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 10  
SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 11  
SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 12  
SEQ ID NO: 5, SEQ ID NO: 7; and SEQ ID NO: 13  
5  
SEQ ID NO: 5, SEQ ID NO: 8; and SEQ ID NO: 9  
SEQ ID NO: 5, SEQ ID NO: 8; and SEQ ID NO: 10  
SEQ ID NO: 5, SEQ ID NO: 8; and SEQ ID NO: 11  
SEQ ID NO: 5, SEQ ID NO: 8; and SEQ ID NO: 12  
10  
SEQ ID NO: 5, SEQ ID NO: 8; and SEQ ID NO: 13  
SEQ ID NO: 5, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 5, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 5, SEQ ID NO: 9; and SEQ ID NO: 13  
15  
SEQ ID NO: 5, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 5, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 5, SEQ ID NO: 10; and SEQ ID NO: 13  
20  
SEQ ID NO: 6, SEQ ID NO: 8; and SEQ ID NO: 9  
SEQ ID NO: 6, SEQ ID NO: 8; and SEQ ID NO: 10  
SEQ ID NO: 6, SEQ ID NO: 8; and SEQ ID NO: 11  
SEQ ID NO: 6, SEQ ID NO: 8; and SEQ ID NO: 12  
SEQ ID NO: 6, SEQ ID NO: 8; and SEQ ID NO: 13  
25  
SEQ ID NO: 6, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 6, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 6, SEQ ID NO: 9; and SEQ ID NO: 13  
30  
SEQ ID NO: 6, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 6, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 6, SEQ ID NO: 10; and SEQ ID NO: 13  
35  
SEQ ID NO: 7, SEQ ID NO: 8; and SEQ ID NO: 9  
SEQ ID NO: 7, SEQ ID NO: 8; and SEQ ID NO: 10  
SEQ ID NO: 7, SEQ ID NO: 8; and SEQ ID NO: 11  
SEQ ID NO: 7, SEQ ID NO: 8; and SEQ ID NO: 12  
SEQ ID NO: 7, SEQ ID NO: 8; and SEQ ID NO: 13  
40  
SEQ ID NO: 7, SEQ ID NO: 9; and SEQ ID NO: 11  
SEQ ID NO: 7, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 7, SEQ ID NO: 9; and SEQ ID NO: 13  
45  
SEQ ID NO: 7, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 7, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 7, SEQ ID NO: 10; and SEQ ID NO: 13  
SEQ ID NO: 8, SEQ ID NO: 9; and SEQ ID NO: 11

SEQ ID NO: 8, SEQ ID NO: 9; and SEQ ID NO: 12  
SEQ ID NO: 8, SEQ ID NO: 9; and SEQ ID NO: 13

5           SEQ ID NO: 8, SEQ ID NO: 10; and SEQ ID NO: 11  
SEQ ID NO: 8, SEQ ID NO: 10; and SEQ ID NO: 12  
SEQ ID NO: 8, SEQ ID NO: 10; and SEQ ID NO: 13

In some embodiments, the vaccine formulation comprises at least two different polypeptides having an amino acid sequence comprising any of SEQ ID 10 NOS: 14-21 or an immunogenic fragment thereof. In certain such embodiments, the vaccine formulation comprises at least two polypeptides, each polypeptide belonging to a different group of (i)-(iii): (i) one of SEQ ID NOS: 14-17 or an immunogenic fragment thereof, (ii) one of SEQ ID NOS: 18-19 or an immunogenic fragment thereof; and (iii) one of SEQ ID NOS: 20-21 or an immunogenic fragment 15 thereof. Examples of such combinations are listed below:

SEQ ID NO: 14 and SEQ ID NO: 18  
SEQ ID NO: 14 and SEQ ID NO: 19  
SEQ ID NO: 14 and SEQ ID NO: 20  
SEQ ID NO: 14 and SEQ ID NO: 21  
20           SEQ ID NO: 15 and SEQ ID NO: 18  
SEQ ID NO: 15 and SEQ ID NO: 19  
SEQ ID NO: 15 and SEQ ID NO: 20  
SEQ ID NO: 15 and SEQ ID NO: 21  
25           SEQ ID NO: 16 and SEQ ID NO: 18  
SEQ ID NO: 16 and SEQ ID NO: 19  
SEQ ID NO: 16 and SEQ ID NO: 20  
SEQ ID NO: 16 and SEQ ID NO: 21  
30           SEQ ID NO: 17 and SEQ ID NO: 18  
SEQ ID NO: 17 and SEQ ID NO: 19  
SEQ ID NO: 17 and SEQ ID NO: 20  
SEQ ID NO: 17 and SEQ ID NO: 21  
35           SEQ ID NO: 18 and SEQ ID NO: 20  
SEQ ID NO: 18 and SEQ ID NO: 21  
40           SEQ ID NO: 19 and SEQ ID NO: 20  
SEQ ID NO: 19 and SEQ ID NO: 21

In some aspects, a vaccine formulation comprising one or more of SEQ ID NOS: 14-21 further comprises a polypeptide having an amino acid sequence comprising any of SEQ ID NOS: 1-13.

In certain embodiments, the vaccine formulation comprises at least three 5 different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 14-21 or an immunogenic fragment thereof. In certain such embodiments, the vaccine formulation comprises three of (i)-(iii): (i) one of SEQ ID NOS: 14-17 or an immunogenic fragment thereof, (ii) one of SEQ ID NOS: 18-19 or an immunogenic fragment thereof; and (iii) one of SEQ ID NOS: 20-21 or an immunogenic fragment 10 thereof. Examples of such combinations are listed below:

SEQ ID NO: 14, SEQ ID NO: 18, and SEQ ID NO: 20  
SEQ ID NO: 14, SEQ ID NO: 18, and SEQ ID NO: 21  
SEQ ID NO: 14, SEQ ID NO: 19, and SEQ ID NO: 20  
SEQ ID NO: 14, SEQ ID NO: 19, and SEQ ID NO: 21

15 SEQ ID NO: 15, SEQ ID NO: 18, and SEQ ID NO: 20  
SEQ ID NO: 15, SEQ ID NO: 18, and SEQ ID NO: 21  
SEQ ID NO: 15, SEQ ID NO: 19, and SEQ ID NO: 20  
SEQ ID NO: 15, SEQ ID NO: 19, and SEQ ID NO: 21

20 SEQ ID NO: 16, SEQ ID NO: 18, and SEQ ID NO: 20  
SEQ ID NO: 16, SEQ ID NO: 18, and SEQ ID NO: 21  
SEQ ID NO: 16, SEQ ID NO: 19, and SEQ ID NO: 20  
SEQ ID NO: 16, SEQ ID NO: 19, and SEQ ID NO: 21

25 SEQ ID NO: 17, SEQ ID NO: 18, and SEQ ID NO: 20  
SEQ ID NO: 17, SEQ ID NO: 18, and SEQ ID NO: 21  
SEQ ID NO: 17, SEQ ID NO: 19, and SEQ ID NO: 20  
SEQ ID NO: 17, SEQ ID NO: 19, and SEQ ID NO: 21

30 A polypeptide may comprise one or more immunogenic portions and one or more non-immunogenic portions. The immunogenic portions may be identified by various methods, including protein microarrays, ELISPOT/ELISA techniques, and / or specific assays on different deletion mutants (e.g., fragments) of the polypeptide in question. Immunogenic portions may also be identified by computer algorithms.  
35 Some such algorithms, like EpiMatrix (produced by EpiVax), use a computational matrix approach. Other computational tools for identifying antigenic epitopes include PEPVAC (Promiscuous EPitope-based VACCine, hosted by Dana Farber

Cancer Institute on the world wide web at immunax.dfcu.harvard.edu/PEPVAC), and MHCpred (which uses a partial least squares approach and is hosted by The Jenner Institute on the world wide web at www.jenner.ac.uk/MHCpred). An immunogenic fragment of a polypeptide described herein comprises at least one immunogenic portion, as measured experimentally or identified by algorithm. Peptides identified by the tools described above include the following:

5 portion, as measured experimentally or identified by algorithm. Peptides identified by the tools described above include the following:

<b>SP2108 Fragments (SEQ ID NOS 34-57, respectively, in order of appearance)</b>	<b>SP0148 Fragments (SEQ ID NOS 58-82, respectively, in order of appearance)</b>	<b>SP1634 Fragments (SEQ ID NOS 83-109, respectively, in order of appearance)</b>	<b>SP0882 Fragments (SEQ ID NOS 110-130, respectively, in order of appearance)</b>	<b>SP0314 Fragments (SEQ ID NOS 131-169, respectively, in order of appearance)</b>
AIIDGPWKA VMMAPYDR V SIAGINYAK VWDPAKNM L QPLPNISQM APYDRVGS APAVIESLV FYYTYGLLA SKYAFAGE TEGAGNLI LADWTNFY Y SLVMYYNK D KEAGVKVTL KSTAVLGT GAKTDDTTK SQKFVDFLV QAFKDAKV N AVIESLVMY DAKTAAND A YGVATIPTL KTAIIIDGP KAYEKEAGV AGNGAYVF	ALGLVAAG V ELTGYEIEV AVNNLSTYK TYLPAEADI RYNMAVNN L DFQQIMVRL EHTDNPTIL APIAQNPNV PSDQQPYV YVYPLLAQG QGLDNLKVI KYLYAAPI GELTGYEI NPNVLVVK K KLSKQFFGD GSPRPFIYE AVNNLSTYK KIFDKIGVE MVRLSDGQF YVYPLLAQG VVQATTSAK TLEKLSKQF VAAGVLAA C LDNLKVIEL NMAVNNLS	RLLDLAPQ V MLEIPAHQI KNFFAHHP K KVILAGHS K SFDNLVSTL YYDLPLNE L YFDLFFGTI ALEYIHHLF LPLNELDIL IPQGSIIGM DPELQKQF A AVYTFDAP G QSLTPEERE AIYAASQI LEIPAHQI LLDLAPQV P WQIEDKHF V TLGRLTQLL LYFDLFFGT SINDLASLK SINDLASLK	HLDNLVLK V DLIAGRVHL ILLPKDYEK EYQDQIGCL YFHDGQNV F NPDISRMIV IPWSENLPD QFGGKGVE Y IGLEYQDQI VYFHDGQN MEVVKPFI YLKMKEHK L KLSPDQRIF RIFIYVGTE FIDETYRTK DTDRSYPVV YIDSSLCYY TQFIGLEYQ KDTDRSYPV LCYYHDLIA NVFNSKESF	MLKDKIAFL SLADYTYK V FLLLGAFYL VLIDGLSQL ILASLGFL GLSQLLPVI FLLNHYMT V MLIPNVDR KLEEMAKQ V VLKRGVYTI KVIAGLLRK TLNYEHMN K NIGYFFFKK KYTDVIEKF KYDDSVSTI TFNQMIKEL DYPETQSVF TPRAINNTL APLLVNGEL YIDHTNVAY KQNGDSYG Y FLLNHYMT V FYLYNGDLS

G AWVIPQAVK	Y	YYDLPLNE L QKVILAGH S GTDDSIIGW TYLSFDNL V FGTILDAGI NQITAVYTF		KSFAPLLV DETvvRTV YIDHTNVAY MLKDKIAFL KLRFKIKTD KLELFYETG KIAFLGSNI SVPRTSYLS FGFGLSLFS STIRSIEQV FRKTTDNPF TVVRTVRDS STIRSIEQV DGLSQLLPV FGFGLSLFS KLVDQGEG F
----------------	---	--	--	--

SP0024 Fragments (SEQ ID NOS 170-193, respectively, in order of appearance)	SP1072 Fragments (SEQ ID NOS 194-227, respectively, in order of appearance)	SP0641 Fragments (SEQ ID NOS 228-264, respectively, in order of appearance)
AIUTCMDSR AQTFENEFP AYVALHGQL DDVIISGAI FENEFPQEY FMQANQAYV ISQQQMGR KPKTRVAIV LHGQLNLPL LHVAQALGL LPLPKKTRV MGTREIVVL MQLLIESPL QANQAYVAL QFMQANQAY QLNLPLKPK QQMGTREIV REIVVLHHT SPLIPDDVI SRLHVAQAL TEDMIRSLV VDVSDQDFL	GIEVEKPLY AEAHILLYRM ALLNQDNMR APPERNYLY AQNSYIHL AVASMGtal AYLLTKTRI DAAKFYHAI DTALEELER EEYQGVPFI EFLEKIAPL EFQVLYDLL EHVEHLKRL ELSEVEMTR ESPLVLNDY GEKTPSFNV GLCPFHGEK IGDMPVQIV ITMPVTKQL KALLNQDNM KRLTKKLVL LTKTRISPI	AAYAPNEVV AGDLRGKII DEIANEVWY DNYLIYGDL DQKEHPEKF DSLTDRLKL EAKNKNKFV EGQGRNRKL EIKGAGDLR EPIAEGQYF EVSELKPHR GAFFDKSKI GDLKWDGLI GEVEKNLEV IHFESVEEM IMFIVGIFL IPGTLNKGI IRYQVFTFK ISDKGGFNW IVSEEDFIL KEIGVEEAI KIVVKDFAR

VSDQDFLPF VTEDMIRSL	LVLVYDGDK MRAEAHLLY NGPEDLAYL QTEEVERAW SEIYLMEGF SPHQALYDM VDKQVIEEI VEMTRNKAL VLYDLLGQY VPFIEAVQI WYQVLAQDL YLMEGFMDV	KKINFQPSL KLKFVYIGK KVYYGNNYK KYWQAIRAL LHIDNTRDF MRFKKEDLK NESVVDNYL NEVWYAGAA NINDIVDGL QYLLKDNII SPRQQGAGL SRSKTLGGY SSLKNTKVL TAAVILAAY WTELPAMGY
------------------------	--	---

Thus, in some aspects, this application provides an immunogenic fragment of an antigen described herein. The fragments, in some instances, are close in size to the full-length polypeptide or the polypeptide of Table 1 or 2. For example, they 5 may lack at most one, two, three, four, five, ten, or twenty amino acids from one or both termini. In certain embodiments, the polypeptide is 100-500 amino acids in length, or 150-450, or 200-400, or 250-250 amino acids in length. In some embodiments, the polypeptide is 100-200, 150-250, 200-300, 250-350, 300-400, 350-450, or 400-500 amino acids in length. The fragments described above or sub-10 fragments thereof (e.g., fragments of 8-50, 8-30, or 8-20 amino acid residues) preferably have one of the biological activities described below, such as increasing the amount of IL-17 released by at least 1.5 fold or 2 fold or more (e.g., either as an absolute measure or relative to an immunologically inactive protein such as ovalbumin). A fragment may be used as the polypeptide in the vaccines described 15 herein or may be fused to another protein, protein fragment or a polypeptide.

In some embodiments, the fragment is a truncated fragment of any of SEQ ID NOS: 1-13 having from 1-5, 1-10, or 1-20 amino acid residues removed from the N-terminus, C-terminus, or both. In some such embodiments, the same number of residues is removed from the N-terminus and the C-terminus, while in other 20 embodiments, a different number of residues is removed from the N-terminus compared to the C-terminus.

In certain aspects, this application provides immunogenic polypeptides with at least 90%, 95%, 97%, 98%, 99%, or 99.5% identity to a polypeptide of Table 1 or 2. The present disclosure also provides a vaccine formulation comprising a pharmaceutically acceptable carrier and one or more polypeptides having an amino acid sequence comprising a sequence at least 90%, 95%, 98%, or 99% identical to any of SEQ ID NOS: 1-11 or an immunogenic fragment thereof, and optionally further comprising a polypeptide having an amino acid sequence comprising a sequence at least 90%, 95%, 98%, or 99% identical to either of SEQ ID NOS: 12 or 13 or an immunogenic fragment thereof. In certain embodiments, the vaccine formulation comprises at least two different polypeptides having an amino acid sequence comprising a sequence at least 90%, 95%, 98%, or 99% identical to any of SEQ ID NOS: 1-13 or an immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising a sequence at least 90%, 95%, 98%, or 99% identical to one of SEQ ID NOS: 1-10 or an immunogenic fragment thereof.

In some embodiments, one or more, e.g. two, three, four, or more polypeptides from Table 1 or 2 or immunogenic fragments or variants thereof are provided in a mixture. In some embodiments, two, three, four, or more polypeptides from Table 1 or 2 or immunogenic fragments or variants thereof are covalently bound to each other, e.g. as a fusion protein.

In some embodiments, the vaccine formulation contains substantially no other *S. pneumoniae* polypeptides other than polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13. In some embodiments, the vaccine formulation contains substantially no other *S. pneumoniae* polypeptides other than polypeptides of Table 1. In some embodiments, the vaccine formulation contains substantially no other *S. pneumoniae* polypeptides other than polypeptides of Tables 1 or 2.

In certain embodiments, vaccine formulations or immunogenic compositions contain substantially no other *S. pneumoniae* polypeptides other than polypeptides having an amino acid sequence comprising any of SEQ ID NO: 1-13. In certain such embodiments, vaccine formulations or immunogenic compositions contain

substantially no other *S. pneumoniae* polypeptides other than polypeptides having an amino acid sequence consisting of any of SEQ ID NO: 1-13. In some embodiments, vaccine formulations or immunogenic compositions contain substantially no other *S. pneumoniae* polypeptides other than polypeptides having an amino acid sequence

5 comprising (or consisting of) any of the amino acid sequences of the polypeptides of Tables 1 and 2. Substantially, in this context, refers to less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, less than 3%, less than 2, or even less than 1% of the other *S. pneumoniae* polypeptides.

In certain embodiments, the vaccine composition induces a T<sub>H</sub>17 cell

10 response at least 1.5-fold than that induced by a control irrelevant antigen (such as the HSV-2 protein ICP47 with the gene name US12) after contacting T<sub>H</sub>17 cells. In some embodiments, the vaccine formulation inhibits infection by *S. pneumoniae* in an uninfected subject. In certain embodiments, the vaccine formulation inhibits *S. pneumoniae* colonization in an individual. In some embodiments, the vaccine

15 formulation inhibits *S. pneumoniae* symptoms.

In certain embodiments, this application provides nucleic acids encoding one or more of the polypeptides described above, such as DNA, RNA, or an analog thereof. The underlying DNA sequences for the polypeptides described above may be modified in ways that do not affect the sequence of the protein product, and such

20 sequences are included in the invention. For instance, the DNA sequence may be codon-optimized to improve expression in a host such as *E. coli*, an insect cell line (e.g., using the baculovirus expression system), or a mammalian (e.g., human or Chinese Hamster Ovary) cell line.

In certain embodiments, this application provides nucleic acids (such as

25 DNA, RNA, or an analog thereof) that are at least 70%, 80%, 90%, 95%, 97%, 98%, 99%, or 100% identical to a gene in Table 1 or 2, or a variant or portion of said gene. In certain embodiments, the nucleic acid is 600-2000, 800-1800, 1000-1600, 1200-1400 nucleotides in length. In some embodiments, the nucleic acid is 600-1600, 800-1800, or 1000-2000 nucleotides in length. The nucleic acids may be

30 used, for example, for recombinant production of the polypeptides of Tables 1 and 2, or immunogenic fragments thereof.

In some embodiments, the vaccine or immunogenic composition may comprise fusion proteins and/or fusion DNA constructs. The polypeptides described herein may be used without modification. In certain embodiments, when smaller related polypeptides are used, such as fragments or the like, and their molecular weight is less than about 5000 daltons, e.g., 1500 to 5000 daltons, modification may be useful in eliciting the desired immune response. For example, the smaller polypeptides can be conjugated to an appropriate immunogenic carrier such as tetanus toxoid, pneumolysin keyhole limpet hemocyanin or the like. In certain embodiments, the vaccine formulation comprises at least one lipidated polypeptide.

5 Conjugation may be direct or indirect (e.g., via a linker). In other embodiments, a construct may comprise a gene or protein from Table 1 or 2 or an immunogenic fragment or variant thereof and a tag. A tag may be N-terminal or C-terminal. For instance, tags may be added to the nucleic acid or polypeptide to facilitate purification, detection, solubility, or confer other desirable characteristics on the

10 protein or nucleic acid. For instance, a purification tag may be a peptide, oligopeptide, or polypeptide that may be used in affinity purification. Examples include His, GST, TAP, FLAG, myc, HA, MBP, VSV-G, thioredoxin, V5, avidin, streptavidin, BCCP, Calmodulin, Nus, S tags, lipoprotein D, and  $\beta$  galactosidase. Particular exemplary His tags include HHHHHH (SEQ ID NO: 32) and

15 MSYYHHHHHH (SEQ ID NO: 33). In other embodiments, the polypeptide is free of tags such as protein purification tags, and is purified by a method not relying on affinity for a purification tag. In some embodiments, the fused portion is short. This, in some instances, the fusion protein comprises no more than 1, 2, 3, 4, 5, 10, or 20 additional amino acids on one or both termini of the polypeptide of Table 1 or

20 25 2.

## B. Immunogenic compositions

The present disclosure also provides pharmaceutical compositions containing immunogenic polypeptides or polynucleotides encoding these immunogenic polypeptides together with a pharmaceutical carrier. Antigens from *S. pneumoniae* were identified by screening immune cells from mice infected with *S. pneumoniae*,

or from healthy human donors. The human donors had presumably been exposed to *S. pneumoniae* at some point during their lifetimes, because *S. pneumoniae* is a very common disease and colonizing pathogen. Briefly, a library of *S. pneumoniae* antigens was expressed in bacteria and mixed with antigen presenting cells (APCs). 5 The APCs, in turn, presented *S. pneumoniae*-derived polypeptides to lymphocytes that had been isolated from mice or from human donors. Lymphocyte responses were assayed for reactivity to *S. pneumoniae*. Human donors, as well as mice immunized with *S. pneumoniae*, produced lymphocytes specific to *S. pneumoniae* antigens. Thus, the present disclosure contemplates compositions of the *S. 10 pneumoniae* antigens that elicit a strong immune response in immunized or infected mice or humans for counteracting infection by *S. pneumoniae*.

Tables 1 and 2 list the protein sequence and corresponding nucleotide sequence for *S. pneumoniae* antigens identified according to the screening methods described herein. The antigens were identified in screens of mouse and human T 15 cells. In the screens of mouse T cells, the identified antigens were subjected to at least two rounds of screening: a genome-wide round to identify pools of 4 antigens that elicited an immune response, followed by a deconvolution round to individually test and identify single antigens that elicited an immune response from a pool identified in the genome-wide round. In contrast, in the screens of human T cells, 20 two different sets of antigen pools were created, such that a polypeptide was combined with different polypeptides between the first and second pools. Consequently, it is possible to determine which polypeptides are antigens by identifying which polypeptides are in positive pools in both the first and second sets. Table 1 lists antigens (and variants thereof) that were identified by one of the above 25 screening methods, and were subsequently subjected to further testing in the mouse model described in Examples 5-8. Thus, compositions according to this disclosure may include one or two or more of the genes listed in Table 1 or 2, or the corresponding gene products.

An immunogenic composition may also comprise portions of said 30 *Streptococcus* polypeptides, for example deletion mutants, truncation mutants, oligonucleotides, and peptide fragments. In some embodiments, the portions of said

polypeptides are immunogenic. The immunogenicity of a portion of a protein is readily determined using the same assays that are used to determine the immunogenicity of the full-length protein. In some embodiments, the portion of the polypeptide has substantially the same immunogenicity as the full-length proteins.

- 5 In some embodiments, the immunogenicity is no more than 10%, 20%, 30%, 40%, or 50% less than that of the full-length protein (e.g., polypeptides of Tables 1 and 2). The peptide fragments may be, for example, linear, circular, or branched.

Some embodiments of the vaccine formulations and immunogenic compositions described herein include an immunogenic polypeptide (e.g., a 10 polypeptide of Table 1 or 2) that contains a membrane translocating sequence (MTS), to facilitate introduction of the polypeptide into the mammalian cell and subsequent stimulation of the cell-mediated immune response. Exemplary membrane translocating sequences include hydrophobic region in the signal sequence of Kaposi fibroblast growth factor, the MTS of  $\alpha$ -synuclein,  $\beta$ -synuclein, 15 or  $\gamma$ -synuclein, the third helix of the Antennapedia homeodomain, SN50, integrin  $\beta$ 3 h-region, HIV Tat, pAntp, PR-39, abaecin, apidaecin, Bac5, Bac7, P. berghei CS protein, and those MTSs described in US Patents 6,248,558, 6,432,680 and 6,248,558.

In certain embodiments, an antigen (e.g., a polypeptide of Table 1 or 2) is 20 covalently bound to another molecule. This may, for example, increase the half-life, solubility, bioavailability, or immunogenicity of the antigen. Molecules that may be covalently bound to the antigen include a carbohydrate, biotin, poly(ethylene glycol) (PEG), polysialic acid, N-propionylated polysialic acid, nucleic acids, polysaccharides, and PLGA. There are many different types of PEG, ranging from 25 molecular weights of below 300 g/mol to over 10,000,000 g/mol. PEG chains can be linear, branched, or with comb or star geometries. In some embodiments, the naturally produced form of a protein is covalently bound to a moiety that stimulates the immune system. An example of such a moiety is a lipid moiety. In some instances, lipid moieties are recognized by a Toll-like receptor (TLR) such as TLR2, 30 and activate the innate immune system.

**C. Antibodies specific to the proteins of Tables 1 and 2**

Another aspect disclosed herein is an antibody preparation generated against an antigenic composition (e.g., one of the proteins listed in Table 1 or 2 or an immunogenic fragment thereof). For instance, this disclosure provides combinations 5 of two, three, four, or five antibodies each recognizing a different protein of Table 1 or 2. Any of a variety of antibodies are included. Such antibodies include, e.g., polyclonal, monoclonal, recombinant, humanized or partially humanized, single chain, Fab, and fragments thereof, etc. The antibodies can be of any isotype, e.g., IgG, various IgG isotypes such as IgG1, IgG2, IgG2a, IgG2b, IgG3, IgG4, etc.; and 10 they can be from any animal species that produces antibodies, including goat, rabbit, mouse, chicken or the like. In some embodiments, Fab molecules are expressed and assembled in a genetically transformed host like *E. coli*. A lambda vector system is available thus to express a population of Fab's with a potential diversity equal to or exceeding that of subject generating the predecessor antibody. See Huse et al. 15 (1989), Science 246, 1275-81.

**D. Components of a vaccine or immunogenic composition comprising *S. pneumoniae* antigens or antibodies recognizing the same**

In certain embodiments, the vaccine or immunogenic composition comprises 20 an antigen and one or more of the following: an adjuvant, stabilizer, buffer, surfactant, controlled release component, salt, preservative, and an antibody specific to said antigen.

**1. *Adjuvants***

The vaccine formulations and immunogenic compositions described herein 25 may include an adjuvant. Adjuvants can be broadly separated into two classes, based on their principal mechanisms of action: vaccine delivery systems and immunostimulatory adjuvants (see, e.g., Singh *et al.*, *Curr. HIV Res.* 1:309-20, 2003). Vaccine delivery systems are often particulate formulations, e.g., emulsions, microparticles, immune-stimulating complexes (ISCOMs), which may be, for 30 example, particles and/or matrices, and liposomes. In contrast, immunostimulatory

adjuvants are sometimes derived from pathogens and can represent pathogen associated molecular patterns (PAMP), *e.g.*, lipopolysaccharides (LPS), monophosphoryl lipid (MPL), or CpG-containing DNA, which activate cells of the innate immune system.

5        Alternatively, adjuvants may be classified as organic and inorganic. Inorganic adjuvants include alum salts such as aluminum phosphate, amorphous aluminum hydroxyphosphate sulfate, and aluminum hydroxide, which are commonly used in human vaccines. Organic adjuvants comprise organic molecules including macromolecules. An example of an organic adjuvant is cholera toxin.

10        Adjuvants may also be classified by the response they induce. In some embodiments, the adjuvant induces the activation of  $T_{H1}$  cells or  $T_{H2}$  cells. In other embodiments, the adjuvant induces the activation of B cells. In yet other embodiments, the adjuvant induces the activation of antigen-presenting cells. These categories are not mutually exclusive; in some cases, an adjuvant activates more  
15        than one type of cell.

      In certain embodiments, the adjuvant induces the activation of  $T_{H17}$  cells. It may promote the  $T_{H17}$  cells to secrete IL-17. In some embodiments, an adjuvant that induces the activation of  $T_{H17}$  cells is one that produces at least a 2-fold, and in some cases a 10-fold, experimental sample to control ratio in the following assay. In  
20        the assay, an experimenter compares the IL-17 levels secreted by two populations of cells: (1) cells treated with the adjuvant and a polypeptide known to induce  $T_{H17}$  activation, and (2) cells treated with the adjuvant and an irrelevant (control) polypeptide. An adjuvant that induces the activation of  $T_{H17}$  cells may cause the cells of population (1) to produce more than 2-fold, or more than 10-fold more IL-17  
25        than the cells of population (2). IL-17 may be measured, for example, by ELISA or Western blot. Certain toxins, such as cholera toxin and labile toxin (produced by enterotoxigenic *E. coli*, or ETEC), activate a  $T_{H17}$  response. Thus, in some embodiments, the adjuvant is a toxin. Cholera toxin was successfully used in the mouse model to induce protective immunity in conjunction with certain  
30        polypeptides from Table 1 (see Examples 5-8). One form of labile toxin is produced by Intercell. Mutant derivates of labile toxin that are active as adjuvants but

significantly less toxic can be used as well. Exemplary detoxified mutant derivatives of labile toxin include mutants lacking ADP-ribosyltransferase activity. Particular detoxified mutant derivatives of labile toxin include LTK7 (Douce *et al.*, “Mutants of Escherichia coli heat-labile toxin lacking ADP-ribosyltransferase activity act as nontoxic, mucosal adjuvants” PNAS Vol. 92, pp. 1644-1648, February 1995) and LTK63 (Williams *et al.*, “Innate Imprinting by the Modified Heat-Labile Toxin of Escherichia coli (LTK63) Provides Generic Protection against Lung Infectious Disease” The Journal of Immunology, 2004, 173: 7435-7443), LT-G192 (Douce *et al.* “Genetically detoxified mutants of heat-labile toxin from Escherichia coli are able to act as oral adjuvants” Infect Immun. 1999 Sep;67(9):4400-6), and LTR72 (“Mucosal adjuvanticity and immunogenicity of LTR72, a novel mutant of Escherichia coli heat-labile enterotoxin with partial knockout of ADP-ribosyltransferase activity.” J Exp Med. 1998 Apr 6;187(7):1123-32).

15            In some embodiments, the adjuvant comprises a VLP (virus-like particle). One such adjuvant platform, Alphavirus replicons, induces the activation of T<sub>H</sub>17 cells using alphavirus and is produced by Alphavax. In certain embodiments of the Alphavirus replicon system, alphavirus may be engineered to express an antigen of interest, a cytokine of interest (for example, IL-17 or a cytokine that stimulates IL-17 production), or both, and may be produced in a helper cell line. More detailed information may be found in U.S. Patent Nos. 5,643,576 and 6,783,939. In some embodiments, a vaccine formulation is administered to a patient in combination with a nucleic acid encoding a cytokine.

25            Certain classes of adjuvants activate toll-like receptors (TLRs) in order to activate a T<sub>H</sub>17 response. TLRs are well known proteins that may be found on leukocyte membranes, and recognize foreign antigens (including microbial antigens). Administering a known TLR ligand together with an antigen of interest (for instance, as a fusion protein) can promote the development of an immune response specific to the antigen of interest. One exemplary adjuvant that activates 30 TLRs comprises Monophosphoryl Lipid A (MPL). Traditionally, MPL has been produced as a detoxified lipopolysaccharide (LPS) endotoxin obtained from gram

negative bacteria, such as *S. minnesota*. In particular, sequential acid and base hydrolysis of LPS produces an immunoactive lipid A fraction (which is MPL), and lacks the saccharide groups and all but one of the phosphates present in LPS. A number of synthetic TLR agonists (in particular, TLR4 agonists) are disclosed in

5 Evans JT et al. "Enhancement of antigen-specific immunity via the TLR4 ligands MPL adjuvant and Ribi.529." *Expert Rev Vaccines* 2003 Apr;2(2):219-29. Like MPL adjuvants, these synthetic compounds activate the innate immune system via TLR. Another type of TLR agonist is a synthetic phospholipid dimer, for example E6020 (Ishizaka ST *et al.* "E6020: a synthetic Toll-like receptor 4 agonist as a

10 vaccine adjuvant." *Expert Rev. Vaccines.* 2007 Oct; 6(5):773-84.). Various TLR agonists (including TLR4 agonists) have been produced and/or sold by, for example, the Infectious Disease Research Institute (IRDI), Corixa, Esai, Avanti Polar Lipids, Inc., and Sigma Aldrich. Another exemplary adjuvant that activates TLRs comprises a mixture of MPL, Trehalose Dicoynomycolate (TDM), and

15 dioctadecyldimethylammonium bromide (DDA). Another TLR-activating adjuvant is R848 (resiquimod).

In some embodiments, the adjuvant is or comprises a saponin. Typically, the saponin is a triterpene glycoside, such as those isolated from the bark of the Quillaja saponaria tree. A saponin extract from a biological source can be further

20 fractionated (e.g., by chromatography) to isolate the portions of the extract with the best adjuvant activity and with acceptable toxicity. Typical fractions of extract from Quillaja saponaria tree used as adjuvants are known as fractions A and C.

A particular form of saponins that may be used in vaccine formulations described herein is immunostimulating complexes (ISCOMs). ISCOMs are an art-  
25 recognized class of adjuvants, that generally comprise Quillaja saponin fractions and lipids (e.g., cholesterol and phospholipids such as phosphatidyl choline). In certain embodiments, an ISCOM is assembled together with a polypeptide or nucleic acid of interest. However, different saponin fractions may be used in different ratios. In addition, the different saponin fractions may either exist together in the same  
30 particles or have substantially only one fraction per particle (such that the indicated ratio of fractions A and C are generated by mixing together particles with the

different fractions). In this context, "substantially" refers to less than 20%, 15%, 10%, 5%, 4%, 3%, 2% or even 1%. Such adjuvants may comprise fraction A and fraction C mixed into a ratio of 70-95 A: 30-5 C, such as 70 A : 30 C to 75 A : 5 C, 75 A : 5 C to 80 A : 20 C, 80 A : 20 C to 85 A : 15 C, 85 A : 15 C to 90 A : 10 C, 90 A : 10 C to 95 A : 5 C, or 95 A : 5 C to 99 A : 1 C.

5 In certain embodiments, combinations of adjuvants are used. Three exemplary combinations of adjuvants are MPL and alum, E6020 and alum, and MPL and an ISCOM.

10 Adjuvants may be covalently bound to antigens. In some embodiments, the adjuvant may comprise a protein which induces inflammatory responses through activation of antigen-presenting cells (APCs). In some embodiments, one or more of these proteins can be recombinantly fused with an antigen of choice, such that the resultant fusion molecule promotes dendritic cell maturation, activates dendritic cells to produce cytokines and chemokines, and ultimately, enhances presentation of 15 the antigen to T cells and initiation of T cell responses (see Wu et al., *Cancer Res* 2005; 65(11), pp 4947-4954). In certain embodiments, a polypeptide described herein is presented in the context of the trivalent *S. pneumoniae* Pneumococcal surface adhesin A: pneumolysin derivative carrying three amino acid substitutions (W433F, D385N, and C428G) which render the molecule nontoxic but do not 20 interfere with TLR4-mediated inflammatory properties-cell wall polysaccharide (PsaA:PdT-CPs) conjugate system described in Lu et al. ("Protection against Pneumococcal colonization and fatal pneumonia by a trivalent conjugate of a fusion protein with the cell wall polysaccharide." *Infect Immun.* 2009 May;77(5):2076-83). The conjugate system is "a fusion protein of PsaA with the pneumolysin 25 nontoxic derivative PdT and then coupled CPs to the fusion protein". In some embodiments, one or more polypeptides described herein is used in place of PsaA in the trivalent conjugate. The trivalent conjugate system typically includes alum and is usually administered parenterally. Other exemplary adjuvants that may be covalently bound to antigens comprise polysaccharides, pneumolysin, synthetic 30 peptides, lipopeptides, and nucleic acids.

Typically, the same adjuvant or mixture of adjuvants is present in each dose of a vaccine. Optionally, however, an adjuvant may be administered with the first dose of vaccine and not with subsequent doses (i.e., booster shots). Alternatively, a strong adjuvant may be administered with the first dose of vaccine and a weaker 5 adjuvant or lower dose of the strong adjuvant may be administered with subsequent doses. The adjuvant can be administered before the administration of the antigen, concurrent with the administration of the antigen or after the administration of the antigen to a subject (sometimes within 1, 2, 6, or 12 hours, and sometimes within 1, 2, or 5 days). Certain adjuvants are appropriate for human patients, non-human 10 animals, or both.

## ***2. Additional components of a vaccine or immunogenic composition***

In addition to the antigens and the adjuvants described above, a vaccine formulation or immunogenic composition may include one or more additional components.

15 In certain embodiments, the vaccine formulation or immunogenic composition may include one or more stabilizers such as sugars (such as sucrose, glucose, or fructose), phosphate (such as sodium phosphate dibasic, potassium phosphate monobasic, dibasic potassium phosphate, or monosodium phosphate), glutamate (such as monosodium L-glutamate), gelatin (such as processed gelatin, hydrolyzed gelatin, or porcine gelatin), amino acids (such as arginine, asparagine, histidine, L-histidine, alanine, valine, leucine, isoleucine, serine, threonine, lysine, 20 phenylalanine, tyrosine, and the alkyl esters thereof), inosine, or sodium borate.

In certain embodiments, the vaccine formulation or immunogenic composition includes one or more buffers such as a mixture of sodium bicarbonate 25 and ascorbic acid. In some embodiments, the vaccine formulation may be administered in saline, such as phosphate buffered saline (PBS), or distilled water.

In certain embodiments, the vaccine formulation or immunogenic composition includes one or more surfactants such as polysorbate 80 (Tween 80), Triton X-100, Polyethylene glycol tert-octylphenyl ether t- 30 Octylphenoxy polyethoxyethanol 4-(1,1,3,3-Tetramethylbutyl)phenyl-polyethylene glycol (TRITON X-100); Polyoxyethylenesorbitan monolaurate Polyethylene glycol

sorbitan monolaurate (TWEEN 20); and 4-(1,1,3,3-Tetramethylbutyl)phenol polymer with formaldehyde and oxirane (TYLOXAPOL). A surfactant can be ionic or nonionic.

5 In certain embodiments, the vaccine formulation or immunogenic composition includes one or more salts such as sodium chloride, ammonium chloride, calcium chloride, or potassium chloride.

In certain embodiments, a preservative is included in the vaccine or immunogenic composition. In other embodiments, no preservative is used. A preservative is most often used in multi-dose vaccine vials, and is less often needed 10 in single-dose vaccine vials. In certain embodiments, the preservative is 2-phenoxyethanol, methyl and propyl parabens, benzyl alcohol, and/or sorbic acid.

In certain embodiments, the vaccine formulation or immunogenic composition is a controlled release formulation.

15 **E. DNA vaccines**

In certain aspects, the vaccine comprises one of the nucleic acids disclosed herein or a nucleic acid corresponding to one of the polypeptides described herein. When a nucleic acid vaccine is administered to a patient, the corresponding gene product (such as a desired antigen) is produced in the patient's body. In some 20 embodiments, nucleic acid vaccine vectors that include optimized recombinant polynucleotides can be delivered to a mammal (including humans) to induce a therapeutic or prophylactic immune response. The nucleic acid may be, for example, DNA, RNA, or a synthetic nucleic acid. The nucleic acid may be single stranded or double stranded.

25 Nucleic acid vaccine vectors (e.g., adenoviruses, liposomes, papillomaviruses, retroviruses, etc.) can be administered directly to the mammal for transduction of cells *in vivo*. The nucleic acid vaccines can be formulated as pharmaceutical compositions for administration in any suitable manner, including parenteral administration.

In determining the effective amount of the vector to be administered in the treatment or prophylaxis of an infection or other condition, the physician evaluates vector toxicities, progression of the disease, and the production of anti-vector antibodies, if any. Often, the dose equivalent of a naked nucleic acid from a vector

5 is from about 1  $\mu$ g to 1 mg for a typical 70 kilogram patient, and doses of vectors used to deliver the nucleic acid are calculated to yield an equivalent amount of therapeutic nucleic acid. Administration can be accomplished via single or divided doses. The toxicity and therapeutic efficacy of the nucleic acid vaccine vectors can be determined using standard pharmaceutical procedures in cell cultures or

10 experimental animals.

A nucleic acid vaccine can contain DNA, RNA, a modified nucleic acid, or a combination thereof. In some embodiments, the vaccine comprises one or more cloning or expression vectors; for instance, the vaccine may comprise a plurality of expression vectors each capable of autonomous expression of a nucleotide coding

15 region in a mammalian cell to produce at least one immunogenic polypeptide. An expression vector often includes a eukaryotic promoter sequence, such as the nucleotide sequence of a strong eukaryotic promoter, operably linked to one or more coding regions. The compositions and methods herein may involve the use of any particular eukaryotic promoter, and a wide variety are known; such as a CMV or

20 RSV promoter. The promoter can be heterologous with respect to the host cell. The promoter used may be a constitutive promoter.

A vector useful in the present compositions and methods can be circular or linear, single-stranded or double stranded and can be a plasmid, cosmid, or episome. In a suitable embodiment, each nucleotide coding region is on a separate vector;

25 however, it is to be understood that one or more coding regions can be present on a single vector, and these coding regions can be under the control of a single or multiple promoters.

Numerous plasmids may be used for the production of nucleic acid vaccines. Suitable embodiments of the nucleic acid vaccine employ constructs using the

30 plasmids VR1012 (Vical Inc., San Diego Calif.), pCMV1.UBF3/2 (S. Johnston, University of Texas) or pcDNA3.1 (InVitrogen Corporation, Carlsbad, Calif.) as the vector. In addition, the vector construct can contain immunostimulatory sequences

(ISS), such as unmethylated dCpG motifs, that stimulate the animal's immune system. The nucleic acid vaccine can also encode a fusion product containing the immunogenic polypeptide. Plasmid DNA can also be delivered using attenuated bacteria as delivery system, a method that is suitable for DNA vaccines that are 5 administered orally. Bacteria are transformed with an independently replicating plasmid, which becomes released into the host cell cytoplasm following the death of the attenuated bacterium in the host cell.

An alternative approach to delivering the nucleic acid to an animal involves the use of a viral or bacterial vector. Examples of suitable viral vectors include 10 adenovirus, polio virus, pox viruses such as vaccinia, canary pox, and fowl pox, herpes viruses, including catfish herpes virus, adenovirus-associated vector, and retroviruses. Exemplary bacterial vectors include attenuated forms of *Salmonella*, *Shigella*, *Edwardsiella ictaluri*, *Yersinia ruckerii*, and *Listeria monocytogenes*. In some embodiments, the nucleic acid is a vector, such as a plasmid, that is capable of 15 autologous expression of the nucleotide sequence encoding the immunogenic polypeptide.

#### **F. Use of Vaccines**

The *S. pneumoniae* vaccines described herein may be used for prophylactic 20 and/or therapeutic treatment of *S. pneumoniae*. Accordingly, this application provides a method for treating a subject suffering from or susceptible to *S. pneumoniae* infection, comprising administering an effective amount of any of the vaccine formulations described herein. In some aspects, the method inhibits *S. pneumoniae* colonization in an individual. In some aspects, the method inhibits *S. pneumoniae* symptoms. The subject receiving the vaccination may be a male or a 25 female, and may be a child or adult. In some embodiments, the subject being treated is a human. In other embodiments, the subject is a non-human animal.

##### **1. Prophylactic use**

In prophylactic embodiments, the vaccine is administered to a subject to 30 induce an immune response that can help protect against the establishment of *S. pneumoniae*, for example by protecting against colonization, the first and necessary

step in disease. Thus, in some aspects, the method inhibits infection by *S. pneumoniae* in a noncolonized or uninfected subject. In another aspect, the method may reduce the duration of colonization in an individual that is already colonized.

In some embodiments, the vaccine compositions of the invention confer 5 protective immunity, allowing a vaccinated individual to exhibit delayed onset of symptoms or reduced severity of symptoms, as the result of his or her exposure to the vaccine. In certain embodiments, the reduction in severity of symptoms is at least 25%, 40%, 50%, 60%, 70%, 80% or even 90%. In particular embodiments, vaccinated individuals may display no symptoms upon contact with *S. pneumoniae*, 10 do not become colonized by *S. pneumoniae*, or both. Protective immunity is typically achieved by one or more of the following mechanisms: mucosal, humoral, or cellular immunity. Mucosal immunity is primarily the result of secretory IgA (sIgA) antibodies on mucosal surfaces of the respiratory, gastrointestinal, and genitourinary tracts. The sIgA antibodies are generated after a series of events 15 mediated by antigen-processing cells, B and T lymphocytes, that result in sIgA production by B lymphocytes on mucosa-lined tissues of the body. Humoral immunity is typically the result of IgG antibodies and IgM antibodies in serum. Cellular immunity can be achieved through cytotoxic T lymphocytes or through delayed-type hypersensitivity that involves macrophages and T lymphocytes, as well 20 as other mechanisms involving T cells without a requirement for antibodies. In particular, cellular immunity may be mediated by T<sub>H</sub>1 or T<sub>H</sub>17 cells.

Essentially any individual has a certain risk of becoming infected with *S. pneumoniae*. However, certain sub-populations have an increased risk of infection. In some embodiments, a vaccine formulation as described herein (e.g., a 25 composition comprising one or more polypeptides from Table 1 or 2, or nucleic acids encoding the polypeptides, or antibodies reactive with the polypeptides) is administered to patients that are immunocompromised.

An immunocompromising condition arising from a medical treatment is likely to expose the individual in question to a higher risk of infection with *S. 30 pneumoniae*. It is possible to treat an infection prophylactically in an individual having the immunocompromised condition before or during treatments known to compromise immune function. By prophylactically treating with an antigenic

composition (e.g., two or more antigens from Table 1 or 2, or nucleic acids encoding the antigens), or with antibodies reactive to two or more antigens from Table 1 or 2, before or during a treatment known to compromise immune function, it is possible to prevent a subsequent *S. pneumoniae* infection or to reduce the risk of the

5 individual contracting an infection due to the immunocompromised condition.

Should the individual contract an *S. pneumoniae* infection e.g., following a treatment leading to an immunocompromised condition it is also possible to treat the infection by administering to the individual an antigen composition.

The following groups are at increased risk of pneumococcal disease or its

10 complications, and therefore it is advantageous for subjects falling into one or more of these groups to receive a vaccine formulation described herein: children, especially those from 1 month to 5 years old or 2 months to 2 years old; children who are at least 2 years of age with asplenia, splenic dysfunction or sickle-cell disease; children who are at least 2 years of age with nephrotic syndrome, chronic

15 cerebrospinal fluid leak, HIV infection or other conditions associated with immunosuppression.

In another embodiment, at least one dose of the pneumococcal antigen composition is given to adults in the following groups at increased risk of pneumococcal disease or its complications: all persons 65 years of age; adults with

20 asplenia, splenic dysfunction or sickle-cell disease; adults with the following conditions: chronic cardiorespiratory disease, cirrhosis, alcoholism, chronic renal disease, nephrotic syndrome, diabetes mellitus, chronic cerebrospinal fluid leak, HIV infection, AIDS and other conditions associated with immunosuppression (Hodgkin's disease, lymphoma, multiple myeloma, immunosuppression for organ

25 transplantation), individuals with cochlear implants; individuals with long-term health problems such as heart disease and lung disease, as well as individuals who are taking any drug or treatment that lowers the body's resistance to infection, such as long-term steroids, certain cancer drugs, radiation therapy; Alaskan natives and certain Native American populations.

## 2. *Therapeutic use*

In therapeutic applications, the vaccine may be administered to a patient suffering from *S. pneumoniae* infection, in an amount sufficient to treat the patient. Treating the patient, in this case, refers to reducing symptoms, bacterial load, or both 5 of *S. pneumoniae* in an infected individual. In some embodiments, treating the patient refers to reducing the duration of symptoms or reducing the intensity of symptoms. In some embodiments, the vaccine reduces transmissibility of *S. pneumoniae* from the vaccinated patient. In certain embodiments, the reductions described above are at least 25%, 30%, 40%, 50%, 60%, 70%, 80% or even 90%.  
10 In therapeutic embodiments, the vaccine is administered to an individual post-infection. The vaccine may be administered shortly after infection, e.g. before symptoms manifest, or may be administered during or after manifestation of symptoms.

A therapeutic *S. pneumoniae* vaccine can reduce the intensity and/or duration 15 symptoms of the various indications of *S. pneumoniae* infection. A *S. pneumoniae* infection can take many forms. In some cases, an infected patient develops pneumonia, acute sinusitis, otitis media (ear infection), meningitis, bacteremia, sepsis, osteomyelitis, septic arthritis, endocarditis, peritonitis, pericarditis, cellulitis, or brain abscess.

## 20 3. *Assaying vaccination efficacy*

The efficacy of vaccination with the vaccines disclosed herein may be determined in a number of ways, in addition to the clinical outcomes described above. First, one may assay IL-17 levels (particularly IL-17A) by stimulating T 25 cells derived from the subject after vaccination. The IL-17 levels may be compared to IL-17 levels in the same subject before vaccination. Increased IL-17 (e.g., IL-17A) levels, such as a 1.5 fold, 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold or more increase, would indicate an increased response to the vaccine. Alternatively (or in combination), one may assay neutrophils in the presence of T cells or antibodies from the patient for pneumococcal killing. Increased pneumococcal 30 killing, such as a 1.5 fold, 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold or more increase, would indicate an increased response to the vaccine. In addition, one

may measure  $T_{H17}$  cell activation, where increased  $T_{H17}$  cell activation, such as a 1.5 fold, 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold or more increase, correlates with an increased response to the vaccine. One may also measure levels of an antibody specific to the vaccine, where increased levels of the specific 5 antibody, such as a 1.5 fold, 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or 100-fold or more increase, are correlated with increased vaccine efficacy. In certain embodiments, two or more of these assays are used. For example, one may measure IL-17 levels and the levels of vaccine-specific antibody. Alternatively, one may follow epidemiological markers such as incidence of, severity of, or duration of 10 pneumococcal infection in vaccinated individuals compared to unvaccinated individuals.

Vaccine efficacy may also be assayed in various model systems such as the mouse model. For instance, BALB/c or C57BL/6 strains of mice may be used. After administering the test vaccine to a subject (as a single dose or multiple doses), 15 the experimenter administers a challenge dose of *S. pneumoniae*. In some cases, the challenge dose is sufficient to cause *S. pneumoniae* colonization (especially nasal colonization) in an unvaccinated animal, and in some cases the challenge dose is sufficient to cause a high rate of lethality in unvaccinated animals. One can then measure the reduction in colonization or the reduction in lethality in vaccinated 20 animals. Examples 5 and 6 show the efficacy of polypeptides of Table 1 in inhibiting *S. pneumoniae* nasal colonization in the mouse model.

## **G. Use of Immunogenic Compositions**

### **1. Defense against *S. pneumoniae* infection**

25 The immunogenic compositions of the present disclosure are designed to elicit an immune response against *S. pneumoniae*. Compositions described herein (e.g., ones comprising one or more polypeptides of Table 1 or 2, or nucleic acids encoding the polypeptides) may stimulate an antibody response or a cell-mediated immune response, or both, in the mammal to which it is administered. In some 30 embodiments, the composition stimulates a  $T_{H1}$ -biased CD4+ T cell response, a  $T_{H17}$ -biased CD4+ T cell response, or a CD8+ T cell response; in the case of a

single component composition, the composition may stimulate an antibody response, a  $T_{H}1$ -biased CD4+ T cell response,  $T_{H}17$ -biased CD4+ T cell response, and/or a CD8+ T cell response.

5 In certain embodiments, the composition (e.g., one comprising one or more polypeptides of Table 1 or 2, or nucleic acids encoding the polypeptides, or antibodies reactive with the peptides) includes a cytokine or nucleotide coding region encoding a cytokine such as IL-17, to provide additional stimulation to the immune system of the mammal. In certain embodiments, the composition comprises a cytokine such as IL-17.

10 While not wishing to be bound by theory, in some embodiments a  $T_{H}17$  cell response is beneficial in mounting an immune response to the compositions disclosed herein, e.g., ones comprising one or more polypeptides of Table 1 or 2. In certain embodiments, an active  $T_{H}17$  response is beneficial in clearing a pneumococcal infection. For instance, mice lacking the IL-17A receptor show 15 decreased whole cell vaccine-based protection from a pneumococcal challenge (Lu *et al.*, "Interleukin-17A mediates acquired immunity to pneumococcal colonization." PLoS Pathog. 2008 Sep 19;4(9)). Furthermore, the same authors showed that the response level of IL-17A was increased in mice treated with a whole-cell vaccine.

20 Thus, herein is provided a method of increasing IL-17 production by administering the compositions described herein (e.g., ones comprising one or more polypeptides of Table 1 or 2) to a subject. Furthermore, this application provides a method of activating  $T_{H}17$  cells by administering said compositions to a subject. In certain embodiments, increased IL-17A levels result in increased pneumococcal killing by neutrophils or neutrophil-like cells, for instance by inducing recruitment 25 and activation of neutrophils or neutrophil-like cells. In certain embodiments, this pneumococcal killing is independent of antibodies and complement. However, specific antibody production and complement activation may be useful additional mechanisms that contribute to clearing of a pneumococcal infection.

30 Immunogenic compositions containing immunogenic polypeptides or polynucleotides encoding immunogenic polypeptides together with a pharmaceutical carrier are also provided.

In some instances, the immunogenic composition comprises one or more nucleic acids encoding one or more polypeptides of SEQ ID NOS: 1-13, such as one or more nucleic acids selected from SEQ ID Nos. 24-31. In some embodiments these nucleic acids are expressed in the immunized individual, producing the 5 encoded *S. pneumoniae* antigens, and the *S. pneumoniae* antigens so produced can produce an immunostimulatory effect in the immunized individual.

Such a nucleic acid-containing immunostimulatory composition may comprise, for example, an origin of replication, and a promoter that drives expression of one or more nucleic acids encoding one or more polypeptides of SEQ 10 ID NOS: 1-13. Such a composition may also comprise a bacterial plasmid vector into which is inserted a promoter (sometimes a strong viral promoter), one or more nucleic acids encoding one or more polypeptides of SEQ ID NOS: 1-13, and a polyadenylation/transcriptional termination sequence. In some instances, the nucleic acid is DNA.

15

## **H. Diagnostic uses**

This application provides, *inter alia*, a rapid, inexpensive, sensitive, and specific method for detection of *S. pneumoniae* in patients. In this respect it should be useful to all hospitals and physicians examining and treating patients with or at 20 risk for *S. pneumoniae* infection. Detection kits can be simple enough to be set up in any local hospital laboratory, and the antibodies and antigen-binding portions thereof can readily be made available to all hospitals treating patients with or at risk for *S. pneumoniae* infection. As used herein, “patient” refers to an individual (such as a human) that either has an *S. pneumoniae* infection or has the potential to 25 contract an *S. pneumoniae* infection. A patient may be an individual (such as a human) that has an *S. pneumoniae* infection, has the potential to contract an *S. pneumoniae* infection, who has recovered from *S. pneumoniae* infection, and/or an individual whose infection status is unknown.

In some embodiments, one may perform a diagnostic assay using two or 30 more antibodies, each of which binds one of the antigens of Table 1 and 2 to detect *S. pneumoniae* in an individual. The instant disclosure also provides a method of

phenotyping biological samples from patients suspected of having a *S. pneumoniae* infection: (a) obtaining a biological sample from a patient; (b) contacting the sample with two or more *S. pneumoniae* -specific antibodies or antigen-binding portions thereof under conditions that allow for binding of the antibody or antigen-binding portion to an epitope of *S. pneumoniae*; where binding indicates the presence of *S. pneumoniae* in the sample. In some embodiments, the binding to the biological sample is compared to binding of the same antibody to a negative control tissue, wherein if the biological sample shows the presence of *S. pneumoniae* as compared to the negative control tissue, the patient is identified as likely having a *S. pneumoniae* infection. In some cases, binding of one antibody indicates the presence of *S. pneumoniae*; in other cases, the binding of two or more antibodies indicates the presence of *S. pneumoniae*. The aforementioned test may be appropriately adjusted to detect other bacterial infections, for instance by using an antibody immunoreactive a homolog (from another bacterial species) of one of the proteins described in Table 1. In some embodiments, the antibodies raised against a *S. pneumoniae* protein in Table 1 or 2 will also bind the homolog in another *Streptococcus* species, especially if the homologs have a high percentage sequence identity.

Alternatively, one may use an antigen of Table 1 or 2 to detect anti-*S. pneumoniae* antibodies in an individual. The instant disclosure also provides a method of phenotyping biological samples from patients suspected of having a *S. pneumoniae* infection: (a) obtaining a biological sample from a patient; (b) contacting the sample with two or more *S. pneumoniae* -specific antigens selected from Table 1 or 2 or portions thereof under conditions that allow for binding of the antigen (or portion thereof) to any host antibodies present in the sample; where binding indicates the presence of anti-*S. pneumoniae* antibodies in the sample. In some embodiments, the binding to the biological sample is compared to binding of the same antigen to a negative control tissue, wherein if the biological sample shows the presence of anti-*S. pneumoniae* antibodies as compared to the negative control tissue, the patient is identified as likely either (1) having a *S. pneumoniae* infection, or (2) having had a *S. pneumoniae* infection in the past. In some cases, detecting one antibody indicates a current or past infection with *S. pneumoniae*; in other cases,

detecting two or more antibodies indicates a current or past infection with *S. pneumoniae*. The aforementioned test may be appropriately adjusted to detect other bacterial infections, for instance by using a homolog (from another bacterial species (e.g., a *Streptococcal* species) of the proteins described in Table 1.

5 In some embodiments, the immune cell response of a mammalian cell may be quantified *ex vivo*. A method for such quantification comprises administering the compositions herein disclosed to a mammalian T cell *ex vivo*, and quantifying the change in cytokine production of the mammalian T cell in response to the composition. In these methods, the cytokine may be, for example, IL-17.

10 The binding of an *S. pneumoniae* antibody to an antigen (e.g., a polypeptide of Table 1 or 2) may be measured using any appropriate method. Such methods include ELISA (enzyme-linked immunosorbent assay), Western blotting, competition assay, and spot-blot. The detection step may be, for instance, chemiluminescent, fluorescent, or colorimetric. One suitable method for measuring 15 antibody-protein binding is the Luminex xMAP system, where peptides are bound to a dye-containing microsphere. Certain systems, including the xMAP system, are amenable to measuring several different markers in multiplex, and could be used to measure levels of antibodies at once. In some embodiments, other systems are used to assay a plurality of markers in multiplex. For example, profiling may be 20 performed using any of the following systems: antigen microarrays, bead microarrays, nanobarcodes particle technology, arrayed proteins from cDNA expression libraries, protein *in situ* array, protein arrays of living transformants, universal protein array, lab-on-a-chip microfluidics, and peptides on pins. Another type of clinical assay is a chemiluminescent assay to detect antibody binding. In 25 some such assays, including the VITROS Eci anti-HCV assay, antibodies are bound to a solid-phase support made up of microparticles in liquid suspension, and a surface fluorometer is used to quantify the enzymatic generation of a fluorescent product.

30 In some embodiments, if the biological sample shows the presence of *S. pneumoniae* (e.g., by detecting one or more polypeptide of Table 1 or 2 or an antibody that binds one of said polypeptides), one may administer a therapeutically

effective amount of the compositions and therapies described herein to the patient. The biological sample may comprise, for example, blood, semen, urine, vaginal fluid, mucus, saliva, feces, urine, cerebrospinal fluid, or a tissue sample. In some embodiments, the biological sample is an organ intended for transplantation. In 5 certain embodiments, before the detection step, the biological sample is subject to culture conditions that promote the growth of *S. pneumoniae*.

The diagnostic tests herein (e.g., those that detect a polypeptide of Table 1 or 2 or an antibody that binds one of said polypeptides) may be used to detect *S. pneumoniae* in a variety of samples, including samples taken from patients and 10 samples obtained from other sources. For example, the diagnostic tests may be used to detect *S. pneumoniae* in food, drink, or ingredients for food and drink; on objects such as medical instruments, medical devices such as cochlear implants and pacemakers, shoes, clothing, furniture including hospital furniture, and drapes including hospital drapes; or in samples taken from the environment such as plant 15 samples. In some embodiments, the tests herein may be performed on samples taken from animals such as agricultural animals (cows, pigs, chickens, goats, horses and the like), companion animals (dogs, cats, birds, and the like), or wild animals. In certain embodiments, the tests herein may be performed on samples taken from cell 20 cultures such as cultures of human cells that produce a therapeutic protein, cultures of bacteria intended to produce a useful biological molecule, or cultures of cells grown for research purposes.

This disclosure also provides a method of determining the location of a *S. pneumoniae* infection in a patient comprising: (a) administering a pharmaceutical composition comprising a labeled *S. pneumoniae* antibody or antigen-binding 25 portion thereof to the patient, and (b) detecting the label, wherein binding indicates a *S. pneumoniae* infection in a particular location in the patient. Such a diagnostic may also comprise comparing the levels of binding in the patient to a control. In certain embodiments, the method further comprises, if the patient has a *S. pneumoniae* infection, treating the infection by administering a therapeutically 30 effective amount of a *S. pneumoniae* -binding antibody or antigen-binding portion thereof to the patient. In certain embodiments, the method further comprises, if the

patient has a *S. pneumoniae* infection, treating the infection by administering a therapeutically effective amount of a *S. pneumoniae* protein of Table 1, or immunogenic portion thereof, to the patient. The method may further comprise determining the location and/or volume of the *S. pneumoniae* in the patient. This 5 method may be used to evaluate the spread of *S. pneumoniae* in the patient and determine whether a localized therapy is appropriate.

In some embodiments, the anti-*S. pneumoniae* antibodies described herein may be used to make a prognosis of the course of infection. In some embodiments, the anti-*S. pneumoniae* antibodies herein may be detected in a sample taken from a 10 patient. If antibodies are present at normal levels, it would indicate that the patient has raised an immune response against anti-*S. pneumoniae*. If antibodies are absent, or present at reduced levels, it would indicate that the patient is failing to raise a sufficient response against anti-*S. pneumoniae*, and a more aggressive treatment would be recommended. In some embodiments, antibodies present at reduced levels 15 refers to antibodies that are present at less than 50%, 20%, 10%, 5%, 2%, or 1% the level of antibodies typical in a patient with a normal immune system. Antibodies may be detected by affinity for any of the antigens described herein (e.g., those in Table 1 and/or 2), for example using ELISA.

In some embodiments, detection of specific *S. pneumoniae* antigens (e.g., 20 those in Table 1 and/or 2) may be used to predict the progress and symptoms of *S. pneumoniae* infection in a patient. It will be understood by one of skill in the art that the methods herein are not limited to detection of *S. pneumoniae*. Other embodiments include the detection of related bacteria including bacteria with proteins homologous to the proteins described in Table 1 or 2. Such related bacteria 25 include, for example, other strains of *Streptococcus*.

## **I. Doses and Routes of Administration**

### ***1. Dosage forms, amounts, and timing***

The amount of antigen in each vaccine or immunogenic composition dose is 30 selected as an effective amount, which induces a prophylactic or therapeutic

response, as described above, in either a single dose or over multiple doses. Preferably, the dose is without significant, adverse side effects in typical vaccinees. Such amount will vary depending upon which specific antigen is employed. Generally, it is expected that a dose will comprise 1-1000  $\mu$ g of protein, in some instances 2-100  $\mu$ g, for instance 4-40  $\mu$ g. In some aspects, the vaccine formulation comprises 1-1000  $\mu$ g of the polypeptide and 1-250  $\mu$ g of the adjuvant. In some embodiments, the appropriate amount of antigen to be delivered will depend on the age, weight, and health (e.g. immunocompromised status) of a subject. When present, typically an adjuvant will be present in amounts from 1  $\mu$ g – 250  $\mu$ g per dose, for example 50-150  $\mu$ g, 75-125  $\mu$ g or 100  $\mu$ g.

In some embodiments, only one dose of the vaccine is administered to achieve the results described above. In other embodiments, following an initial vaccination, subjects receive one or more boost vaccinations, for a total of two, three, four or five vaccinations. Advantageously, the number is three or fewer. A boost vaccination may be administered, for example, about 1 month, 2 months, 4 months, 6 months, or 12 months after the initial vaccination, such that one vaccination regimen involves administration at 0, 0.5-2 and 4-8 months. It may be advantageous to administer split doses of vaccines which may be administered by the same or different routes.

The vaccines and immunogenic compositions described herein may take on a variety of dosage forms. In certain embodiments, the composition is provided in solid or powdered (e.g., lyophilized) form; it also may be provided in solution form. In certain embodiments, a dosage form is provided as a dose of lyophilized composition and at least one separate sterile container of diluent.

In some embodiments, the composition will be administered in a dose escalation manner, such that successive administrations of the composition contain a higher concentration of composition than previous administrations. In some embodiments, the composition will be administered in a manner such that successive administrations of the composition contain a lower concentration of composition than previous administrations.

In therapeutic applications, compositions are administered to a patient suffering from a disease in an amount sufficient to treat the patient. Therapeutic applications of a composition described herein include reducing transmissibility, slowing disease progression, reducing bacterial viability or replication, or inhibiting 5 the expression of proteins required for toxicity, such as by 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20% or 10% of the levels at which they would occur in individuals who are not treated with the composition.

In prophylactic embodiments, compositions are administered to a human or other mammal to induce an immune response that can inhibit the establishment of an 10 infectious disease or other condition. In some embodiments, a composition may partially block the bacterium from establishing an infection.

In some embodiments, the compositions are administered in combination with antibiotics. This co-administration is particularly appropriate when the pharmaceutical composition is administered to a patient who has recently been 15 exposed (or is suspected of having been recently exposed) to *S. pneumoniae*. Many antibiotics are used to treat pneumococcal infections, including penicillin, amoxicillin, amoxicillin/clavulanate, cefuroxime, cefotaxime, ceftriaxone, and vancomycin. The appropriate antibiotic may be selected based on the type and severity of the infection, as well as any known antibiotic resistance of the infection 20 (Jacobs MR "Drug-resistant Streptococcus pneumoniae: rational antibiotic choices" Am J Med. 1999 May 3;106(5A):19S-25S).

## ***2. Routes of administration***

The vaccine formulations and pharmaceutical compositions herein can be delivered by administration to an individual, typically by systemic administration 25 (e.g., intravenous, intraperitoneal, intramuscular, intradermal, subcutaneous, subdermal, transdermal, intracranial, intranasal, mucosal, anal, vaginal, oral, buccal route or they can be inhaled) or they can be administered by topical application. In some embodiments, the route of administration is intramuscular. In other embodiments, the route of administration is subcutaneous. In yet other 30 embodiments, the route of administration is mucosal. In certain embodiments, the route of administration is transdermal or intradermal

Certain routes of administration are particularly appropriate for vaccine formulations and immunogenic compositions comprising specified adjuvants. In particular, transdermal administration is one suitable route of administration for *S. pneumoniae* vaccines comprising toxins (e.g. cholera toxin or labile toxin); in other 5 embodiments, the administration is intranasal. Vaccines formulated with Alphavirus replicons may be administered, for example, by the intramuscular or the subcutaneous route. Vaccines comprising Monophosphory Lipid A (MPL), Trehalose Dicoynomycolate (TDM), and dioctadecyldimethylammonium bromide (DDA) are suitable (*inter alia*) for intramuscular and subcutaneous administration. 10 A vaccine comprising resiquimod may be administered topically or subcutaneously, for example.

### ***3. Formulations***

The vaccine formulation or immunogenic composition may be suitable for administration to a human patient, and vaccine or immunogenic composition 15 preparation may conform to USFDA guidelines. In some embodiments, the vaccine formulation or immunogenic composition is suitable for administration to a non-human animal. In some embodiments, the vaccine or immunogenic composition is substantially free of either endotoxins or exotoxins. Endotoxins may include pyrogens, such as lipopolysaccharide (LPS) molecules. The vaccine or 20 immunogenic composition may also be substantially free of inactive protein fragments which may cause a fever or other side effects. In some embodiments, the composition contains less than 1%, less than 0.1%, less than 0.01%, less than 0.001%, or less than 0.0001% of endotoxins, exotoxins, and/or inactive protein fragments. In some embodiments, the vaccine or immunogenic composition has 25 lower levels of pyrogens than industrial water, tap water, or distilled water. Other vaccine or immunogenic composition components may be purified using methods known in the art, such as ion-exchange chromatography, ultrafiltration, or distillation. In other embodiments, the pyrogens may be inactivated or destroyed prior to administration to a patient. Raw materials for vaccines, such as water, 30 buffers, salts and other chemicals may also be screened and depyrogenated. All materials in the vaccine may be sterile, and each lot of the vaccine may be tested for sterility. Thus, in certain embodiments the endotoxin levels in the vaccine fall

below the levels set by the USFDA, for example 0.2 endotoxin (EU)/kg of product for an intrathecal injectable composition; 5 EU/kg of product for a non-intrathecal injectable composition, and 0.25-0.5 EU/mL for sterile water.

In certain embodiments, the preparation comprises less than 50%, 20%, 10%, 5 or 5% (by dry weight) contaminating protein. In certain embodiments, the desired molecule is present in the substantial absence of other biological macromolecules, such as other proteins (particularly other proteins which may substantially mask, diminish, confuse or alter the characteristics of the component proteins either as purified preparations or in their function in the subject reconstituted mixture). In 10 certain embodiments, at least 80%, 90%, 95%, 99%, or 99.8% (by dry weight) of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). In some embodiments, the vaccine or immunogenic composition comprising purified subunit proteins contains less than 5%, 2%, 1%, 0.5%, 0.2%, 15 0.1% of protein from host cells in which the subunit proteins were expressed, relative to the amount of purified subunit. In some embodiments, the desired polypeptides are substantially free of nucleic acids and/or carbohydrates. For instance, in some embodiments, the vaccine or immunogenic composition contains less than 5%, less than 2%, less than 1%, less than 0.5%, less than 0.2%, or less than 20 0.1% host cell DNA and/or RNA. In certain embodiments, at least 80%, 90%, 95%, 99%, or 99.8% (by dry weight) of biological macromolecules of the same type are present in the preparation (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present).

It is preferred that the vaccine or immunogenic composition has low or no 25 toxicity, within a reasonable risk-benefit ratio. In certain embodiments, the vaccine or immunogenic composition comprises ingredients at concentrations that are less than LD<sub>50</sub> measurements for the animal being vaccinated. LD<sub>50</sub> measurements may be obtained in mice or other experimental model systems, and extrapolated to 30 humans and other animals. Methods for estimating the LD<sub>50</sub> of compounds in humans and other animals are well-known in the art. A vaccine formulation or immunogenic composition, and any component within it, might have an LD<sub>50</sub> value in rats of greater than 100 g/kg, greater than 50g/kg, greater than 20 g/kg, greater

than 10 g/kg, greater than 5 g/kg, greater than 2 g/kg, greater than 1 g/kg, greater than 500 mg/kg, greater than 200 mg/kg, greater than 100 mg/kg, greater than 50 mg/kg, greater than 20 mg/kg, or greater than 10 mg/kg. A vaccine formulation or immunogenic composition that comprises a toxin such as botulinum toxin (which 5 can be used as an adjuvant) should contain significantly less than the LD<sub>50</sub> of botulinum toxin.

The formulations suitable for introduction of the vaccine formulations or pharmaceutical composition vary according to route of administration.

Formulations suitable for parenteral administration, such as, for example, by 10 intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, intranasal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending 15 agents, solubilizers, thickening agents, stabilizers, and preservatives. The formulations can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials.

Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. In the case of adoptive 20 transfer of therapeutic T cells, the cells can be administered intravenously or parenterally.

Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the polypeptides or packaged nucleic acids suspended in diluents, such as water, saline or PEG 400; (b) capsules, sachets or 25 tablets, each containing a predetermined amount of the active ingredient, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; and (d) suitable emulsions. Tablet forms can include one or more of lactose, sucrose, mannitol, sorbitol, calcium phosphates, corn starch, potato starch, tragacanth, microcrystalline cellulose, acacia, gelatin, colloidal silicon dioxide, croscarmellose sodium, talc, 30 magnesium stearate, stearic acid, and other excipients, colorants, fillers, binders, diluents, buffering agents, moistening agents, preservatives, flavoring agents, dyes,

disintegrating agents, and pharmaceutically compatible carriers. Lozenge forms can comprise the active ingredient in a flavor, usually sucrose and acacia or tragacanth, as well as pastilles comprising the active ingredient in an inert base, such as gelatin and glycerin or sucrose and acacia emulsions, gels, and the like containing, in 5 addition to the active ingredient, carriers known in the art. The pharmaceutical compositions can be encapsulated, e.g., in liposomes, or in a formulation that provides for slow release of the active ingredient.

The antigens, alone or in combination with other suitable components, can be made into aerosol formulations (e.g., they can be "nebulized") to be administered via 10 inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like. Aerosol formulations can be delivered orally or nasally.

Suitable formulations for vaginal or rectal administration include, for example, suppositories, which consist of the polypeptides or packaged nucleic acids 15 with a suppository base. Suitable suppository bases include natural or synthetic triglycerides or paraffin hydrocarbons. In addition, it is also possible to use gelatin rectal capsules which consist of a combination of the polypeptides or packaged nucleic acids with a base, including, for example, liquid triglycerides, polyethylene glycols, and paraffin hydrocarbons.

20

## **J. Preparation and Storage of Vaccine Formulations and Immunogenic Compositions**

The *S. pneumoniae* vaccines and immunogenic compositions described herein may be produced using a variety of techniques. For example, a polypeptide 25 may be produced using recombinant DNA technology in a suitable host cell. A suitable host cell may be bacterial, yeast, mammalian, or other type of cell. The host cell may be modified to express an exogenous copy of one of the relevant polypeptide genes. Typically, the gene is operably linked to appropriate regulatory sequences such as a strong promoter and a polyadenylation sequence. In some 30 embodiments, the promoter is inducible or repressible. Other regulatory sequences may provide for secretion or excretion of the polypeptide of interest or retention of

the polypeptide of interest in the cytoplasm or in the membrane, depending on how one wishes to purify the polypeptide. The gene may be present on an extrachromosomal plasmid, or may be integrated into the host genome. One of skill in the art will recognize that it is not necessary to use a nucleic acid 100% identical to the naturally-occurring sequence. Rather, some alterations to these sequences are tolerated and may be desirable. For instance, the nucleic acid may be altered to take advantage of the degeneracy of the genetic code such that the encoded polypeptide remains the same. In some embodiments, the gene is codon-optimized to improve expression in a particular host. The nucleic acid may be produced, for example, by 5 PCR or by chemical synthesis.

10

Once a recombinant cell line has been produced, a polypeptide may be isolated from it. The isolation may be accomplished, for example, by affinity purification techniques or by physical separation techniques (e.g., a size column).

In a further aspect of the present disclosure, there is provided a method of 15 manufacture comprising mixing one or more polypeptides or an immunogenic fragment or variant thereof with a carrier and/or an adjuvant.

In some embodiments, antigens for inclusion in the vaccine formulations and immunogenic compositions may be produced in cell culture. One method comprises providing one or more expression vectors and cloning nucleotides encoding one or 20 more polypeptides selected from polypeptides having an amino acid sequence of Table 1 or Table 2, then expressing and isolating the polypeptides.

The immunogenic polypeptides described herein, and nucleic acid compositions that express the polypeptides, can be packaged in packs, dispenser devices, and kits for administering nucleic acid compositions to a mammal. For 25 example, packs or dispenser devices that contain one or more unit dosage forms are provided. Typically, instructions for administration of the compounds will be provided with the packaging, along with a suitable indication on the label that the compound is suitable for treatment of an indicated condition, such as those disclosed herein.

## V. Examples

### **Example 1. Antigen identification and pooled murine screens**

Each open reading frame predicted in the *S. pneumoniae* TIGR4 genome was cloned into an expression vector comprising a tag that is able to be presented by the 5 major histocompatibility complex (MHC). Each construct was then expressed in *E. coli*, and full-length expression validated by a surrogate assay that identifies the tag in the context of MHC. The screen is described in more detail in International Application WO 2010/002993. In order to facilitate screening the large library, the library was pooled such that four induced library clones were present in each well.

10 In order to screen T cells from mice immunized against *S. pneumoniae*, an aliquot of the pooled library was added to peritoneal-derived macrophages. The macrophages were allowed to bind the tagged *S. pneumoniae* antigens via the MHC. After 2 hr at 37°C, the macrophages were washed with PBS. The macrophages were then fixed with 1% paraformaldehyde for 15 min and washed extensively with PBS. 10<sup>5</sup> T

15 cells were added to each well in 200 µL of RP-10 media. The T cells had previously been isolated from mice that had been immunized 2 times with killed *S. pneumoniae* bacteria with cholera toxin adjuvant. The assay plates were incubated for 72 hrs at 37°C. The amount of IL-17 in the supernatant of each well was determined through the use of an IL-17 ELISA assay. The threshold for a positive result was set at two

20 standard deviations above the mean of all samples.

### **Example 2. Deconvolution of the positive murine pools**

A secondary screen was used to determine which antigen(s) out of the four clones in each well induced the positive response observed in the pooled screen described in Example 1. All the clones in each positive pool were pulsed 25 individually onto peritoneal macrophages in duplicate wells. T cells isolated from immunized mice from the same genetic background as the initial screen were used to screen the pulsed macrophages using the IL-17 assay described in Example 1. Individual antigens that induced an average response in the duplicate wells greater than two standard deviations above the mean of negative control samples were 30 considered positive responses. The library plasmids present in these positive clones were sequenced to confirm the identity of the antigen. The antigens SP\_1574,

SP\_1655, SP\_2106, SP\_0148, SP\_1473, SP\_0605, SP\_1177, SP\_0335, SP\_0906, SP\_1828, SP\_2157, SP\_1229, SP\_1128, SP\_1836, SP\_1865, SP\_0904, SP\_0882, SP\_0765, SP\_1634, SP\_0418, SP\_1923, SP\_1313, SP\_0775, SP\_0314, SP\_0912, SP\_0159, SP\_0910, SP\_2148, SP\_1412, SP\_0372, SP\_1304, SP\_2002, SP\_0612, 5 SP\_1988, SP\_0484, SP\_0847, SP\_1527, SP\_0542, SP\_0441, SP\_0350, SP\_0014, SP\_1965, SP\_0117, and SP\_2108 were confirmed using this method.

### **Example 3. Antigen identification and pooled human screens**

CD4+ T cells and CD14+ monocytes were isolated from peripheral blood acquired from human donors. The monocytes were differentiated into dendritic cells 10 by culturing them in GM-CSF and IL-4 containing media, essentially as described in Tedder TF and Jansen PJ (1997 "Isolation and generation of human dendritic cells." *Current Protocols in Immunology Supp* 23: 7.32.1-7.32.16). After five days in culture, the dendritic cells were seeded into 384 well plates. The CD4+ T cells were expanded in culture to ensure sufficient quantities.

15 Each open reading frame predicted in the *S. pneumoniae* TIGR4 genome was cloned into an expression vector comprising a tag that is able to be presented by the major histocompatibility complex (MHC). Each construct was then expressed in *E. coli*, and full-length expression validated by a surrogate assay that identifies the tag in the context of MHC. In order to facilitate screening the large library, the library 20 was pooled such that four induced library clones were present in each well. In order to screen the human T cells, an aliquot of the pooled library was added to the seeded dendritic cells in 384-well plates. After 2 hr at 37°C, the dendritic cells were fixed with 1% paraformaldehyde for 15 min and washed extensively with phosphate buffer and lysine buffer. 40,000 of the CD4+ T cells in 70 µL of RP-10 media were 25 added to each well of a 384-well plate. The assay plates were incubated for 3 days at 37°C. The amount of IL-17 in the supernatant of each well was determined through the use of an IL-17 ELISA assay. In different iterations of the screen, the threshold for a positive result was set at two standard deviations above the mean of all samples, two standard deviations above the mean of negative controls, or 1.78 30 times the median absolute deviation of the data set. Positive pools were then deconvoluted as described in Example 4.

**Example 4. Deconvolution of the positive human pools**

For all antigens, deconvolution was performed by comparing the results of two pool screens. In this method, two different sets of pools were prepared, so that a polypeptide was with three different polypeptides between the first and second 5 pools. Consequently, it is possible to determine which polypeptides are antigens by identifying which polypeptides are in positive pools in both the first and second sets. In this deconvolution method, a pool was identified as positive if it was at least 1.78 times the median absolute deviation of the data set.

An antigen was identified as a positive hit if it was positive in at least two 10 repeated secondary screens. The antigens SP2108, SP0641, SP1393, SP0024, SP0641.1, SP1072, SP1384 and SP2032 were identified using the above approach.

**Example 5****SP2108, SP0148 and SP1634 polypeptides**

15 The SP2108 polypeptide (SEQ ID NO: 9), SP0148 polypeptide (SEQ ID NO: 7) and SP1634 polypeptide (see Table 2) were formulated as vaccine compositions using 4 µg of the polypeptide in combination with 1 µg cholera toxin. For combinations, 4 µg of each polypeptide was used. The compositions were administered intranasally to C57BL/6 mice three times, one week apart. The 20 subjects were then allowed to rest for 3 weeks, and bled at that time for immunogenicity. For this assay, heparinized whole blood was collected from the retrograde orbital sinus. The total PBMC were stimulated with either killed whole cells (WCC) or a combination of the three polypeptides in round bottomed tubes for three days. The supernatants were then harvested and evaluated by ELISA for IL-17 25 levels. Cholera toxin alone (CT) or an unrelated antigen from HSV (003) were used as negative controls. Results are shown in FIGS. 1 and 2. The subjects were allowed to rest an additional 2 weeks, at which time they were challenged with intranasal administration of *S. pneumoniae*. The subjects were sacrificed a week later, and the number of colony-forming units (CFU) was counted from nasal 30 washes. Results are shown in FIG. 3.

**Example 6****SP0882 and SP0314 polypeptides**

This example used the same protocols as Example 5, except that only two doses of the vaccine composition were administered. In these experiments, the SP0882 polypeptide (SEQ ID NO: 2) and SP0314 polypeptides (see Table 2) were used in conjunction with the three polypeptides tested in Example 5. Results of the immunogenicity assay are shown in FIGS. 4 and 5. Results of the colonization assay are shown in FIG. 6.

10

**Example 7****SP1072, SP0641N, and SP0024 polypeptides**

This example used a protocol similar to that of Example 5, except that two doses of the vaccine composition were administered, one week apart. Four weeks after the last immunization, the mice were challenged intranasally with live type 6B *S. pneumoniae*. One week later the bacterial burden was assessed in each mouse by plating a nasal lavage on selective media and counting CFU. The CFU isolated from each mouse is plotted for each immunized cohort. The results are shown in FIG. 7. Statistically significant results are indicated in the figure (\* = p-value < 0.05).

20

**Example 8****SP0148, SP0314, SP0882, and SP2108 polypeptides tested in the BALB/c mouse**

To determine whether similar immune responses were seen across different mouse genotypes, several polypeptides were administered to BALB/c mice. Using a protocol similar to that of Example 5, the mice were immunized, challenged with *S. pneumoniae*, and the number of CFU was recorded. The results of this experiment are shown in FIG. 8.

## SEQUENCES

## Polypeptide Sequences

SEQ ID NO: 1  
SP0024

5 >gi|14971488|gb|AAK74215.1| conserved hypothetical protein  
*Streptococcus pneumoniae* TIGR4  
 MSYFEQFMQANQAYVALHGQLNLPLKPKTRVAITCMDSRLHVAQALGLALGDAHILRNAGGRVTEDM  
 IRSILVISQQQMGTREIVVLHHTDCGAQTFENEFPQEYLKEELGVDVSDQDFLPFQDIEESVREDMQLL  
 IESPLIPDDVIIISGAIYNNVDTGSMTVVEL

10 SEQ ID NO: 2  
SP0882

>gi|14972356|gb|AAK75009.1| conserved hypothetical protein  
*(Streptococcus pneumoniae* TIGR4)

15 MNQSYFYLKMKEHKLKVPTGKERRVRILLPKDYEKTDRSYPVVFHDGQNVFNSKESFIGHSWKII  
 PAIKRNPDISRMIIVVAIDNDGMGRMNEYAAWKFQESPPIPGQQFGGKGVEYAEFVMEVVKPFIDETYRT  
 KADCQHTAMIGSSLGGNITQFIGLEYQDQIGCLGVFSSANWLHQEAFNRYFECQKLSPDQRIFIFYVGT  
 EEADDTDKTLMDGNIKQAYIDSSLCKYYHDLIAGGVHLDNLVLKVQSGAIHSEIPWSENLPDCLRFFAE  
 KW

20 SEQ ID NO: 3  
SP0882N

MNQSYFYLKMKEHKLKVPTGKERRVRILLPKDYEKTDRSYPVVFHDGQNVFNSKESFIGHSWKII  
 PAIKRNPDISRMIIVVAIDNDGMGRMNEYAAWKFQESPPIPGQQFGGKGVEYAEFVMEVVKPFI

25 SEQ ID NO: 4  
SP0882 with exogenous leader

MSSKFMKSAAVLGTATLASLLLACMNQSYFYLKMKEHKLKVPTGKERRVRILLPKDYEKTDRSYP  
 VVFHDGQNVFNSKESFIGHSWKIIPAIKRNPDISRMIIVVAIDNDGMGRMNEYAAWKFQESPPIPGQQF  
 30 GGKGVEYAEFVMEVVKPFIDETYRTKADCQHTAMIGSSLGGNITQFIGLEYQDQIGCLGVFSSANWLH  
 QEAfnRYFECQKLSPDQRIFIFYVGTEEADDTDKTLMDGNIKQAYIDSSLCKYYHDLIAGGVHLDNLVLK  
 VQSGAIHSEIPWSENLPDCLRFFAEKW

35 SEQ ID NO: 5  
SP0882N with exogenous leader

MSSKFMKSAAVLGTATLASLLLACMNQSYFYLKMKEHKLKVPTGKERRVRILLPKDYEKTDRSYP  
 VVFHDGQNVFNSKESFIGHSWKIIPAIKRNPDISRMIIVVAIDNDGMGRMNEYAAWKFQESPPIPGQQF  
 GGKGVEYAEFVMEVVKPFI

40 SEQ ID NO: 6  
SP0148 lacking signal sequence

MCSGGAKKEGEAASKKEIIVATNGSPKPFIFYEENGELTGYIEEVVRAIFKDSDKYDVKEKTEWSGVF  
 AGLDADRYNMAVNNLSYTKERAEKYLYAAPIAQNPNVLvvKKDDSSIKSLLDDIGGKSTEVVQATTSAK  
 45 QLEAYNAEHTDNPTILNYTKADFQQIMVRLSDGQFDYKIFDKIGVETVIKNQGLDNLKIELPSDQQP  
 YVYPLLAQGQDELKSFVDKRIKELYKDGTLEKLSKQFFGDTYLPAEADIK

50 SEQ ID NO: 7  
SP0148 including signal sequence

MKKIVKYSSLAALALVAAGVLAACSGGAKKEGEAASKKEIIVATNGSPKPFIFYEENGELTGYIEEVVR  
 AIFKDSDKYDVKEKTEWSGVFAGLDADRYNMAVNNLSYTKERAEKYLYAAPIAQNPNVLvvKKDDSS  
 IKSLDDIGGKSTEVVQATTSAKQLEAYNAEHTDNPTILNYTKADFQQIMVRLSDGQFDYKIFDKIGVE

TVIKNQGLNDLKVIELPSDQQPYVYPLLAQGQDELKSFVDKRIKELYKDGTLEKLSKQFFGDTYLPAE  
ADIK

SEQ ID NO: 8

5 SP1072

>gi|14972547|gb|AAK75185.1| DNA primase Streptococcus pneumoniae TIGR4

MVDKQVIEEIKNNANIVEIGDVISLQKAGRNYLGLCPFHGEKTPSFNVVEDKQFYHCFGCRSGDVF  
KFIEEYQGVPFIEAVQILGQRVGIEVEKPLYSEQKSASPHQALYDMHEDAACKFYHAILMTTMGEEAR  
NYLYQRGLTDEVLKHFwigLAPPERNYLYQRLSDQYREEDLLDSGLFYLSANQFVDTFHNRIMFPLT  
NDQGKVIASFGRIWQKTDQSQTSKYKNSRSTAIFNKSYEYHMDRAKRSSGKASEIYLMEGFMDVIAAY  
RAGIENAVASMGTALSREHVEHLKRLTKLVLVYDGDKAGQAATLKALDEIGDMPVQIVSMPDNLDPD  
EYLQKNGPEDLAYLLTKTRISPIEFYIHQYKPENSENLQAQIEFLEKIAPIPLIVQEKSIAAQNSYIHL  
ADSLASFDTQIEQIVNESRQVQRQNRMEGISRPTITMPVTQLSAIMRAEHLLYRMMESPLVLD  
YRLREDFAFATPEFQVLYDLLGQYGNLPPAEVLAEQTEEVERAWYQVLAQDLPAEISPQELSEVEMTRN  
KALLNQDNMRRIKKVQEASHVGDDTDALEELERLISQKRRME

SEQ ID NO: 9

SP2108 including signal sequence

20 SP2108

>gi|14973620|gb|AAK76167.1| maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein (*Streptococcus pneumoniae* TIGR4)

MSSKFMKSAAVLGTATLASLLLACGSKTADKPADGSSEVKELTVYVDEGYKSYIEEVAKAYEKEAGVKVTLKTGDAKGGLDKLSDLNQSGNVPDVMMAPYDRVGSGLSGDQGLSEVKLSDGAKTDDTTKSLVTAANGKVYGAPAVIESLVMYYNKDLVKDAPKTFADELNLAQDSKYAFAGEDGKTTAFLADWTNFYTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIVGINYAKSWYEWPKGMDTEGAGNLIQTQFQEGKTAIIIDGPWKAQAFKDAKVNYGVATIPTLPNGKEYAAGGGKAWVIPQAVKNLEASQKFVDFLIVATEQQKVLYDKTNEIPANTEARSYAEGKNDLTTAVIKQFKNTQPLPNISQMSAVWDPAKNMLFDAVSGQKDRAFTAANDAVTLIKETIKOKFGE

SEQ ID NO: 10

### SP2108 lacking signal sequence

35 MCGSKTADKPADSGSSEVKELTVYDEGYKSYIEEVAKAYEKEAGVKVTLKTGALGGLDKSLDNQS  
GNVPDVMMAPYDRVGSGLSGQI SEVKLSDGAKTDDTKSLVTAANGKVYGAPAVIESLVMYYNKDLV  
KDAPKTFADELENLAKDSKYAFAGEDGKTTAFLADWTNFYYTYGLLAGNGAYVFGQNGKDAKDIGLAND  
GSIVGINYAKSWYEKWPKGMDTEGAGNLIQTQFQEGKAAIIDGPWKAQAFKDAKVNYGVATIPTLP  
40 NGKEYAAFGGGKAWVIPQAVKNLEASQKFVDFLVATEQQKVLYDKTNEIPANTEARSYAEGKNDELTT  
AVIKQFKNTQPLPNISQMSAVWDPAKNMFLDAVSGQDKAKTAANDAVTLIKETIKQKFG

SEQ ID NO: 11

SP0641M

45 MSGTSMATPIVAASTVLIRPKLKEMLERPVLKNLKDDKIDLTSLTKIALQNTARPMMDATSWKEKSQ  
YFASPRQQGAGLINVANALRNEVVATFKNTDSKGLVNSYGSISLKEIKGDKKYFTIKLHNTSNRPLTF  
KVSASAITTDSLTDRLKDETYKDEKSPDGKQIVPEIHKPEVKGANITFEHDTFTIGANSSFDLNAV  
NVGEAKNKNKFVESFIHFESVEEALNSNGKKINFQPSLSMPIMGFAGWNHEPILDKWAEEGSRS  
KTLGGYDDDGPKIPGTLNKGIGGEHGDKFNPAGVIQRKDKNTTSLDQNPPELFAFNNEGINAPSSS  
GSKIANIYPLDSNGNPQDAQLERGLTPSPIVLRSAAEGLISIVNTNKEGENQRDLKVISREHFIRGIL  
NSKSNDAKGIKSSKLKVWGLKWDGLIYNPRGREENAPESKDQNQDPATKIRGQFEPPIAEGQYFYKFKY  
RLTKDYPWQVSYIPVKIDNTAPKIVSVDFSNPEKIKLITKDTYHKVKDQYKNETLFARDQKEHPEKFD  
EIANEVWYAGAALVNEDGEVEKNLEVTYAGEGQGRNRKLDKGNTIYEIKGAGDLRGKIIIEVIALDGS  
SNFTKIHRIKFANOADEKGMSYYLVDPDODSSKYO

55 SEO ID NO: 12

SP0641

>gi|14972117|gb|AAK74791.1| serine protease, subtilase family  
[*Streptococcus pneumoniae* TIGR4]

MKKSTVLSLTAAVILAAYAPNEVVLA DTSS EDALNISDKEKVAENKEKHENIHSAMETSQDFKEKK  
 TAVIKEKEVVKNPVIDNNTSNEEAKIKEENSNKSQGDYTDASFVNKNTENPKKEDKVVYIAEFKDKES  
 GEKAIKELSSLKNTKVLYTYDRIFNGSAIETTPDNLDKIKQIEGIGISSVERAQKVQPMNNHARKEIGVE  
 EAIDYLKSINAPFGKNFGRGMVISNIDTGTDRHKAMRIDDDAKASMRFKEDLKGTDKNYWLSDKI  
 5 PHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIFS YKMSDA  
 GSGFAGDETMFHAIEDSIKHNV DVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASS  
 SSWDLVANNHLKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITT  
 EDGTLKPSKLKFVYIGKGQDQDLIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYN  
 10 RDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMINPDKKTEVKRNNKEDFKDKLEQYYPI  
 NSNKPNVGDEKEIDFKFAPDTDKELYKEDII VPAGSTSWGPRIDILLKPDVSAPGKNIKSTLN  
 VINGK STYGYMSGTSMATPIVAASTVLIRPKLKEMLERPVLKLNKGDDKIDLTSLT  
 KIALQNTARPMMDATSW KEKSQYFASPRQQGAGLINVANALRNEVVATFKNTDSKGLVNSYGSISLKEIKGD  
 KKYFTIKLHNTSN RPLTFKVSASAITTDSLTDRLKLDETYKDEKSPDGKQIVPEI  
 HPEVKVGANITFEHDFTIGANSSFD LNAVINVGEAKNKNKFVESFIHESVEE  
 MEALNSNGKKINFQPSLSMPLMGFAGNWNHEPILD  
 KWAWE 15 EGSRSKTLGGYDDDGGKPKIPGTLNKGIGGEHGIDKFNPAGV  
 IONRDKNTTSLDQNPELFAFNNEGIN APSSSGSKIANIYPLDSNGNPQDAQLERGL  
 TPSPLVLSAEEGLISIVNTNKEGENQRDLKVISREHF  
 IRGILNSKSNDAKGIKSSKLKVWGDLKWDGLIYNPRGREENAP  
 ESKDNQDPATKIRGQFEP  
 IAEGQYF YKFKYRLTKDYPQVSYIPVKIDNTAPKIVS  
 VDFSNPEKIKLITKDTYHKV  
 DQYKNETLFARDQEH PEKFDEIANEVWYAGAALVN  
 EDEGEVEKNLEV  
 TYAGEGGQGRNRKLDKGNTIYEIKGAGDLRGKII  
 EVI 20 ALDGSSNFTKIHRIKFANQADEKG  
 MISYLVDPDQDSSKYQKLGEIAESFK  
 FNKLN  
 GNGKEGSLKKDTTG  
 VEHHHQENEESIKEKSSFTIDRN  
 NI  
 STIRDFENKDLKKL  
 I  
 KKKFREV  
 DDFTSETG  
 KRMEE  
 DYK  
 DKG  
 NII  
 AYDDG  
 T  
 DLEYETE  
 KLD  
 EIK  
 SKIYGV  
 LSPSKD  
 GHFEIL  
 GKIS  
 NVSK  
 NAKV  
 YGN  
 NYSIE  
 IKAT  
 KY DFHS  
 K  
 T  
 M  
 FD  
 LY  
 AN  
 IND  
 IV  
 D  
 GLA  
 F  
 AG  
 D  
 M  
 R  
 L  
 F  
 V  
 K  
 D  
 N  
 D  
 Q  
 K  
 K  
 A  
 E  
 I  
 K  
 R  
 M  
 P  
 E  
 K  
 I  
 K  
 E  
 T  
 K  
 S  
 E  
 Y  
 P  
 Y  
 V  
 S  
 S  
 Y  
 G  
 N  
 V  
 I  
 E  
 L  
 G  
 E  
 D  
 L  
 S  
 K  
 N  
 K  
 P  
 D  
 N  
 L  
 T  
 K  
 M  
 E  
 S  
 G  
 K  
 I  
 Y  
 S  
 D  
 E  
 K  
 Q  
 Q  
 Y  
 L  
 K  
 D  
 N  
 I  
 I  
 L  
 R  
 K  
 G  
 Y  
 A  
 L  
 K  
 V  
 T  
 T  
 Y  
 N  
 P  
 G  
 K  
 T  
 D  
 M  
 L  
 E  
 G  
 N  
 G  
 V  
 S  
 K  
 E  
 D  
 25 IAKI  
 Q  
 K  
 A  
 P  
 N  
 L  
 R  
 A  
 L  
 S  
 E  
 T  
 I  
 Y  
 A  
 D  
 S  
 R  
 N  
 V  
 E  
 D  
 G  
 R  
 S  
 T  
 Q  
 S  
 V  
 L  
 M  
 A  
 L  
 D  
 G  
 F  
 N  
 I  
 I  
 R  
 Y  
 Q  
 V  
 F  
 T  
 F  
 K  
 M  
 N  
 D  
 K  
 G  
 E  
 A  
 I  
 D  
 K  
 D  
 G  
 N  
 L  
 V  
 T  
 D  
 S  
 S  
 K  
 L  
 V  
 L  
 F  
 G  
 K  
 D  
 D  
 K  
 E  
 Y  
 T  
 G  
 E  
 D  
 K  
 F  
 N  
 V  
 E  
 A  
 I  
 K  
 E  
 D  
 G  
 S  
 M  
 L  
 F  
 I  
 D  
 T  
 K  
 P  
 V  
 N  
 L  
 S  
 M  
 D  
 K  
 N  
 Y  
 F  
 N  
 P  
 S  
 K  
 S  
 N  
 K  
 I  
 Y  
 V  
 R  
 N  
 P  
 E  
 F  
 Y  
 I  
 R  
 G  
 K  
 I  
 S  
 D  
 K  
 G  
 G  
 F  
 N  
 W  
 E  
 L  
 R  
 V  
 N  
 E  
 S  
 V  
 V  
 D  
 N  
 Y  
 L  
 I  
 Y  
 G  
 D  
 L  
 H  
 I  
 D  
 N  
 T  
 R  
 D  
 F  
 N  
 I  
 K  
 L  
 N  
 V  
 K  
 D  
 G  
 D  
 I  
 M  
 D  
 W  
 G  
 M  
 K  
 D  
 Y  
 K  
 A  
 N  
 G  
 F  
 P  
 D  
 K  
 V  
 T  
 D  
 M  
 D  
 G  
 N  
 V  
 L  
 Q  
 T  
 G  
 Y  
 S  
 D  
 L  
 N  
 A  
 K  
 A  
 V  
 G  
 V  
 H  
 Y  
 Q  
 F  
 L  
 D  
 N  
 V  
 K  
 P  
 E  
 V  
 N  
 I  
 D  
 P  
 K  
 G  
 N  
 T  
 S  
 I  
 E  
 Y  
 A  
 D  
 G  
 K  
 S  
 V  
 V  
 F  
 N  
 I  
 N  
 D  
 K  
 R  
 N  
 N  
 G  
 F  
 D  
 G  
 E  
 I  
 Q  
 E  
 Q  
 H  
 I  
 Y  
 I  
 N  
 G  
 K  
 E  
 Y  
 T  
 S  
 F  
 N  
 D  
 I  
 K  
 Q  
 I  
 I  
 D  
 K  
 T  
 L  
 N  
 I  
 K  
 I  
 V  
 V  
 K  
 D  
 F  
 A  
 R  
 N  
 T  
 T  
 V  
 K  
 E  
 F  
 I  
 L  
 N  
 K  
 D  
 T  
 G  
 E  
 V  
 S  
 E  
 L  
 K  
 P  
 H  
 R  
 V  
 T  
 V  
 T  
 I  
 Q  
 N  
 G  
 30 KEM  
 S  
 S  
 T  
 I  
 V  
 S  
 E  
 E  
 D  
 F  
 I  
 L  
 P  
 V  
 Y  
 K  
 G  
 E  
 L  
 E  
 K  
 G  
 Y  
 Q  
 F  
 D  
 G  
 W  
 E  
 I  
 S  
 G  
 F  
 E  
 G  
 K  
 D  
 A  
 G  
 Y  
 V  
 I  
 N  
 L  
 S  
 K  
 D  
 T  
 F  
 I  
 K  
 P  
 V  
 F  
 K  
 I  
 E  
 E  
 K  
 K  
 E  
 E  
 NK  
 P  
 T  
 F  
 D  
 V  
 S  
 K  
 K  
 D  
 N  
 P  
 Q  
 V  
 N  
 H  
 S  
 Q  
 L  
 N  
 E  
 S  
 H  
 R  
 K  
 E  
 D  
 L  
 Q  
 R  
 E  
 E  
 H  
 S  
 Q  
 K  
 S  
 D  
 S  
 T  
 K  
 D  
 V  
 T  
 A  
 T  
 V  
 L  
 D  
 K  
 N  
 N  
 I  
 S  
 S  
 K  
 T  
 T  
 N  
 N  
 P  
 N  
 K  
 L  
 P  
 KT  
 GT  
 AS  
 GA  
 Q  
 T  
 L  
 L  
 A  
 A  
 G  
 I  
 M  
 F  
 I  
 V  
 G  
 I  
 F  
 L  
 G  
 L  
 K  
 K  
 K  
 N  
 Q  
 D

35 SEQ ID NO: 13  
 SP0641N  
 MVVLADTSS EDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVKNPVIDNNTS  
 EEA  
 KIKEENSNKSQGDYTDASFVNKNTENPKKEDKVVYIAEFKDKESGEKA  
 I  
 KELSSLKNTKVLYTYDR  
 40 IFNGSAIETTPDNLDKIKQIEGIGISSVERAQKVQPMNNHARKEIGVE  
 EAIDYLKSINAPFGKNFGRGM  
 VISNIDTGTDRHKAMRIDDDAKASMRFKEDLKGT  
 DKNYWLSDK  
 K  
 I  
 P  
 HAF  
 NY  
 YNGG  
 KIT  
 V  
 E  
 K  
 Y  
 D  
 G  
 R  
 D  
 Y  
 F  
 D  
 P  
 H  
 G  
 M  
 H  
 I  
 A  
 G  
 I  
 L  
 A  
 G  
 N  
 D  
 T  
 E  
 Q  
 D  
 I  
 K  
 N  
 F  
 N  
 G  
 I  
 D  
 G  
 I  
 A  
 P  
 N  
 A  
 Q  
 I  
 F  
 S  
 Y  
 K  
 M  
 Y  
 S  
 D  
 A  
 G  
 S  
 G  
 F  
 A  
 G  
 D  
 E  
 T  
 M  
 F  
 H  
 A  
 I  
 E  
 D  
 S  
 I  
 K  
 H  
 N  
 V  
 D  
 V  
 S  
 S  
 G  
 F  
 T  
 G  
 T  
 G  
 L  
 V  
 G  
 E  
 K  
 Y  
 W  
 Q  
 A  
 I  
 R  
 A  
 L  
 R  
 K  
 A  
 G  
 I  
 P  
 M  
 V  
 V  
 A  
 T  
 G  
 N  
 Y  
 A  
 T  
 S  
 A  
 S  
 S  
 S  
 W  
 D  
 L  
 V  
 A  
 N  
 H  
 L  
 K  
 M  
 T  
 D  
 T  
 G  
 N  
 V  
 R  
 T  
 A  
 A  
 H  
 E  
 D  
 A  
 I  
 A  
 V  
 A  
 S  
 A  
 K  
 N  
 Q  
 T  
 V  
 E  
 F  
 D  
 K  
 V  
 N  
 I  
 G  
 G  
 E  
 S  
 F  
 K  
 Y  
 R  
 N  
 I  
 G  
 A  
 F  
 F  
 D  
 K  
 S  
 K  
 I  
 T  
 T  
 N  
 E  
 D  
 G  
 T  
 K  
 A  
 P  
 S  
 K  
 L  
 K  
 F  
 V  
 Y  
 I  
 G  
 K  
 Q  
 D  
 Q  
 D

50 SEQ ID NO: 14  
 SP0882 consensus  
 MNQSYFYLKMKEHKLKVPTGKERRVRILLPKDYEKDTDRSY  
 P  
 V  
 V  
 Y  
 F  
 H  
 D  
 G  
 Q  
 N  
 V  
 F  
 N  
 S  
 K  
 E  
 S  
 F  
 I  
 Y  
 G  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K  
 I  
 I  
 P  
 A  
 I  
 K  
 R  
 N  
 P  
 D  
 I  
 S  
 R  
 M  
 I  
 V  
 V  
 A  
 I  
 D  
 N  
 D  
 G  
 M  
 G  
 R  
 M  
 N  
 E  
 Y  
 A  
 A  
 W  
 K  
 F  
 Q  
 E  
 S  
 I  
 P  
 G  
 Q  
 Q  
 F  
 G  
 G  
 K  
 V  
 E  
 Y  
 A  
 E  
 Y  
 H  
 S  
 W  
 K<br

EAFNRYFECQKLSPDQRIFIYVGTEEADTDKTLMDGNIKQAYIDSSLCKYYHDLIAGGVH  
 I H R

5 LDNLVLKVQSGAIHSEIPWSENLPDCLRFFAEKW

SEQ ID NO: 15  
 SP0882N consensus

10 MNQSYFYLKMKEHKLKVPYTGKERRVRILLPKDYEKDTDRSYPVVFHDGQNVFNSKESF  
 I Y

IGHSWKIIIPAIKRNPDISRMIVVAIDNDGMGRMNEYAAWKFQESPIPGQQFGGKGVEYAE  
 Y H E E

15 FVMEVVVKPFI

SEQ ID NO: 16

20 SP0882 consensus with exogenous leader

MSSKFMKSAAVLGTATLASLLLACMNQSYFYLKMKEHKLKVPYTGKERRVRILLPKDYE  
 T T V I

25 KDTDRSYPVVFHDGQNVFNSKESFIGHSWKIIIPAIKRNPDISRMIVVAIDNDGMGRMNE  
 Y Y H

YAAWKFQESPIPGQQFGGKGVEYAEFVMEVVVKPFI DETYRTKADCQHTAMIGSSLGGNIT  
 E E

30 QFIGLEYQDQIGCLGVFSSANWLHQEAFNRYFECQKLSPDQRIFIYVGTEEADTDKTL  
 EK I H

35 DGNIKQAYIDSSLCKYYHDLIAGGVHLDNLVLKVQSGAIHSEIPWSENLPDCLRFFAEKW  
 R

SEQ ID NO: 17

SP0882N consensus with exogenous leader

40 MSSKFMKSAAVLGTATLASLLLACMNQSYFYLKMKEHKLKVPYTGKERRVRILLPKDYE  
 T T V I

45 KDTDRSYPVVFHDGQNVFNSKESFIGHSWKIIIPAIKRNPDISRMIVVAIDNDGMGRMNE  
 Y Y H

YAAWKFQESPIPGQQFGGKGVEYAEFVMEVVVKPFI  
 E E

50

SEQ ID NO: 18

0148 consensus lacking signal sequence

55 MCSGGAKKEGEAASKKEIIVATNGSPKPFIYEENGELTGYEIEVVRAIFKDSDKYDVKE  
 Q S R N N X

KTEWSGVFLAGDADRYNMAVNNLSYTKERAEKYLYAAPIAQNPNVLVVKDDSSIKS LDD  
 I E

IGGKSTEVVQATTSAKQLEAYNAEHTDNPTILNYTKADLQQIMVRLSDGQFDYKIFDKIG  
F

5 VETVIKNQGLDNLKVIELPSDQQPYVYPLLAQGQDELKSFVDKRIKELYKDGTLKLSKQ  
Y S

FFGDTYLPAEADIK

10

SEQ ID NO: 19  
SP0148 consensus including signal sequence

15 MKKIVKYSSLAALALVAAGVLAACSGGAKKEGEAASKKEIIVATNGSPKPFYEENGELT  
G L Q S R N

GYEIEVVRAIFKDSDKYDVKFEKTEWSGVAGLDADRYNMAVNNLSYTKERAEKYLYAAP  
N X I

20

IAQNPNVLVVKDDSSIKSLLDIGGKSTEVVQATTSAKQLEAYNAEHTDNPTILNYTKAD  
E

LQQIMVRLSDGQFDYKIFDKIGVETVIKNQGLDNLKVIELPSDQQPYVYPLLAQGQDELK  
F Y S

25

SFVDKRIKELYKDGTLKLSKQFFGDTYLPAEADIK

30

SEQ ID NO: 20  
SP2108 consensus lacking signal sequence

MCGSKTADKPADSGSSEVKELTVYVDEGYKSYIEEVAKAYEKEAGVKVTLKTGDALGGLD  
A I

35

KLSLDNQSGNVPDVMMAPYDRVGSGLSDGQLSEVKLSDGAKTDDTTKSLVTAANGKVYGA  
I X T

PAVIESLVMYYNKDLVKDAPKTFADELENLAKDSKYAFAGEDGKTTAFLADWTNFYYTYGL  
A

40

LAGNGAYVFGQNGKDAKDIGLANDSIVGINYAKSWYEKWPKGMDTEGAGNLIQTQFQE  
G P A X H

45

GKTAIIIDGPWKAQAFKDAKVNYGVATIPTLPGKEYAAFGGGKAWVIPQAVKNLEASQK  
A

FVDFLVATEQQKVLYDKTNEIPANTEARSYAEGKNDELTTAVIKQFKNTQPLPNISQMSA  
S A S

50

VWDPAKNMLFDAVSGQKDAKTAANDAVTLIKETIKQKFGE

SEQ ID NO: 21

SP2108 consensus including signal sequence

55

MSSKFMKSAAVLGTATLASLLLACGSKTADKPADSGSSEVKELTVYVDEGYKSYIEEVA  
T T V A

KAYEKEAGVKVTLKTGDALGGLDKLSLDNQSGNVPDVMMAPYDRVGSGLSGQLEVKLS  
 I I X

5 DGAKTDDTKSLVTAANGKVYGAPAVIESIVMYYNKDLVKDAPKTFADLENLAKDSKYAF  
 T

AGEDGKTTAFLADWTNFYYTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIVGINYAKSWY  
 A G P A X

10 EKWPKGMDTEGAGNLIQTQFQEGKTAIIIDGPWKAQAFKDAKVNYGVATIPTLPNGKEY  
 H

15 AAFGGGKAWVIPQAVKNLEASQKFVDFLVATEQQKVLYDKTNEIPANTEARSYAEGKND  
 A S A

LTTAVIKQFKNTQPLPNISQMSAWDPAKNMLFDAVSGQKDAKTAANDAVTLIKETIKQK  
 S

20 FGE

SEQ ID NO: 22  
 SP1634  
 >gi|14973124|gb|AAK75714.1| hypothetical protein SP\_1634

25 Streptococcus pneumoniae TIGR4  
 MANIFDYLKVAYDSYYDPLNELDILTIEITYLSFDNLVSTLPQRLLDAPQVPRDPTMLTSKNRL  
 QLLDELAQHKRFKNCKLSHFINDIDPELQKQFAAMTYRVSLDTYLIVFRGTDSSIIIGWKEDFHLTYMK  
 EIPAQKHALRYLKNFFAHHPKQKVILAGHSKGGNLAIYAAASQIEQSLQNQITAVYTFDAPGLHQELTQ  
 TAGYQRIMDRSKIFIPQGSIIIGMMLEIPAHQIIIVQSTALGGIAQHDTFSWQIEDKHFVQLDKTNSDSQ  
 30 QVDTTFKEWVATVPDEELQLYFDLFFGTILDAGISSINDLASLKALEYIHHLFVQAQSLTPEERETLG  
 RLTQLLIDTRYQAWKNR

SEQ ID NO: 23  
 SP0314  
 >gi|14971788|gb|AAK74491.1| hyaluronidase Streptococcus pneumoniae  
 TIGR4

40 MQTGTKKLIVSLSLVLSGFLLNHMTIGAEETTTNTIQQSQKEVQYQQRDTKNLVENGDFGQTEDGS  
 SPWTGSKAQGWSAWVDQKNSADASTRVIEAKDGAITISSHEKLRAALHRMVPIEAKKKYKLRFKIKTD  
 NKIGIAKVRIEESGKDRLWNSATTSGTKDWQTIEADYSPTLDVDKIKLELFYETGTGTVSKDIEL  
 VEVADQLSEDSQTDKQLEEKIDLPIGKKHVFSLADYTYKVENPDVASVKNGILEPLKEGTTNVIVSKD  
 GKEVKKIPLKILASVKDAYTDRLDDWNGIIAGNQYYDSKNEQMAKLNQELEGKVADSLSSISSQADRT  
 YLWEKFNSYKTSANLTATYRKLEEMAKQVTNPSSRYQDETVRTVRSMEWMHKHVYNSEKSIVGNW  
 WDYEIGTPRAINNTLSLMKEYFSDEEIKKYTDVIEKFVDPDEHFRKTTDNPFKALGGNLVDMGRVKVI  
 45 AGLLRKDDQEISSTIRSIEQVFKLVQDGEGFYQDGSYIDHTNVAYTGYGNVLIDGLSQLLPVIQKTK  
 NPIDKDKMQTMYHWIDKSFAPLLVNGELMDMSRGRSISRANSEGHVAAVEVIRGIHRIADMSEGETKQ  
 CLQSLVKTIVQSDSYYDVFKNLKYKDISLMSLSDAGVASVPRPSYLSAFNMDKTAMYNAEKGFG  
 FGLSLFSSRTINYEHMNKENKRGWYTSQGMFYLYNGDLHYSDGYWPTVNPYKMPGTETDAKRADSD  
 TGKVLPSAFVGTSKLDDANATATMDFTNWNQTLTAHKSWFMLKDKIAFLGSNIQNTSTDAAATTIDQR  
 KLESGNPYKVVNDKEASLTEQEKDYPETQSVFLESFDKKNIGYFFFKKSSISMSKALQKGAWKDIN  
 50 EGQSDKEVENEFLETLISQAHKQNRDSYGYMLIPNVDRATFNQMIKELESSLIENNETLQSVYDAKQGVW  
 GIVKYDDSVSTISNQFQVLKRGVYTIRKEDEYKIAYYNPETQESAPDQEVFKKLEQAAQPQVQNSKE  
 KEKSEEKNHSDQKNLPQTGEGQSIASLGFLLLGAFYLFRRGKNN

## Nucleic Acid Sequences

SEQ ID NO: 24  
 SP0882N DNA

5 ATGAAATCAATCCTACTTTATCTAAAAATGAAAGAACACAAACTCAAGGTTCTTATACAGGTAAGGA  
 GCGCCGTGTACGTATTCTTCTTAAAGATTATGAGAAAGATAACAGACCGTTCTATCCTGTTGAT  
 ACTTTATGACGGGCAAATGTTTAATAGCAAAGAGTCTTCATTGGACATTATGGAAAGATTATC  
 CCAGCTATCAAACGAAATCCGGATATCAGTCGATGATTGTCGTTGCTATTGACAATGATGGTATGGG  
 10 GCGGATGAATGAGTATGCGGTTGGAAGTTCCAAGAATCTCCTATCCCAGGGCAGCAGTTGGTGGTA  
 AGGGTGTGGAGTATGCTGAGTTGTCATGGAGGTGGTCAAGCCTTTATC

SEQ ID NO: 25  
 SP0882 with exogenous leader

15 ATGTCATCTAAATTTATGAAGAGCGCTGCGGTGCTTGGAACTGCTACACTTGCTAGCTTGCTTTGGT  
 AGCTTGCATGAATCAATCCTACTTTATCTAAAAATGAAAGAACACAAACTCAAGGTTCTTATACAG  
 20 GTAAGGAGGCCGTGTACGTATTCTTCTTAAAGATTATGAGAAAGATAACAGACCGTTCTATCCT  
 GTTGTATACTTCATGACGGGCAAATGTTTAATAGCAAAGAGTCTTCATTGGACATTATGGAA  
 GATTATCCCAGCTATCAAACGAAATCCGGATATCAGTCGATGATTGTCGTTGCTATTGACAATGATG  
 GTATGGGGCGGATGAATGAGTATGCGGTTGGAAGTTCCAAGAATCTCCTATCCCAGGGCAGCAGTT  
 25 GGTGGTAAGGGTGTGGAGTATGCTGAGTTGTCATGGAGGTGGTCAAGCCTTTATCGATGAGACCTA  
 TCGTACAAAAGCAGACTGCCAGCATACGGCTATGATTGGTTCTACTAGGAGGCAATTACCCAGT  
 TTATCGGTTTGAATACCAAGACCAAATTGGTTGCTGGGCGTTTTCATCGCAAATGGCTCCAC  
 CAAGAACGCTTAAACCGCTATTCGAGTGCCAGAAACTATCGCTGACCAGCGCATCTTCATCTATGT  
 AGGAACAGAACAGAACGAGATGATAACAGACAAGACCTTGATGGATGGCAATATCAAACAGCCTATATCG  
 ACTCGTCGCTTGCTATTACCATGATTGATAGCAGGGGGAGTACATCTGGATAATCTTGTGCTAAAA  
 30 GTTCAGTCTGGTGCCATCCATAGTGAAATCCCTGGTCAGAAAATCTACCAGATTGTCAGATT  
 TGCAAGAAAATGGTAA

SEQ ID NO: 26  
 SP0882N with exogenous leader

35 ATGTCATCTAAATTTATGAAGAGCGCTGCGGTGCTTGGAACTGCTACACTTGCTAGCTTGCTTTGGT  
 AGCTTGCATGAATCAATCCTACTTTATCTAAAAATGAAAGAACACAAACTCAAGGTTCTTATACAG  
 GTAAGGAGGCCGTGTACGTATTCTTCTTAAAGATTATGAGAAAGATAACAGACCGTTCTATCCT  
 GTTGTATACTTCATGACGGGCAAATGTTTAATAGCAAAGAGTCTTCATTGGACATTATGGAA  
 GATTATCCCAGCTATCAAACGAAATCCGGATATCAGTCGATGATTGTCGTTGCTATTGACAATGATG  
 GTATGGGGCGGATGAATGAGTATGCGGTTGGAAGTTCCAAGAATCTCCTATCCCAGGGCAGCAGTT  
 40 GGTGGTAAGGGTGTGGAGTATGCTGAGTTGTCATGGAGGTGGTCAAGCCTTTATC

SEQ ID NO: 27  
 SP0148 lacking signal sequence

45 ATGTGCTCAGGGGGTCTAAGAAAAGAAGGAGAACAGCTAGCAAGAAAGAAATCATCGTGCACCAA  
 TGGATCACCAAAGCCATTATCTATGAAGAAAATGGCAATTGACTGGTTACGAGATTGAAGTCGTT  
 GCGCTATCTTAAAGATTCTGACAAATATGATGTCAAGTTGAAAAGACAGAACGGTCAAGGTGTCTT  
 GCTGGTCTTGACGCTGATCGTTACAATATGGCTGTCAACAATCTTAGCTACACTAAAGAACGTGCGGA  
 GAAATACCTCATGCCGACCAATTGCCAAAATCTAATGTCCTTGCTGAAGAAAGATGACTCTA  
 50 GTATCAAGTCCTCGATGATATGGTGGAAAATCGACGGAAGTCGTTCAAGGCCACTACATCAGCTAAG  
 CAGTTAGAACATACATGCTGAACACACGGACAACCCAACATCCTTAACATATACTAAGGCAGACTT  
 CCAACAAATCATGGTACGTTGAGCGATGGACAATTGACTATAAGATTTGATAAAATGGTGT  
 AAACAGTGATCAAGAACCAAGGTTGGACAATTGAAAGTTATCGAACTTCCAAGCGACCAACACCG  
 TACGTTACCCACTTCTGCTCAGGGTCAAGATGAGTTGAAATCGTTGAGACAAACGCATCAAAGA  
 ACTTATAAAAGATGAACTCTGAAAAATTGCTAAACAATTCTGGAGACACTTATCACGGCAG  
 AAGCTGATATAAATAA

SEQ ID NO: 28  
 SP0148 including signal sequence

55 ATGAAAAAAATCGTAAATACTCATCTCTGCGAGCCCTGCTCTGTTGCTGCAGGTGTGCTTGC  
 TTGCTCAGGGGGTCTAAGAAAAGAAGGAGAACAGCTAGCAAGAAAGAAATCATCGTTGCAACCAATG  
 GATCACCAAAGCCATTATCTATGAAGAAAATGGCAATTGACTGGTTACGAGATTGAAGTCGTT  
 GCTATCTTAAAGATTCTGACAAATATGATGTCAAGTTGAAAGACAGAACGGTCAAGGTGTCTTGC

5 TGGTCTTGACGCTGATCGTTACAATATGGCTGTCAACAACTCTAGCTACACTAAAGAACGTGCGGAGA  
AATAACCTCTATGCCGACCAATTGCCAAATCTTAAATGCTTGTGAGAAGAGATGACTCTAGT  
ATCAAGTCTCTCGATGATATCGGGAAAATCGACGGAAGTCGTCAAGCCACTACATCAGCTAAGCA  
GTTAGAACGATACAATGCTGAACACACGGACAACCCAACTATCCTTAACCTACTAAAGGCAGACTTCC  
ACAAATCATGGTACGTTGAGCGATGGACAATTGACTATAAGATTTGATAAAATCGGTGTTGAA  
ACAGTGTCAAGAACCAAGGTTGGACAACCTGAAAGTTATCGAACTTCCAAGCGACCAACAAACCGTA  
CGTTACCCACTTCTGCTAGGGTCAAGATGAGTTGAAATCGTTGAGACAAACGCATCAAAGAAC  
TTTATAAAGATGGAACCTTGAAACAAATTGTCTAAACATTCTCGGAGACACTTATCACCGGCAGAA  
GCTGATATTAAATAA  
10 SEQ ID NO: 29  
SP2108 lacking signal sequence  
ATGTGCGGAAGCAAAACTGCTGATAAGCCTGCTGATTCTGGTTCATCTGAAGTCAAAGAACTCACTGT  
ATATGTAGACGAGGGATATAAGAGCTATATTGAAGAGGTTGCTAAAGCTTATGAAAAAGAGCTGGAG  
15 TAAAAGTCACTCTTAAACTGGTGTAGCTAGGAGGTCTTGATAAAACTTCTTGTACAACCAAACT  
GGTAATGTCCTGTATGTTATGATGGCTCCATACGACCGTGTAGGTAGCCTTGGTTCTGACGGACA  
TTCAGAAGTGAATTGAGCGATGGTGTAAACAGACGACACAACAAATCTTGTAAACAGCTGCTA  
ATGGTAAAGTTACGGTGTCCCTGCCATTGAGTCACTTGTTATGTACTACAACAAAGACTTGGTG  
AAAGATGCTCCAAAACATTGCTGACTTGGAAACCTTGCTAAAGATAGCAAATACGCATTGCTGG  
20 TGAAGATGGTAAACTACTGCCTCCTAGCTGACTGGACAAACTCTACTATACATATGGACTTCTTG  
CCGTAACGGTGCTACGTCTTGGCCAAACGGTAAAGACGCTAAAGACATCGGTCTGCAAACGAC  
GGTTCTATCGTAGGTATCAACTACGCTAAATCTGGTACGAAAATGGCTAAAGGTATGCAAGATA  
AGAAGGTGCTGGAAACTTAATCCAAACTCAATTCAAAGAGTAAACAGCTGCTATATCGACGGAC  
25 CTTGGAAAGCTCAAGCCTTAAAGATGCTAAAGTAAACTACGGAGTTGCAACTATCCCACCTCTTCCA  
AATGGAAAAGAATATGCTGATTGGTGGTAAAGCTTGGTCAATTCTCAAGCCGTTAAGAACCT  
TGAAGCTTCTCAAAAATTGAGACTTCTTGTGCAACTGAACAAACAAAGTATTATATGATAAGA  
CTAACGAAATCCCAGCTAAACTGAGGCTCGTTCATACGCTGAAGGTAACAGATGAGTTGACA  
30 GCTGTTATCAAACAGTCAGAACACTCAACCAACTGCCAAACATCTCTCAAATGTCTGCAGTTGGGA  
TCCAGCGAAAATATGCTTTGATGCTGTAAGTGGTCAAAAAGATGCTAAACAGCTGCTAACGATG  
CTGTAACATTGATCAAAGAACATCAAACAAAATTGGTGAATAA  
SEQ ID NO: 30  
SP0641M  
35 ATGTCAGGAACACTGATGGCGACTCCAATCGTGGCAGCTTCTACTGTTTGATTAGACCGAAATTAAA  
GGAAATGCTGAAAGACCTGTATTGAAAATCTTAAAGGGAGATGACAAAATAGATCTTACAAGTCTTAA  
CAAAAATTGCCCTACAAAATACTGCGCAGCTATGATGGATGCAACTTCTTGGAAAGAAAAAGTC  
40 TACTTGTGATCACCTAGACAACAGGGAGCAGGCTTAATTATGTGGCAATGCTTGGAGAAATGAAGT  
TGTAGCAACTTCAAAAACACTGATTCTAAAGGTTGGTAAACTCATATGGTCCATTCTTCTTAAAG  
AAATAAAAGGTATAAAAATACTTACAATCAAGCTTCAAATACATCAAACAGACCTTGTACTTT  
AAAGTTTCAGCATCAGCGATAACTACAGATTCTTAACGACAGATTAAACTTGTGAAACATATAA  
AGATGAAAATCTCCAGATGGTAAAGCAAATTGTCAGGAAATTCAACCGAGAAAAGTC  
45 ATATCACATTGAGCATGATACTTCACTATAGGCGCAAATTCTAGCTTGTGATTGAAATGCGGTATA  
AAATGTTGGAGAGGCCAAAACAAAATAATTGAGATCATTATTGAGTCAGTGGAAAGA  
AATGGAAGCTCTAAACTCCAACGGGAAGAAAATAACTTCAAACCTTCTTGTGATGCCCTAAATGG  
GATTGCTGGAAATTGGAACCACGAACCAATCCTGATAAATGGCTGGGAAGAAGGGTCAAGATCA  
50 AAAACACTGGGAGGTTATGATGATGGTAAACCGAAAATTCCAGGAACCTTAAATAAGGGAAATTGG  
TGGAGAACATGGTATAGATAAATTAAATCCAGCAGGAGTTACAAAATAGAAAAGATAAAAATACAA  
CATCCCTGGATCAAATCCAGAATTATTGCTTCAATAACGAAGGGATCAACGCTCCATCATCAAGT  
GGTCTAAGATTGCTAACATTATCCTTGTAGATTCTAAATGGAAATCCTCAAGATGCTCAACTTGAAG  
AGGATTAACACCTTCTCCACTTGTATTAGAAGTGCAGAAGAAGGATTGATTCAATAGTAAATACAA  
55 ATAAAGAGGGAGAAAATCAAAGAGACTTAAAGTCATTGAGAGAACACTTATTAGAGGAATTG  
AATTCTAAAGCAATGCAAAGGGAAATCAAATCATCTAAACTAAAAGTTGGGTGACTGAAAGTG  
GGATGGACTCATCTATAATCCTAGAGGGTAGAGAAGAAAATGCACCGAGAAAAGTAAGGATAATCAAGATC  
CTGCTACTAAGATAAGAGGTCAATTGAAACCGATTGCGGAAGGTCAATTCTTAAATTAAAT  
AGATTAACATAAGATTACCCATGGCAGGTTCCATTATTCTGTAAAATTGATAACACCGCCCTAA  
GATTGTTCGGTTGATTGTTCAAATCCTGAAAAAATTAGTTGATTACAAAGGATACTTATCATAAG  
TAAAAGATCAGTATAAGAATGAAACGCTATTGCGAGAGATCAAAAGAACATCCTGAAAAATTGAC  
GAGATTGCGAACGAGTTGGTATGCTGGCGCCGCTTGTAAATGAAGATGGAGAGGGTTGAAAAAAA

TCTTGAAGTAACCTACGCAGGTGAGGGTCAAGGAAGAAATAGAAAACCTGATAAAGACGGAAATACCA  
TTTATGAAATTAAAGGTGCGGGAGATTAAAGGGGAAAATCATTGAAGTCATTGCATTAGATGGTCT  
AGCAATTTCACAAAGATTCACTAGAATTAAATTGCTAATCAGGCTGATGAAAAGGGGATGATTCCTA  
TTATCTAGTAGATCCTGATCAAGATTCACTAAATATCAA

5

SEQ ID NO: 31

SP0641N

ATGGTAGTCTTACGCAGACACATCTAGCTCTGAAGATGCTTAAACATCTCTGATAAAGAAAAAGTAGC

AGAAAATAAAGAGAAACATGAAAATATCCATAGTCTATGGAAACTTCACAGGATTTAAAGAGAAGA

10

AAACAGCAGTCATTAAGGAAAAGAAGTTGTTAGTAAAATCCTGTGATAGACAATAACACTAGCAAT

GAAGAAGCAAAATCAAAGAAGAAAATTCCAATAATCCCAAGGAGATTACGGACTCATTTGTGAA

TAAAAACACAGAAAATCCCAAAAAGAAGATAAAGTTGCTATATTGCTGAATTAAAGATAAAGAAT

CTGGAGAAAAGCAATCAAGGAACATATCCAGTCTTAAGAATACAAAAGTTTATATACTTATGATAGA

ATTTTAAACGGTAGTGCATAGAAACAACCTCCAGATAACTTGGACAAAATTAAACAAATAGAAGGTAT

15

TTCATCGGTTGAAAGGGCACAAAAGTCCAACCCATGATGAATCATGCCAGAAAGGAAATTGGAGTTG

AGGAAGCTATTGATTACCTAAAGTCTATCAATGCTCCGTTGGGAAAATTGATGGTAGAGGTATG

GTCATTCAAAATATCGATACTGGAACAGATTATAGACATAAGGCTATGAGAATCGATGATGATGCCAA

AGCCTCAATGAGATTTAAAAAGAAGACTTAAAGGCACTGATAAAAATTATTGGTTGAGTGATAAAA

TCCCTCATGCGTTCAATTATTATAATGGTGGCAAAACTACTGTAGAAAATATGATGATGGAAGGGAT

TATTTGACCCACATGGGATGCATATTGCAAGGGATTCTGCTGGAAATGATACTGAACAAGACATCAA

AAACCTTAAACGGCATAGATGGAATTGCACCTAATGCACAAATTCTCTTACAAAATGTATTCTGACG

CAGGAICTGGGTTGCGGGTGTGAAACAATGTTCATGCTATTGAAGATTCTATCAAACACAACGTT

GATGTTGTTCGGTATCATCTGGTTTACAGGAACAGGTCTTGTAGGTGAGAAAATTGGCAAGCTAT

TCGGGCATTAAAGAAAAGCAGGCATCCTAACATGGTGTGCTACGGTAACTATGCGACTCTGCTTCAA

GTCAGCACATGAAGATGCGATAGCGGTGCTCTGCTAAAATCAAACAGTTGAGTTGATAAAAGTAA

CATAGGTTGGAGAAAGTTAAATACAGAAATATAGGGCCTTTGATAAGAGTAAATCACACAA

ATGAAGATGGAACAAAAGCTCTAGTAAATTAAATTGTTATATAGGCAAGGGGCAAGACCAAGAT

20

TTGATAGGTTGGATCTAGGGCAAAATTGCACTGAGTAATGGATAGAATTATACAAAGGATTAAAAAA

TGCTTTAAAAAGCTATGGATAAGGGTGCACGCCATTATGGTTGTAATACTGTAATTACTACA

ATAGAGATAATTGGACAGAGCTTCCAGCTATGGGATATGAAGCGGATGAAGGTACTAAAAGTCAGTG

TTTCAATTTCAGGAGATGATGGTAAAGCTATGGAACATGATTAATCCTGATAAAAAAAACTGAAGT

CAAAGAAATAATAAAGAAGATTAAAGATAAATTGGAGCAATACTATCCAATTGATATGGAAAGTT

TTAATTCCAACAAACCGAATGAGGTGACAAAAAGAGATTGACTTTAAGTTGCACCTGACACAGAC

25

AAAGAACTCTATAAAGAAGATATCATCGTCCAGCAGGATCTACATCTGGGGGCCAAGAATAGATT

ACTTTAAACCCGATGTTCAAGCACCTGGTAAAATTAAATCCACGCTTAATGTTATTAATGGCA

AATCAACTTATGGC

SEQ ID NO: 32

HHHHHH

SEQ ID NO: 33

MSYYHHHHHH

## CLAIMS

We claim:

1. A vaccine formulation comprising a pharmaceutically acceptable carrier and one or more polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-11 or an immunogenic fragment thereof, and optionally further comprising a polypeptide having an amino acid sequence comprising either of SEQ ID NOS: 12 or 13 or an immunogenic fragment thereof.
2. The vaccine formulation of claim 1, wherein the vaccine formulation comprises at least two different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13 or an immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising one of SEQ ID NOS: 1-10 or an immunogenic fragment thereof.
3. The vaccine formulation of claim 2, which comprises at least two polypeptides, each polypeptide belonging to a different group of (i)-(vi):
  - 15 (i) SEQ ID NO: 1 or an immunogenic fragment thereof,
  - (ii) one of SEQ ID NOS: 2-5 or an immunogenic fragment thereof,
  - (iii) one of SEQ ID NOS: 6-7 or an immunogenic fragment thereof,
  - (iv) SEQ ID NO: 8 or an immunogenic fragment thereof,
  - (v) one of SEQ ID NOS: 9-10 or an immunogenic fragment thereof, and
  - 20 (vi) one of SEQ ID NO: 11-13 or an immunogenic fragment thereof.
4. The vaccine formulation of claim 1, wherein the vaccine formulation comprises at least three different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13 or an immunogenic fragment thereof, wherein at least one of said polypeptides has an amino acid sequence comprising one of SEQ ID NOS: 1-10.

5. The vaccine formulation of claim 4, which comprises at least three polypeptides, each polypeptide belonging to a different group of (i)-(vi):

(i) SEQ ID NO: 1 or an immunogenic fragment thereof,

(ii) one of SEQ ID NOS: 2-5 or an immunogenic fragment thereof,

5 (iii) one of SEQ ID NOS: 6-7 or an immunogenic fragment thereof,

(iv) SEQ ID NO: 8 or an immunogenic fragment thereof,

(v) one of SEQ ID NOS: 9-10 or an immunogenic fragment thereof, and

(vi) one of SEQ ID NO: 11-13 or an immunogenic fragment thereof.

6. The vaccine formulation of any of claims 1-5, wherein the fragment is a 10 truncated fragment of any of SEQ ID NOS: 1-13 having from 1-20 amino acid residues removed from the N-terminus, C-terminus, or both.

7. The vaccine formulation of claim 1, wherein the vaccine formulation comprises one or more polypeptides having an amino acid sequence consisting of any of SEQ ID NOS: 1-11.

15 8. The vaccine formulation of claim 1, which comprises a polypeptide having an amino acid sequence comprising SEQ ID NO: 6.

9. The vaccine formulation of claim 1, which comprises a polypeptide having an amino acid sequence comprising SEQ ID NO: 7.

10. The vaccine formulation of claim 1, which comprises a polypeptide having 20 an amino acid sequence comprising SEQ ID NO: 9.

11. The vaccine formulation of claim 1, which comprises a polypeptide having an amino acid sequence comprising SEQ ID NO: 10.

12. The vaccine formulation of claim 1, wherein the vaccine formulation comprises a polypeptide consisting of SEQ ID NO: 6 and a polypeptide consisting of SEQ ID NO: 9.
13. The vaccine formulation of claim 1, wherein the vaccine formulation comprises a polypeptide consisting of SEQ ID NO: 7 and a polypeptide consisting of SEQ ID NO: 10.
14. The vaccine formulation of any of claims 1-13, which contains substantially no other *S. pneumoniae* polypeptides other than polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 1-13.
- 10 15. The vaccine formulation of claim 1, which comprises a pharmaceutically acceptable carrier and one or more polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 2, 7, 9, 22, and 23 or an immunogenic fragment thereof.
16. A vaccine formulation comprising a pharmaceutically acceptable carrier and 15 a polypeptide having an amino acid sequence consisting of SEQ ID NO: 11 or an immunogenic fragment thereof.
17. A vaccine formulation comprising a pharmaceutically acceptable carrier and a polypeptide having an amino acid sequence comprising SEQ ID NO: 12.
18. A vaccine formulation comprising a pharmaceutically acceptable carrier and 20 one or more polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 14-21 or an immunogenic fragment thereof.
19. The vaccine formulation of claim 18, wherein the vaccine formulation comprises at least two different polypeptides having an amino acid sequence comprising any of SEQ ID NOS: 14-21 or an immunogenic fragment thereof.
- 25 20. The vaccine formulation of claim 19, which comprises at least two polypeptides, each polypeptide belonging to a different group of (i)-(iii):
  - (i) one of SEQ ID NOS: 14-17 or an immunogenic fragment thereof,

(ii) one of SEQ ID NOS: 18-19 or an immunogenic fragment thereof; and

(iii) one of SEQ ID NOS: 20-21 or an immunogenic fragment thereof.

21. The vaccine formulation of claim 18, wherein the vaccine formulation further comprises a polypeptide having an amino acid sequence comprising any of 5 SEQ ID NOS: 1-13.

22. The vaccine formulation of claim 18, wherein the fragment is a truncated fragment of any of SEQ ID NOS: 14-21 wherein from 1-20 amino acid residues are removed from the N-terminus, C-terminus, or both.

23. The vaccine formulation of claim 18, which comprises a polypeptide having 10 an amino acid sequence comprising any of SEQ ID NOS: 14-17.

24. The vaccine formulation of claim 18, which comprises a polypeptide having an amino acid sequence comprising either of SEQ ID NOS: 18-19.

25. The vaccine formulation of claim 18, which comprises a polypeptide having an amino acid sequence comprising either of SEQ ID NOS: 20-21.

15 26. The vaccine formulation of any of claims 1-25, wherein the polypeptide is conjugated to an immunogenic carrier.

27. The vaccine formulation of any of claims 1-25, which comprises at least one lipidated polypeptide.

28. The vaccine formulation of any of claims 1-27, further comprising an 20 adjuvant.

29. The vaccine formulation of claim 28, wherein the adjuvant is an agonist of toll-like receptors (TLRs).

30. The vaccine formulation of claim 28, wherein the adjuvant is alum.

31. The vaccine formulation of claim 28, wherein the vaccine formulation 25 comprises 1-1000 µg of the polypeptide and 1-250 µg of the adjuvant.

32. The vaccine formulation of any of claims 1-31, which induces a T<sub>H</sub>17 cell response at least 1.5-fold after contacting T<sub>H</sub>17 cells.
33. The vaccine formulation of any of claims 1-31, wherein the vaccine formulation inhibits infection by *S. pneumoniae* in an uninfected subject.
- 5 34. The vaccine formulation of any of claims 1-31, wherein the vaccine formulation inhibits *S. pneumoniae* colonization in an individual.
35. The vaccine formulation of any of claims 1-31, wherein the vaccine formulation inhibits *S. pneumoniae* symptoms.
- 10 36. A method for treating a subject suffering from or susceptible to *S. pneumoniae* infection, comprising administering an effective amount of a vaccine formulation according to any of claims 1-45.
37. The method of claim 36, wherein the method inhibits infection by *S. pneumoniae* in an uninfected subject.
- 15 38. The method of claim 36, wherein the method inhibits *S. pneumoniae* colonization in an individual.
39. The method of claim 36, wherein the method inhibits *S. pneumoniae* symptoms.
40. The method of claim 36, wherein the method treats a subject with one dose.
41. The method of claim 36, wherein the method treats a subject within three  
20 doses.
42. The method of claim 36, wherein the subject is a human.
43. An immunogenic composition comprising a pharmaceutically acceptable carrier and two or more polypeptides having amino acid sequences comprising any of SEQ ID NOS: 1-23 and SP1574, SP1655, SP2106, SP1473, SP0605, SP1177,  
25 SP0335, SP0906, SP1828, SP2157, SP1229, SP1128, SP1836, SP1865, SP0904, SP0765, SP1634, SP0418, SP1923, SP1313, SP0775, SP0314, SP0912, SP0159,

SP0910, SP2148, SP1412, SP0372, SP1304, SP2002, SP0612, SP1988, SP0484, SP0847, SP1527, SP0542, SP0441, SP0350, SP0014, SP1965, SP0117, SP0981, SP2229, SP2136, SP1179, SP1174, SP2216, SP1393, SP0641.1, SP1384, and SP2032, or an immunogenic fragment thereof.