

## FUSION PROTEINS AND COMBINATION VACCINES COMPRISING HAEMOPHILUS INFLUENZAE PROTEIN E AND PILIN A

This application claims priority to United States patent application number 61/474779 filed April 13, 2011 and United States patent application number 61/534012 filed September 13, 2011.

### FIELD OF THE INVENTION

[01] The present invention relates to compositions comprising *Haemophilus influenzae* (*H. influenzae*) Protein E and Pilin A. More particularly, the present application relates to fusion proteins and immunogenic compositions comprising Protein E and Pilin A, vaccines comprising such immunogenic compositions and therapeutic uses of the same.

### BACKGROUND OF THE INVENTION

[02] Protein E (PE) is an outer membrane lipoprotein with adhesive properties. It plays a role in the adhesion/invasion of non-typeable *Haemophilus influenzae* (NTHi) to epithelial cells. (J. Immunology 183: 2593-2601 (2009); The Journal of Infectious Diseases 199:522-531 (2009), Microbes and Infection 10:87-96 (2008)). It is highly conserved in both encapsulated *Haemophilus influenzae* and non-typeable *H. influenzae* and has a conserved epithelial binding domain. (The Journal of Infectious Diseases 201:414-419 (2010)). Thirteen different point mutations have been described in different *Haemophilus* species when compared with *Haemophilus influenzae* Rd as a reference strain. Its expression is observed on both logarithmic growing and stationary phase bacteria. (WO2007/084053).

[03] Protein E is also involved in human complement resistance through binding vitronectin. (Immunology 183: 2593-2601 (2009)). PE, by the binding domain PKRYARSVRQ YKILNCANYH LTQVR (SEQ ID NO. 1, corresponding to amino acids 84-108 of SEQ ID NO. 4), binds vitronectin which is an important inhibitor of the terminal complement pathway. (J. Immunology 183:2593-2601 (2009)).

[04] Pilin A (PilA) is likely the major pilin subunit of *H. influenzae* Type IV Pilus (Tfp) involved in twitching motility (Infection and Immunity, 73: 1635-1643 (2005)). NTHi PilA is a conserved adhesin expressed *in vivo*. It has been shown to be involved in NTHi adherence, colonization and biofilm formation. (Molecular Microbiology 65: 1288-1299 (2007)).

[05] Non-typeable *Haemophilus influenzae* is an important and common respiratory pathogen that causes otitis media in infants and children. NTHi is, after *Streptococcus pneumoniae*, the most common cause of acute otitis media in children (J. Immunology 183: 2593-2601 (2009), Pediatrics 113:1451-1465 (2004)). It is an important cause of sinusitis in children and adults. (Current Infectious Disease Reports 11:177-182 (2009)). It has been associated with increased risk of exacerbations in chronic obstructive pulmonary disease (COPD) in adults. (Journal of Chronic Obstructive Pulmonary Disease 3:109-115 (2006)). In addition, non-typeable *H. influenzae* causes community-acquired pneumonia in adults and may cause pneumonia in children in developing countries. (Current Infectious Disease Reports 11:177-182 (2009)).

[06] A need for vaccines for NTHi exists.

#### BRIEF SUMMARY OF THE INVENTION

[07] As a first aspect, the present invention provides fusion proteins of formula (I).

[08]  $(X)_m - (R_1)_n - A - (Y)_o - B - (Z)_p$  (formula I)

[09] wherein:

X is a signal peptide or MHHHHHH (SEQ ID NO. 2);

m is 0 or 1;

R<sub>1</sub> is an amino acid;

n is 0, 1, 2, 3, 4, 5 or 6;

A is Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof, or PilA from *Haemophilus influenzae* or an immunogenic fragment thereof;

Y is selected from the group consisting of GG, SG, SS, GGG and (G)<sub>h</sub> wherein h is 4, 5, 6, 7, 8, 9, or 10;

o is 0 or 1;

B is PilA from *Haemophilus influenzae* or an immunogenic fragment thereof, or Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof;

Z is GGHHHHHH (SEQ ID NO. 3); and

p is 0 or 1.

[10] As a second aspect, the present invention provides immunogenic compositions comprising fusion proteins of formula (I). The composition may further comprise a pharmaceutically acceptable adjuvant. The composition may comprise an excipient.

[11] In a third aspect, the present invention provides a method for the treatment or prevention of a condition or disease caused wholly or in part by *Haemophilus influenzae*. The method comprises administering to a subject in need thereof a therapeutically effective amount of the fusion protein of formula (I).

[12] In a fourth aspect, the present invention provides a method for the treatment or prevention of otitis media. The method comprises administering to a subject in need thereof a therapeutically effective amount of the fusion protein of formula (I).

[13] In a fifth aspect, the present invention provides a method for the treatment or prevention of exacerbations in chronic obstructive pulmonary disease. The method comprises administering to a subject in need thereof a therapeutically effective amount of the fusion protein of formula (I).

[14] In a sixth aspect, the present invention provides a method for the treatment or prevention of pneumonia. The method comprises administering to a subject in need thereof a therapeutically effective amount of the fusion protein of formula (I).

[15] In a seventh aspect, the present invention provides a pharmaceutical composition comprising a fusion protein of formula (I) for use in the treatment or prevention of a condition or disease caused wholly or in part by *Haemophilus influenzae*. Pharmaceutical compositions may further comprise a pharmaceutically acceptable adjuvant.

[16] In an eighth aspect, the present invention provides nucleic acids encoding the proteins of the invention.

[17] In a ninth aspect, the present invention provides a process of producing nucleic acids of the invention.

[18] Further aspects of the present invention are described in the detailed description of particular embodiments, examples and claims which follow.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. SDS-PAGE of induced bacterial extracts for fusion protein constructs LVL291, LVL268 and LVL269. Insoluble fraction (I), Soluble fraction (S) and Culture Media fraction (M) were loaded for LVL291, LVL268 and LVL269 before and after induction (ind).

Figure 2. SDS-PAGE and Western blot related to purification extracts for fusion protein constructs LVL291, LVL268 and LVL269. Flow through fraction (Ft), Wash fraction (W) and Elution fraction (E) were loaded for purification of LVL291, LVL268 and LVL269. Anti-his tag was used to probe extracts.

Figure 3. SDS-PAGE of induced bacterial and purification extracts for fusion protein constructs LVL291 and LVL315. Culture Media fraction (M), Soluble fraction (Sol), Insoluble fraction (Ins), Flow through fraction (Ft), Wash fraction #1 (W1), Wash fraction #2 (W2) and Elution fraction (E) were loaded for LVL291 and LVL315.

Figure 4. SDS-PAGE of induced bacterial and purification extracts for fusion protein construct LVL312. Culture Media fraction (M), Soluble fraction (Sol), Insoluble fraction (Ins), Flow Through fraction (Ft), Wash fraction #1 (W1), Wash fraction #2 (W2) and Elution fraction (E) were loaded for LVL312.

Figure 5. SDS-PAGE of induced (1mM and 10 $\mu$ M IPTG) bacterial extracts for fusion protein construct LVL317. Extracts from before (NI) and after induction (In), Soluble fraction (S), Insoluble fraction (I).

Figure 6. SDS-PAGE of induced (1mM and 10 $\mu$ M IPTG) bacterial extracts for fusion protein construct LVL318. Extracts from before (NI) and after induction (In), Culture Media fraction (M), Soluble fraction (S), Insoluble fraction (I).

Figure 7. CD spectra of PE, PilA and PE-PilA fusion proteins.

Figure 8. Combination of PE and PilA CD spectrum.



Figure 9. PilA thermal denaturation curve.

Figure 10. PE denaturation curve.

Figure 11. PE-PilA fusion protein thermal denaturation curve.

Figure 12. Typical SP Sepharose™ Fast Flow chromatogram.

Figure 13. Typical Q Sepharose™ Fast Flow chromatogram.

Figure 14. SDS-PAGE of In-process samples from purification process of PE-PilA fusion protein.

Figure 15. Western Blot of In-process samples of purification process from PE-PilA fusion protein. Blot using rabbit polyclonal anti-PE.

Figure 16. Western Blot of In-process samples of purification process from PE-PilA fusion protein. Blot using rabbit polyclonal anti-*E.coli* (BLR).

Figure 17. Thermal transition of PE-PilA fusion protein and PE and PilA proteins. Curves: PilA (1), Protein E (Prot E, PE) (2), PE-PilA Purified Bulk not diluted, 737µg/ml (3), and PE-PilA Purified Bulk diluted at Final Container concentration 60µg/ml (4).

Figure 18. Antibody responses against LVL291 PE-PilA fusion protein and against monovalent PE and PilA in the Balb/c mouse model.

Figure 19. Effect of PE-PilA fusion protein vaccination on NTHi strain 86-028NP bacterial clearance in mouse nasopharynx.

Figure 20. Effect of PE-PilA fusion protein vaccination on NTHi strain 3224A bacterial clearance in mouse nasopharynx.

Figure 21. Effect of PilA vaccination on bacterial clearance in mouse nasopharynx.

Figure 22. Effect of PE vaccination on bacterial clearance in mouse nasopharynx.

Figure 23. (a) LVL317 PE-PilA fusion protein binding to vitronectin and (b) LVL317 and LVL735 PE-PilA fusion protein bound to vitronectin.

Figure 24. Inhibition of vitronectin binding by polyclonal antibodies against PE-PilA fusion protein.

Figure 25. SDS-PAGE of soluble fractions of induced bacterial extracts for fusion protein constructs LVL291, LVL702, LVL736, LVL737, LVL738, LVL739, LVL740 and pET26b vector (negative control). (a) Experiment 1 (b) Experiment 2 (c) Experiment 3. PE-PilA fusion protein indicated by arrow.

Figure 26. The average band percentage of fusion protein in the soluble fraction from Experiments 1, 2 and 3.

Figure 27. PE and PilA antibody response to LVL317 and LVL735.

Figure 28. Effect of LVL735 and LVL317 vaccination on bacterial clearance in a mouse model of non-typeable *Haemophilus influenzae* nasopharyngeal colonization.

## DETAILED DESCRIPTION OF THE INVENTION

[19] Unless otherwise explained or defined herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. For example, definitions of common terms in molecular biology can be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

[20] The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Additionally, numerical limitations given with respect to concentrations or levels of a substance, such as an antigen may be approximate. Thus, where a concentration is indicated to be (for example) approximately 200 pg, it is intended that the concentration includes values slightly more or slightly less than ( “about” or “~”) 200 pg.

[21] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below.

[22] The term “comprises” means “includes”. Thus, unless the context requires otherwise, the word “comprises,” and variations such as “comprise” and “comprising” will be understood to imply the inclusion of a stated compound or composition (e.g., nucleic acid, polypeptide, antigen) or step, or group of compounds or steps, but not to the exclusion of any other compounds, composition, steps, or groups thereof. The abbreviation, “e.g.” is derived from the Latin *exempli gratia*, and is used herein to indicate a non-limiting example. Thus, the abbreviation “e.g.” is synonymous with the term “for example.”

[23] In order to facilitate review of the various embodiments of this disclosure, the following explanations of terms are provided. Additional terms and explanations are provided in the context of this disclosure.

[24] A “subject” as used herein is a mammal, including humans, non-human primates, and non-primate mammals such as members of the rodent genus (including but not limited to mice and rats) and members of the order Lagomorpha (including but not limited to rabbits).

[25] As used herein “Protein E”, “protein E”, “Prot E”, and “PE” mean Protein E from *H. influenzae*. Protein E may consist of or comprise the amino acid sequence of SEQ ID NO. 4 (MKKIILTLSL GLLTACSAQI QKAEQNDVKL APPTDVRSGY IRLVKNVNY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TSLTTPDTTL YNAAQIICAN  
YGEAFSVDKK) as well as sequences with at least or exactly 75%, 77%, 80%, 85%, 90%,

95%, 97%, 99% or 100% identity, over the entire length, to SEQ ID NO. 4. Comparison of 53 sequences of Protein E from *Haemophilus influenzae* (Table 1, SEQ ID NO. 5 – SEQ ID NO. 57) demonstrated approximately 77% to approximately 100% identity to Protein E as set forth in SEQ ID NO. 4. For example, in the amino acid sequence of Protein E, amino acid #20 may be isoleucine (I) or threonine (T); amino acid #23 may be alanine (A) or valine (V); amino acid #24 may be lysine (K) or glutamic acid (E); amino acid #31 may be alanine (A) or threonine (T); amino acid #32 may be proline (P) or alanine (A); amino acid #34 may be threonine (T) or alanine (A); amino acid #37 may be arginine (R) or glutamine (Q); amino acid #47 may be valine (V) or alanine (A); amino acid #57 may be tryptophane (W) or may be absent (-); amino acid #70 may be alanine (A) or threonine (T); amino acid #93 may be glutamine (Q) or absent (-); amino acid #109 may be threonine (T) or isoleucine (I); amino acid #119 may be glycine (G) or serine (S); amino acid #153 may be glutamic acid (E) or lysine (K); amino acid #156 may be serine (S) or leucine (L); amino acid #160 may be lysine (K) or asparagine (N); amino acid #161 may be lysine (K), isoleucine (I) or absent (-); amino acids #162 - #195 may be absent, or as set forth in SEQ ID NO. 15 (with (-) indicating amino acid #166 is absent) or as set forth in SEQ ID NO. 16; or any combination thereof.

[26] Protein E may consist of or comprise an amino acid sequence that differs from SEQ ID NO. 4 at any one or more amino acid selected from the group consisting of: amino acid #20, amino acid #23, amino acid #24, amino acid #31, amino acid #32, amino acid #34, amino acid #37, amino acid #47, amino acid #57, amino acid #70, amino acid #93, amino acid #109, amino acid #119, amino acid #153, amino acid #156, amino acid #160, amino acid #161 and amino acids #162-#195, wherein amino acid #20 is threonine (T); amino acid #23 is valine (V); amino acid #24 is lysine (K); amino acid #31 is threonine (T); amino acid #32 is alanine (A); amino acid #34 is alanine (A); amino acid #37 is glutamine (Q); amino acid #47 is alanine (A); amino acid #57 is absent (-); amino acid #70 is threonine (T); amino acid #93 is absent (-); amino acid #109 is isoleucine (I); amino acid #119 is serine (S); amino acid #153 is lysine (K); amino acid #156 is leucine (L); amino acid #160 is asparagine (N); amino acid #161 is lysine (K) or isoleucine (I); or amino acids #162 - #195 are as set forth in SEQ ID NO. 15 (with (-) indicating amino acid #166 is absent) or as set forth in SEQ ID NO. 16.

[27] Table 1: Protein E amino acid sequences from 53 strains of *Haemophilus influenzae* (SEQ ID NO. 5 - SEQ ID NO. 57). - indicates amino acid is absent.

| Strain Name | Protein E sequence                                                                                                                                                                     |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3224A       | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.5)  |
| RdKW20      | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQIRTDYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.6)   |
| 86-028NP    | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.7)  |
| R2846       | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.8)  |
| R2866       | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.9)  |
| 3655        | MKKIILTSLGLLTACSAQIQKAEQNDMKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.10) |
| PittAA      | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.11) |
| PittEE      | MKKIILTSLGLLTACSAQIQKAEQNDMKLAPPTDVRSGYIRLVKNVNYIDSESI-VDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.12) |
| PittHH      | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.13) |

|          |                                                                                                                                                                                                                               |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PittII   | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.14)                                        |
| R3021    | MKKIILTSLGLLTACSAQTQKAEQNDVKLTPPTDVQSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRIDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKNNKICT-LISLNFIQLLGCREYSIFLQLLLFYC<br>WHF (SEQ ID NO.15) |
| 22.4-21  | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKKIKKICTLISLNFIQLLGCREYSIFLQLLLFYC<br>WHF (SEQ ID NO.16) |
| 3219C    | MKKIILTSLGLLTACSAQIQKAEQNDMKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.17)                                        |
| 3185     | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.18)                                        |
| 3241A    | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.19)                                        |
| 038144S1 | MKKIILTSLGLLTACSAQTQKVEQNDVKLTAPT D VRSGFVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFLVDKK (SEQ ID NO.20)                                      |
| 810956   | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.21)                                        |
| 821246   | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQIRTD FYDEFWGQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKAFSVDKK (SEQ ID NO.22)                                       |

|           |                                                                                                                                                                                         |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 840645    | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.23)   |
| 902550Z19 | MKKIILTSLGLLTACSAQTQKVEQNDVKLTPTDVRSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.24)   |
| A840177   | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.25)   |
| A860514   | MKKIILTSLGLLTACSAQTQKVEQNDVKLTAPT DVRSGYVRLVKNANYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.26) |
| A950014   | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRIDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.27)  |
| 306543X4  | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.28)  |
| A930105   | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.29)  |
| 901905U   | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.30)  |
| A920030   | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEAFSVDKK (SEQ ID NO.31)   |
| 3221B     | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK                                                       |

|            |                                                                                                                                                                                        |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|            | KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.32)                                                                                                                                      |
| 27W116791N | MKKIILTSLGLLTACSAQTQKVEQNDVKLTPPTDVRSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.33) |
| N218       | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.34) |
| N163       | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGEAFSVDKK (SEQ ID NO.35) |
| N162       | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGEAFSVDKK (SEQ ID NO.36) |
| N107       | MKKIILTSLGLLTACSAQTQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQIRTDYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.37)  |
| N91        | MKKIILTSLGLLTACSAQTQKVEQNDVKLTAPADVRSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.38) |
| D211PG     | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVR-YKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGEAFSVDKK (SEQ ID NO.39) |
| D211PD     | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVR-YKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGEAFSVDKK (SEQ ID NO.40) |
| D201PG     | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDKK (SEQ ID NO.41) |



|        |                                                                                                                                                                                        |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| D201PD | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.42) |
| D198PG | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.43) |
| D198PD | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.44) |
| D195PD | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.45) |
| D189PG | MKKIILTSLGLLTACSAQTQKVEQNDVKLTPPTDVRSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.46) |
| D189PD | MKKIILTSLGLLTACSAQTQKVEQNDVKLTPPTDVRSGYVRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.47) |
| D129CG | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGEA FSVDKK (SEQ ID NO.48) |
| D124PG | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.49) |
| D124PD | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK<br>KHTLSLTPD TTLYNAAQIICANYGKA FSVDKK (SEQ ID NO.50) |
| D58PG  | MKKIILTSLGLLTACSAQTQKAEQNDVKLTPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWQGLRAAPKKQK                                                      |

|       |                                                                                                                                                                                         |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.51)                                                                                                                                       |
| D330D | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.52) |
| BS433 | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDTVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.53) |
| BS432 | MKKIILTSLGLLTACSAQTQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQIRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.54) |
| 1714  | MKKIILTSLGLLTACSAQIQKAKQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGEAFSVDDK (SEQ ID NO.55) |
| 1128  | MKKIILTSLGLLTACSAQIQKAEQNDVKLAPPTDVRSGYIRLVKNVNYIDSESIWVDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.56) |
| BS430 | MKKIILTSLGLLTACSAQIQKAEQNDMKLAPPTDVRSGYIRLVKNVNYIDSESI-VDNQEPQ<br>IVHFDAVVNLDKGLYVYPEPKRYARSVRQYKILNCANYHLTQVRTDFYDEFWGQGLRAAPKKQK<br>KHTLSLTPD T TLYNAAQIICANYGKAFSVDDK (SEQ ID NO.57) |

[28] Protein E may be Protein E from *H. influenzae* strain 3224A, RdKW20, 86-028NP, R2846, R2866, 3655, PittAA, PittEE, PittHH, PittII, R3021, 22.4-21, 3219C, 3185, 3241A, 038144S1, 810956, 821246, 840645, 902550Z19, A840177, A860514, A950014, 306543X4, A930105, 901905U, A920030, 3221B, 27W116791N, N218, N163, N162, N107, N91, D211PG, D211PD, D201PG, D201PD, D198PG, D198PD, D195PD, D189PG, D189PD, D129CG, D124PG, D124PD, D58PG, D330D, BS433, BS432, 1714, 1128 or BS430. Protein E may be Protein E as set forth in any of SEQ ID NO. 5 – SEQ ID NO. 57.

[29] Protein E may be a sequence with at least 95% identity, over the entire length, to any of SEQ ID NO. 4 – SEQ ID NO. 57. Protein E may be a sequence with at least 95% identity,

over the entire length, to any of the sequences set forth in Table 1, SEQ ID NO. 5 – SEQ ID NO. 57.

[30] Immunogenic fragments of Protein E comprise immunogenic fragments of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino acids of SEQ ID NO. 4. The immunogenic fragments may elicit antibodies which can bind SEQ ID NO. 4.

[31] Immunogenic fragments of Protein E may comprise immunogenic fragments of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino acids of any of SEQ ID NO. 4 – SEQ ID NO. 57. The immunogenic fragments may elicit antibodies which can bind the full length sequence from which the fragment is derived.

[32] Immunogenic fragments of Protein E comprise immunogenic fragments of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino acids of SEQ ID NO. 5 – SEQ ID NO. 57. The immunogenic fragments may elicit antibodies which can bind the full length sequence from which the fragment is derived.

[33] As used herein "PilA" means Pilin A from *H. influenzae*. PilA may consist of or comprise the protein sequence of SEQ ID NO. 58 (MKLTTQQTLLK KGFTLLIELMI VIAIIAILAT IAIPSYQNYT KKAHVSELLQ ASAPYKADVE LCVYSTNETT NCTGGKNGIA ADITTAKGYV KSVTTSNGAI TVKGDGTLAN MEYILQATGN AATGVTWTTT CKGTDASLFP ANFCGSVTQ) as well as sequences with 80% to 100% identity to SEQ ID NO. 58. For example, PilA may be at least 80%, 85%, 90%, 95%, 97% or 100% identical to SEQ ID NO. 58. Full length comparison of 64 sequences of PilA from *Haemophilus influenzae* (Table 2, SEQ ID NO. 58 – SEQ ID NO. 121) demonstrated approximately 80% to 100% identity to PilA as set forth in SEQ ID NO. 58. For example, in the amino acid sequence of PilA, amino acid #6 may be glutamine (Q) or leucine (L); amino acid #7 may be glutamine (Q) or threonine (T); amino acid #37 may be glutamine (Q) or lysine (K); amino acid #44 may be alanine (A) or serine (S); amino acid #57 may be alanine (A) or serine (S); amino acid #67 may be asparagine (N) or glycine (G); amino acid #68 may be glutamic acid (E) or lysine (K); amino acid #69 may be threonine (T) or proline (P); amino acid #71 may be lysine (K), asparagine (N), serine (S) or threonine (T); amino acid #73 may be threonine (T), serine (S) or methionine (M); amino acid #76 may be lysine (K), serine (S) or asparagine (N); amino acid #84 may be threonine (T) or lysine (K); amino acid #86 may be alanine (A) or valine (V); amino acid #91 may be lysine (K) or alanine (A); amino acid #94 may be threonine (T),

isoleucine (I) or lysine (K); amino acid #96 may be serine (S) or glutamine (Q); amino acid #97 may be asparagine (N) or serine (S); amino acid #99 may be alanine (A) or glycine (G); amino acid #103 may be alanine (A) or lysine (K); amino acid #109 may be aspartic acid (D), alanine (A) or threonine (T); amino acid #110 may be glycine (G), asparagine (N), or arginine (R); amino acid #112 may be serine (S) or glutamic acid (E); amino acid #114 may be threonine (T) or isoleucine (I); amino acid #116 may be threonine (T) or glutamine (Q); amino acid #118 may be glutamic acid (E), threonine (T), alanine (A), lysine (K) or serine (S); amino acid #121 may be serine (S) or alanine (A); amino acid #122 may be alanine (A) or threonine (T); amino acid #123 may be lysine (K), threonine (T) or alanine (A); amino acid #128 may be lysine (K) or threonine (T); amino acid #135 may be aspartic acid (D) or glutamic acid (E); amino acid #136 may be alanine (A) or threonine (T); amino acid #145 may be glycine (G) or arginine (R); amino acid #149 may be glutamine (Q) or lysine (K); or any combination thereof.

[34] Pil A may consist of or comprise an amino acid sequence that differs from SEQ ID NO. 58 at any or more amino acid selected from the group consisting of amino acid #6, amino acid #7, amino acid #37, amino acid #44, amino acid #57, amino acid #67, amino acid #68, amino acid #69, amino acid #71, amino acid #73, amino acid #76, amino acid #84, amino acid #86, amino acid #91, amino acid #94, amino acid #96, amino acid #97, amino acid #99, amino acid #103, amino acid #109, amino acid #110, amino acid #112, amino acid #114, amino acid #116, amino acid #118 amino acid, #121, amino acid #122, amino acid #123, amino acid #128, amino acid #135, amino acid #136, amino acid #145 and amino acid #149, wherein amino acid #6 is leucine (L); amino acid #7 is threonine (T); amino acid #37 is lysine (K); amino acid #44 is serine (S); amino acid #57 is serine (S); amino acid #67 is glycine (G); amino acid #68 is lysine (K); amino acid #69 is proline (P); amino acid #71 is lysine (K), serine (S) or threonine (T); amino acid #73 is serine (S) or methionine (M); amino acid #76 is serine (S) or asparagine (N); amino acid #84 is lysine (K); amino acid #86 is valine (V); amino acid #91 is alanine (A); amino acid #94 is isoleucine (I) or lysine (K); amino acid #96 is glutamine (Q); amino acid #97 is serine (S); amino acid #99 is glycine (G); amino acid #103 is alanine (A); amino acid #109 is aspartic acid (D) or threonine (T); amino acid #110 is glycine (G) or arginine (R); amino acid #112 is serine (S); amino acid #114 is threonine (T); amino acid #116 is threonine (T); amino acid #118 is glutamic acid (E), alanine (A), lysine (K) or serine (S); amino acid #121 is serine (S); amino acid #122 is threonine (T); amino acid #123 is lysine (K) or alanine (A); amino acid #128 is lysine (K); amino acid #135 is glutamic acid (E); amino acid #136 is threonine (T); amino acid #145 is arginine (R); amino acid #149 is lysine (K).

[35] Table 2: Pilin A amino acid sequences from 64 strains of *Haemophilus influenzae* (SEQ ID NO. 58 – SEQ ID NO. 121).

| Strain Name | PilA sequence                                                                                                                                                                  |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 86-028NP    | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.58)  |
| NTHi3219C   | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTKCTGGKNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWK<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.59)  |
| NTHi3224A   | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.60)  |
| NTHi12      | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYKQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSSCSGGSNGIAADITTAKGYVASVITQSGGITVKGDGTLANMEYILQAAGNAAAGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.61) |
| NTHi44      | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.62)  |
| NTHi67      | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.63)  |
| 1054MEE     | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.64)  |
| 1729MEE     | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.65)  |
| 1728MEE     | MKLTTQQTLLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY                                                                                                              |

|         |                                                                                                                                                                              |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|         | STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.66)                                                                     |
| 1885MEE | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEITNCMGGKNGIAADITTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAAAGVTWT<br>TTCKGTDASLFPANFCGSITQ (SEQ ID NO.67) |
| 1060MEE | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKASVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.68)  |
| RdKW20  | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTSCTGGKNGIAADIKTAKGYVASVITQSGGITVKGNGLANMEYILQAKGNAAAGVTWT<br>TTCKGTDASLFPANFCGSVTK (SEQ ID NO.69)  |
| 214NP   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSSCSGGSNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.70) |
| 1236MEE | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTSCTGGKNGIAADIKTAKGYVASVITQSGGITVKGNGLANMEYILQAKGNAAAGVTWT<br>TTCKGTDASLFPANFCGSVTK (SEQ ID NO.71)  |
| 1714MEE | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.72)  |
| 1128MEE | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKASVSELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.73)  |
| R2846   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.74) |
| R2866   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTEASLFPANFCGSVTQ (SEQ ID NO.75) |

|          |                                                                                                                                                                              |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3655     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKASVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.76)  |
| PittAA   | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.77)  |
| PittGG   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.78)  |
| PittII   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGDGLANMEYILQATGNAATGVTWT<br>TTCKGTEASLFPANFCGSVTQ (SEQ ID NO.79)  |
| R3021    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGDGLANMEYILQATGNAATGVTWT<br>TTCKGTEASLFPANFCGSVTQ (SEQ ID NO.80)  |
| 22.4-21  | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWK<br>TTCKGTDASLFPANFCGSVTK (SEQ ID NO.81) |
| 3185A    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEATKCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.82)  |
| 3221B    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEATKCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.83)  |
| 3241A    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.84)  |
| 038144S1 | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAISELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT                                          |

|           |                                                                                                                                                                               |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | TTCKGTDASLFPANFCRSVTK (SEQ ID NO.85)                                                                                                                                          |
| 821246    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTEASLFPANFCGSVTQ (SEQ ID NO.86)  |
| 840645    | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.87)  |
| 902550Z19 | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTVKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.88)  |
| A840177   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.89)  |
| A920030   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGNGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.90)  |
| A950014   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDMSYTLTAEGDSAKGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.91)   |
| 901905U   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSSCSGGSNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.92)  |
| A920029   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKSDVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVITQSGGITVKGNGTLTNMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSITQ (SEQ ID NO.93)  |
| A930105   | MKLTTLQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGNGNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.94) |



|          |                                                                                                                                                                               |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 306543X4 | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSSCSGGSNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.95)  |
| N218     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEATKCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDTSLFPANFCGSVTQ (SEQ ID NO.96)  |
| N163     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.97)  |
| N162     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.98)   |
| N120     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.99)   |
| N107     | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.100)  |
| N92      | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.101) |
| N91      | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.102)  |
| D219PG   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEATKCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.103) |
| D211PG   | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT                                          |

|        |                                                                                                                                                                               |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|        | TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.104)                                                                                                                                         |
| D211PD | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.105) |
| D204CD | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILXATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.106) |
| D198PG | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.107) |
| D198PD | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.108) |
| D195PD | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.109) |
| D195CD | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.110) |
| D189PG | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTSCTGGKNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWK<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.111) |
| D189PD | MKLTTQOTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTSCTGGKNGIAADITTAKGYVKSVTTSNGAITVAGNGTLDGMSYTLTAEGDSAKGVTWK<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.112) |
| D124PG | MKLTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.113) |

|        |                                                                                                                                                                                |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| D124PD | MKLTTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.114) |
| D124CG | MKLTTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.115) |
| D58PG  | MKLTTQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNETTNCTGGKNGIAADITTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTEASLFPANFCGSVTQ (SEQ ID NO.116)   |
| BS433  | MKLTTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGNNGIAADIKTAKGYVASVKTQSGGITVKGDGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.117) |
| BS432  | MKLTTQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGTLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.118)   |
| BS430  | MKLTTQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STNEATKCTGGKNGIAADITTAKGYVKSVTTSNGAITVKGDGTLANMEYILQASGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.119)   |
| 1714   | MKLTTTLQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYKADVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGTLANMEYILQATGNAATGVTWT<br>TTCKGTDASLFPANFCGSVTQ (SEQ ID NO.120) |
| 1128   | MKLTTQTLKKGFTLIELMIVIAIIAILATIAIPSYQNYTKKASVSELLQASAPYKSDVELCVY<br>STGKPSTCSGGSNGIAADITTAKGYVASVKTQSGGITVKGNGTLANMEYILQAKGNATAGVTWT<br>TTCKGTDASLFPANFCRSVTK (SEQ ID NO.121)   |

[36] PilA may be PilA from *H. influenzae* strain NTHi3219C, NTHi3224A, NTHi12, NTHi44, NTHi67, 1054MEE, 1729MEE, 1728MEE, 1885MEE, 1060MEE, RdKW20, 214NP, 1236MEE, 1714MEE, 1128MEE, 86-028NP, R2846, R2866, 3655, PittAA, PittGG, PittII, R3021, 22.4-21, 3185A, 3221B, 3241A, 038144S1, 821246, 840645, 902550Z19, A840177, A920030,

A950014, 901905U, A920029, A930105, 306543X4, N218, N163, N162, N120, N107, N92, N91, D219PG, D211PG, D211PD, D204CD, D198PG, D198PD, D195PD, D195CD, D189PG, D189PD, D124PG, D124PD, D124CG, D58PG, BS433, BS432, BS430, 1714 or 1128. An amino acid sequence for PilA from *H. influenzae* strain D204CD is set forth in SEQ ID NO. 106, wherein X at position #116 is either glutamine (Q) or leucine (L); ambiguity as to the amino acid at position #116 could be cleared up by technical resolution of the second nucleotide encoding amino acid #116, clarifying the PilA sequence for strain D204CD. PilA may be PilA as set forth in any of SEQ ID NO. 58 – SEQ ID NO. 121.

[37] PilA may be a sequence with at least 95% identity, over the entire length, to any of SEQ ID NO. 58 – SEQ ID NO. 121 (as set out in Table 2).

[38] Immunogenic fragments of PilA comprise immunogenic fragments of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino acids of SEQ ID NO. 58 – SEQ ID NO. 121. The immunogenic fragments may elicit antibodies which can bind the full length sequence from which the fragment is derived.

[39] For example, immunogenic fragments of PilA comprise immunogenic fragments of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino acids of SEQ ID NO. 58. The immunogenic fragments may elicit antibodies which can bind SEQ ID NO. 58.

[40] Identity between polypeptides may be calculated by various algorithms. For example, the Needle program, from the EMBOSS package (Free software; EMBOSS: The European Molecular Biology Open Software Suite (2000). Trends in Genetics 16(6): 276—277) and the Gap program from the GCG<sup>®</sup> package (Accelrys Inc.) may be used. This Gap program is an implementation of the Needleman-Wunsch algorithm described in: Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453. The BLOSUM62 scoring matrix has been used, and the gap open and extension penalties were respectively 8 and 2.

[41] Looking at the computed alignment, identical residues between two compared sequences can be observed. A percentage of identity can be computed by (1) calculating the number of identities divided by the length of the alignment, multiplied by 100 (for example, for the Needle program analysis), (2) calculating the number of identities divided by the length of the longest sequence, multiplied by 100, (3) calculating the number of identities divided by the length of the shortest sequence, multiplied by 100, or (4) calculating the number of

identities divided by the number of aligned residues, multiplied by 100 (a residue is aligned if it is in front of another) (for example, for the Gap program analysis).

[42] As used herein, "adjuvant" means a compound or substance that, when administered to a subject in conjunction with a vaccine, immunotherapeutic, or other antigen- or immunogen-containing composition, increases or enhances the subject's immune response to the administered antigen or immunogen (as compared to the immune response that would be obtained in the absence of adjuvant). This is to be distinguished from "adjuvant therapy", defined by the National Cancer Institute of the United States Institutes of Health in the context of cancer treatment as additional treatment given after the primary treatment, to lower the risk that the cancer will recur.

[43] Conservative substitutions are well known and are generally set up as the default scoring matrices in sequence alignment computer programs. These programs include PAM250 (Dayhoff M.O. *et al.*, (1978), "A model of evolutionary changes in proteins", In "Atlas of Protein sequence and structure" 5(3) M.O. Dayhoff (ed.), 345-352), National Biomedical Research Foundation, Washington, and Blosum 62 (Steven Henikoff and Jorja G. Henikoff (1992), "Amino acid substitution matrices from protein blocks"), *Proc. Natl. Acad. Sci. USA* 89 (Biochemistry): 10915-10919. The invention further provides fusion proteins of formula (I) containing conservative amino acid substitutions. For example, the fusion proteins of formula (I) may contain a conservative substitution of any amino acid from PE or PilA of *H. influenzae* as described in any of the sequences set forth herein (for example, any PE sequence set forth in SEQ ID NO. 4 – SEQ ID NO. 57 and/or any PilA sequence set forth in SEQ ID NO. 58 – SEQ ID NO. 121)

[44] As used herein "signal peptide" refers to a short (less than 60 amino acids, for example, 3 to 60 amino acids) polypeptide present on precursor proteins (typically at the N terminus), and which is typically absent from the mature protein. The signal peptide (sp) is typically rich in hydrophobic amino acids. The signal peptide directs the transport and/or secretion of the translated protein through the membrane. Signal peptides may also be called targeting signals, transit peptides, localization signals, or signal sequences. For example, the signal sequence may be a co-translational or post-translational signal peptide.

[45] A heterologous signal peptide may be cleaved from a fusion protein construct by signal peptide peptidases during or after protein transportation or secretion. For example, the signal

peptide peptidase is signal peptide peptidase I. A “heterologous” signal peptide is one which is not associated with the protein as it exists in nature.

[46] As used herein “treatment” means the prevention of occurrence of symptoms of the condition or disease in a subject, the prevention of recurrence of symptoms of the condition or disease in a subject, the delay of recurrence of symptoms of the condition or disease in a subject, the decrease in severity or frequency of symptoms of the condition or disease in a subject, slowing or eliminating the progression of the condition and the partial or total elimination of symptoms of the disease or condition in a subject.

[47] As used herein, “optionally” means that the subsequently described event(s) may or may not occur, and includes both event(s) that occur and events that do not occur.

[48] The pathogenesis of disease caused by NTHi begins with nasopharyngeal colonization. Mechanisms to adhere to and maintain long-term residence within the nasopharyngeal micro-environment are considered ‘virulence determinants’ for NTHi. (Vaccine 28: 279-289 (2010)).

[49] The importance of NTHi being able to adhere to the mucosal epithelial surfaces of a human host is reflected in the multiplicity of adhesins expressed by NTHi. For example, some NTHi express pili. Other adhesive structures belong to the autotransporter family of proteins; these include Hap, HMW1/HMW2 and Hia/Hsf proteins. Further outer membrane proteins, such as the P2 protein, P5 protein and OapA have been described as adhesions for *Haemophilus influenzae*. (Cellular Microbiology 4:191-200 (2002), Microbes and Infection 10: 87-96 (2008), Vaccine 28: 279-289 (2010)).

[50] Otitis media is a major cause of morbidity in 80% of all children less than 3 years of age. (Expert Rev. Vaccines 5:517-534 (2006)). More than 90% of children develop otitis media before age 7 (Current Opinion in Investigational Drugs 4:953-958 (2003)). In 2000, there were 16 million visits made to office-based physicians for otitis media in the United States and approximately 13 million antibacterial prescriptions dispensed. (Pediatrics 113:1451-1465 (2004)). In European countries, the reported acute otitis media rates range between 0.125 to 1.24 per child-year. (Expert Review of Vaccines 8:1479-1500 (2009)). Otitis media is a costly infection and the most common reason children receive antibiotics. (Current Infectious Disease Reports 11:177-182 (2009)). Bacteria are responsible for approximately 70% of cases of acute otitis media, with *Streptococcus pneumoniae*, non-typeable *Haemophilus influenzae*, and *Moraxella catarrhalis* predominating as the causative agents (Expert Review

of Vaccines 5:517-534 (2006)). A subset of children experience recurrent and chronic otitis media and these otitis prone children have protracted middle-ear effusions that are associated with hearing loss and delays in speech and language development. (Current Infectious Disease Reports 11:177-182 (2009)).

[51] Following the introduction of the heptavalent pneumococcal vaccine in many countries, some studies have demonstrated a significant increase in the proportion of acute otitis media caused by *H. influenzae*, with *H. influenzae* becoming the predominant pathogen. (Pediatric Infectious Disease Journal 23:824-828; Pediatric Infectious Disease Journal 23:829-833 (2004)).

[52] Since otitis media is a multifactorial disease, the feasibility of preventing otitis media using a vaccination strategy has been questioned. (Current Infectious Disease Reports 11:177-182 (2009)). However, the results from one study suggest that it is possible for an antigen to induce at least partial protection against non-typeable *H. influenzae*. (Lancet 367:740-748 (2006)). One approach to developing vaccine antigens is to use antigenically conserved regions of genetically heterogeneous but abundantly expressed surface molecules. Another approach is to identify surface proteins that demonstrate sequence or functional epitope conservation. A third consideration for a vaccine antigen could be to select an antigen that is expressed during infection and colonization in a human host. Murphy (Curr. Infect. Disease Reports 11:177-182 (2009)) states that, despite the existence of several potential non-typeable *H. influenzae* candidate antigens, one cannot predict with certainty whether the candidate antigen will be effective. (Current Infectious Disease Reports 11:177-182 (2009)). Some of the proteins described as potential vaccine antigens are: *Haemophilus* adhesin protein (Hap), High molecular-weight (HMW) proteins 1 and 2, *H. influenzae* adhesin (Hia), D15 protein, HtrA heat shock protein, P2 surface protein, lipoprotein D, P5 fimbriae derived peptides, outer membrane protein P4, outer membrane protein (OMP) 26 (OMP26), P6 protein, Protein E, Type IV pilus, lipooligosaccharide and phosphoryl choline. (Current Infectious Disease Reports 11:177-182 (2009); Expert Review of Vaccines 5:517-534 (2006)).

[53] The chinchilla model is a robust and validated animal model of otitis media and its prevention (Expert Review of Vaccines 8:1063-1082 (2009)). While the chinchilla model may mimic the natural course of human infection, others have suggested that results in the chinchilla model may vary from one laboratory to the next. (Current Opinion in Investigational Drugs 4:953-958 (2003)).

[54] Various other rodents have also been used for the induction of otitis media and are summarized in Vaccine 26:1501-1524 (2008). The murine animal model is often studied in otitis media research.

[55] The presence of bactericidal antibody is associated with protection from otitis media due to non-typeable *H. influenzae*. (Current Opinion in Infectious Disease 16:129-134 (2003)). However, an immune response need not be bactericidal to be effective against NTHi. Antibodies that merely react with NTHi surface adhesins can reduce or eliminate otitis media in the chinchilla. (Current Opinion in Investigational Drugs 4:953-958 (2003)).

[56] Chronic obstructive pulmonary disease is a chronic inflammatory disease of the lungs and a major cause of morbidity and mortality worldwide. Approximately one in 20 deaths in 2005 in the US had COPD as the underlying cause. (Drugs and Aging 26:985-999 (2009)). It is projected that in 2020 COPD will rise to the fifth leading cause of disability adjusted life years, chronic invalidating diseases, and to the third most important cause of mortality (Lancet 349:1498-1504 (1997)).

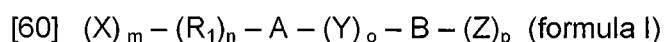
[57] The course of COPD is characterized by progressive worsening of airflow limitation and a decline in pulmonary function. COPD may be complicated by frequent and recurrent acute exacerbations (AE), which are associated with enormous health care expenditure and high morbidity. (Proceedings of the American Thoracic Society 4:554–564 (2007)). One study suggests that approximately 50% of acute exacerbations of symptoms in COPD are caused by non-typeable *Haemophilus influenzae*, *Moraxella catarrhalis*, *Streptococcus pneumoniae*, and *Pseudomonas aeruginosa*. (Drugs and Aging 26:985-999 (2009)). *H. influenzae* is found in 20-30% of exacerbations of COPD; *Streptococcus pneumoniae*, in 10-15% of exacerbations of COPD; and *Moraxella catarrhalis*, in 10-15% of exacerbations of COPD. (New England Journal of Medicine 359:2355-2365 (2008)). *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Moraxella catarrhalis* have been shown to be the primary pathogens in acute exacerbations of bronchitis in Hong Kong, South Korea, and the Philippines, while *Klebsiella spp.*, *Pseudomonas aeruginosa* and *Acinetobacter spp.* constitute a large proportion of pathogens in other Asian countries/regions including Indonesia, Thailand, Malaysia and Taiwan (Respirology, (2011) 16, 532-539; doi:10.1111/j.1440.1843.2011.01943.x). In Bangladesh, 20% of patients with COPD showed positive sputum culture for *Pseudomonas*, *Klebsiella*, *Streptococcus pneumoniae* and *Haemophilus influenzae*, while 65% of patients with AECOPD showed positive cultures for



*Pseudomonas*, *Klebsiella*, *Acinetobacter*, *Enterobacter*, *Moraxella catarrhalis* and combinations thereof. (Mymensingh Medical Journal 19:576-585 (2010)). However, it has been suggested that the two most important measures to prevent COPD exacerbation are active immunizations and chronic maintenance of pharmacotherapy. (Proceedings of the American Thoracic Society 4:554–564 (2007)).

[58] There is a need for effective vaccines against NTHi. Using antigens that may act at different steps in pathogenesis may improve the efficacy of a vaccine. The inventors have found that PilA and PE may be beneficially present in the immunogenic compositions of the invention as fusion proteins.

[59] The present invention relates to fusion proteins of formula (I).



[61] wherein:

X is a signal peptide or MHHHHHH (SEQ ID NO. 2);

m is 0 or 1;

R<sub>1</sub> is an amino acid;

n is 0, 1, 2, 3, 4, 5 or 6;

A is Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof, or PilA from *Haemophilus influenzae* or an immunogenic fragment thereof;

Y is selected from the group consisting of GG, SG, SS and (G)<sub>h</sub> wherein h is 4, 5, 6, 7, 8, 9, or 10;

o is 0 or 1;

B is PilA from *Haemophilus influenzae* or an immunogenic fragment thereof, or Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof;

Z is GGHHHHHH (SEQ ID NO: 3); and

p is 0 or 1.

[62] In one embodiment, the fusion proteins of formula (I) are defined wherein X is selected from the group consisting of the signal sequence from CcmH (cytochrome c membrane protein H), DsbA (periplasmic protein disulfide isomerase I), DsbB (disulfide bond membrane protein B), FlgI (flagellar peptidoglycan ring protein), FocC (F1c Chaperone protein), MalE (maltose transporter subunit E), NadA (quinolinate synthase subunit A), NikA (nickel ABC transporter component A), NspA (Neisserial surface protein A), Omp26 (outer membrane

protein 26), OmpA (outer membrane protein A), OspA (outer surface protein A), pelB (pectate lyase B), PhoA (bacterial alkaline phosphatase), PhtD (pneumococcal histidine triad protein D), PhtE (pneumococcal histidine triad protein E), SfmC (periplasmic pilin chaperone), Sip1 (surface immunogenic protein), TolB (Tol-Pal Cell Envelope Complex Component B), TorA (trimethylamine N-oxide reductase system subunit A), TorT (trimethylamine N-oxide reductase system periplasmic protein T) and Yral (putative periplasmic pilin chaperone); or any subgroup thereof. In one embodiment, X is a co-translational signal peptide or a post-translational signal peptide. In one embodiment X is the signal sequence from FlgI (*flgI sp*). In another particular embodiment, X is the signal sequence from pelB (*pelB sp*). In another embodiment, X is a post-translational signal peptide. In another embodiment, X is selected from the group consisting of the signal sequence from FlgI, NadA and pelB.

[63] In one embodiment, the fusion proteins of formula (I) are defined wherein m is 1. In another embodiment, m is 0.

[64] In one particular embodiment,  $R_1$  and n are defined wherein  $(R_1)_n$  is 1 to 6 amino acids enriched in small, usually hydrophilic, amino acids. Hydrophilic amino acids include glutamic acid (E), aspartic acid (D) and asparagine (N).

[65] In one embodiment, the fusion proteins of formula (I) are defined wherein n is selected from the group consisting of 0, 1, 2 and 6. In one particular embodiment,  $R_1$  and n are defined wherein  $(R_1)_n$  is selected from the group consisting of D, E, ATNDDD (SEQ ID NO. 178) and MD, or any subset thereof.

[66] In one particular embodiment, n is selected from the group consisting of 1, 2 and 6. In one particular embodiment, n is 0.

[67] In one embodiment, the fusion proteins of formula (I) are defined wherein A is Protein E from *H. influenzae*. In another embodiment, the fusion proteins of formula (I) are defined wherein A is Protein E as encoded by an amino acid sequence selected from the group consisting of SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8, SEQ ID NO. 9, SEQ ID NO. 10, SEQ ID NO. 11, SEQ ID NO. 12, SEQ ID NO. 13, SEQ ID NO. 14, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 18, SEQ ID NO. 19, SEQ ID NO. 20, SEQ ID NO. 21, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 25, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 29, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID

NO. 36, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO.39, SEQ ID NO. 40, SEQ ID NO. 41, SEQ ID NO. 42, SEQ ID NO. 43 SEQ ID NO. 44, SEQ ID NO. 45, SEQ ID NO. 46, SEQ ID NO. 47, SEQ ID NO. 48, SEQ ID NO. 49, SEQ ID NO. 50, SEQ ID NO. 51, SEQ ID NO. 52, SEQ ID NO. 53, SEQ ID NO. 54, SEQ ID NO. 55, SEQ ID NO. 56 and SEQ ID NO. 57; or any subset of SEQ ID NO. 5 through SEQ ID NO. 57. In another embodiment, the fusion proteins of formula (I) are defined wherein A is Protein E, wherein Protein E is approximately 75% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO: 4. In another embodiment, A is Protein E wherein Protein E is approximately 90% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO: 4. In another embodiment, A is Protein E wherein Protein E is at least 95% identical to the Protein E amino acid sequence set forth in SEQ ID NO: 4. In additional embodiment, A is Protein E wherein Protein E is at least 95% identical to Protein E as set for in any of SEQ ID NO. 4 – SEQ ID NO. 57. In a particular embodiment, A is Protein E having the amino acid sequence set forth in SEQ ID NO. 4.

[68] In another embodiment, the fusion proteins of formula (I) are defined wherein A is an immunogenic fragment of Protein E from *H. influenzae*. In another embodiment, A is an immunogenic fragment of Protein E wherein Protein E has an amino acid sequence selected from the group consisting of SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8, SEQ ID NO. 9, SEQ ID NO. 10, SEQ ID NO. 11, SEQ ID NO. 12, SEQ ID NO. 13, SEQ ID NO. 14, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 18, SEQ ID NO. 19, SEQ ID NO. 20, SEQ ID NO. 21, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 25, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 29, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 36, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO.39, SEQ ID NO. 40, SEQ ID NO. 41, SEQ ID NO. 42, SEQ ID NO. 43 SEQ ID NO. 44, SEQ ID NO. 45, SEQ ID NO. 46, SEQ ID NO. 47, SEQ ID NO. 48, SEQ ID NO. 49, SEQ ID NO. 50, SEQ ID NO. 51, SEQ ID NO. 52, SEQ ID NO. 53, SEQ ID NO. 54, SEQ ID NO. 55, SEQ ID NO. 56 and SEQ ID NO. 57; or any subset of SEQ ID NO. 4 through SEQ ID NO. 57. In another embodiment, A is an immunogenic fragment of Protein E, wherein Protein E is approximately 75% to 100% identical to the amino acid sequence set forth in SEQ ID NO: 4. In another embodiment, A is an immunogenic fragment of Protein E, wherein Protein E is approximately 90% to 100% identical to SEQ ID NO. 4. In an additional embodiment, A is an immunogenic fragment of Protein E, wherein Protein E is at least 95% identical to any of SEQ ID NO. 4 – SEQ ID NO. 57. More specifically, in one embodiment, A is an immunogenic fragment of Protein E,

wherein Protein E is 93% to 100% identical to SEQ ID NO. 124. In a particular embodiment, A is an immunogenic fragment of Protein E wherein Protein E is SEQ ID NO. 4.

[69] In another embodiment, A is an immunogenic fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 17-160 of SEQ ID NO. 4 (SEQ ID NO. 122), amino acids 18-160 of SEQ ID NO. 4 (SEQ ID NO. 123), amino acids 19-160 of SEQ ID NO. 4 (SEQ ID NO. 124), amino acids 20-160 of SEQ ID NO. 4 (SEQ ID NO. 125) and amino acids 22-160 of SEQ ID NO. 4 (SEQ ID NO. 126). In another embodiment, A is an immunogenic fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 17-160 of SEQ ID NO. 4 (SEQ ID NO. 122), amino acids 18-160 of SEQ ID NO. 4 (SEQ ID NO. 123), amino acids 19-160 of SEQ ID NO. 4 (SEQ ID NO. 124), amino acids 20-160 of SEQ ID NO. 4 (SEQ ID NO. 125), amino acids 22-160 of SEQ ID NO. 4 (SEQ ID NO. 126), amino acids 23-160 of SEQ ID NO. 4 (SEQ ID NO. 179) and amino acids 24-160 of SEQ ID NO. 4 (SEQ ID NO. 180). In a further embodiment, A is an immunogenic fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 17-160 of SEQ ID NO. 4 (SEQ ID NO. 122), amino acids 18-160 of SEQ ID NO. 4 (SEQ ID NO. 123), amino acids 20-160 of SEQ ID NO. 4 (SEQ ID NO. 125), amino acids 22-160 of SEQ ID NO. 4 (SEQ ID NO. 126), amino acids 23-160 of SEQ ID NO. 4 (SEQ ID NO. 179) and amino acids 24-160 of SEQ ID NO. 4 (SEQ ID NO. 180). More specifically, in one embodiment, A is SEQ ID NO. 124, amino acids 19-160 of SEQ ID NO. 4. In an additional embodiment, A is SEQ ID NO. 125, amino acids 20-160 of SEQ ID NO. 5. In another embodiment, A is immunogenic fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 23-160 of SEQ ID NO. 4 (SEQ ID NO. 179) and amino acids 24-160 of SEQ ID NO. 4 (SEQ ID NO. 180).

[70] Protein E - SEQ ID NO. 4

MKKIILTLSL GLLTACSAQI QKAEQNDVKL APPTDVRSYG IRLVKNVNYY  
LDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLEQVRED FYDEFWGQGI RAAPKKQKKH TSLTPDPTTL YNAAQIICAN  
YGEAFSVDKK

[71] Amino acids 17-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 122

SAQI QKAEQNDVKL APPTDVRSYG IRLVKNVNYY

IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK

[72] Amino acids 18-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 123

AQI QKAEQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK

[73] Amino acids 19-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 124

QI QKAEQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK

[74] Amino acids 20-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 125

I QKAEQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK

[75] Amino acids 22-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 126

KAEQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK

[76] Amino acids 23-160 of Protein E from SEQ ID NO. 4 - SEQ ID NO. 179

AEQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA

NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAF'SVDKK

[77] Amino acids 24-160 Protein E from SEQ ID NO. 4 - SEQ ID NO. 180

EQNDVKL APPTDVRSGY IRLVKNVNY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAF'SVDKK

[78] In another embodiment, the fusion proteins of formula (I) are defined wherein A is PilA from *H. influenzae*. In another embodiment, the fusion proteins of formula (I) are defined wherein A is PilA from *H. influenzae* having an amino acid sequence selected from the group consisting of SEQ ID NO. 58, SEQ ID NO. 59, SEQ ID NO. 60, SEQ ID NO. 61, SEQ ID NO. 62, SEQ ID NO. 63, SEQ ID NO. 64, SEQ ID NO. 65, SEQ ID NO. 66, SEQ ID NO. 67, SEQ ID NO. 68, SEQ ID NO. 69, SEQ ID NO. 70, SEQ ID NO. 71, SEQ ID NO. 72, SEQ ID NO. 73, SEQ ID NO. 74, SEQ ID NO. 75, SEQ ID NO. 76, SEQ ID NO. 77, SEQ ID NO. 78, SEQ ID NO. 79, SEQ ID NO. 80, SEQ ID NO. 81, SEQ ID NO. 82, SEQ ID NO. 83, SEQ ID NO. 84, SEQ ID NO. 85, SEQ ID NO. 86, SEQ ID NO. 87, SEQ ID NO. 88, SEQ ID NO. 89, SEQ ID NO. 90, SEQ ID NO. 91, SEQ ID NO. 92, SEQ ID NO. 93, SEQ ID NO. 94, SEQ ID NO. 95, SEQ ID NO. 96, SEQ ID NO. 97, SEQ ID NO. 98, SEQ ID NO. 99, SEQ ID NO. 100, SEQ ID NO. 101, SEQ ID NO. 102, SEQ ID NO. 103, SEQ ID NO. 104, SEQ ID NO. 105, SEQ ID NO. 106, SEQ ID NO. 107, SEQ ID NO. 108, SEQ ID NO. 109, SEQ ID NO. 110, SEQ ID NO. 111, SEQ ID NO. 112, SEQ ID NO. 113, SEQ ID NO. 114, SEQ ID NO. 115, SEQ ID NO. 116, SEQ ID NO. 117, SEQ ID NO. 118, SEQ ID NO. 119, SEQ ID NO. 120 and SEQ ID NO. 121; or any subset of SEQ ID NO. 58 through SEQ ID NO. 121. In another embodiment, A is PilA wherein PilA is approximately 80% to 100% identical to SEQ ID NO. 58. In another embodiment, A is PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121. In a particular embodiment, A is PilA of SEQ ID NO. 58.

[79] In another embodiment, the fusion proteins of formula (I) are defined wherein A is an immunogenic fragment of PilA from *H. influenzae*. In another embodiment, A is an immunogenic fragment of PilA wherein PilA is approximately 80% to 100% identical to SEQ ID NO. 58. For example, A is an immunogenic fragment of PilA wherein PilA has an amino acid sequence selected from the group consisting of SEQ ID NO. 58, SEQ ID NO. 59, SEQ ID NO.

60, SEQ ID NO. 61, SEQ ID NO. 62, SEQ ID NO. 63, SEQ ID NO. 64, SEQ ID NO. 65, SEQ ID NO. 66, SEQ ID NO. 67, SEQ ID NO. 68, SEQ ID NO. 69, SEQ ID NO. 70, SEQ ID NO. 71, SEQ ID NO. 72, SEQ ID NO. 73, SEQ ID NO. 74, SEQ ID NO. 75, SEQ ID NO. 76, SEQ ID NO. 77, SEQ ID NO. 78, SEQ ID NO. 79, SEQ ID NO. 80, SEQ ID NO. 81, SEQ ID NO. 82, SEQ ID NO. 83, SEQ ID NO. 84, SEQ ID NO. 85, SEQ ID NO. 86, SEQ ID NO. 87, SEQ ID NO. 88, SEQ ID NO. 89, SEQ ID NO. 90, SEQ ID NO. 91, SEQ ID NO. 92, SEQ ID NO. 93, SEQ ID NO. 94, SEQ ID NO. 95, SEQ ID NO. 96, SEQ ID NO. 97, SEQ ID NO. 98, SEQ ID NO. 99, SEQ ID NO. 100, SEQ ID NO. 101, SEQ ID NO. 102, SEQ ID NO. 103, SEQ ID NO. 104, SEQ ID NO. 105, SEQ ID NO. 106, SEQ ID NO. 107, SEQ ID NO. 108, SEQ ID NO. 109, SEQ ID NO. 110, SEQ ID NO. 111, SEQ ID NO. 112, SEQ ID NO. 113, SEQ ID NO. 114, SEQ ID NO. 115, SEQ ID NO. 116, SEQ ID NO. 117, SEQ ID NO. 118, SEQ ID NO. 119, SEQ ID NO. 120 and SEQ ID NO. 121; or any subset SEQ ID NO. 58 through SEQ ID NO. 121. In an additional embodiment, A is an immunogenic fragment of PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121. In a particular embodiment, A is an immunogenic fragment of PilA from *H. influenzae* strain 86-028NP wherein PilA is SEQ ID NO. 58.

[80] PilA from *H. influenzae* strain 86-028NP - SEQ ID NO. 58

MKLTTQOTLK KGFTLIELMI VIAIIAILAT IAIPSYQNYT KKAHVSELLQ  
ASAPYKADVE LCVYSTNETT NCTGGKNGIA ADITTAKGYV KSVTTSNGAI  
TVKGDGTLAN MEYILQATGN AATGVTWTTT CKGTDASLFP ANFCGSVTQ

[81] In another embodiment, A is an immunogenic fragment of PilA approximately 75% to 100% identical to SEQ ID NO. 127. More specifically, in one embodiment A is SEQ ID NO. 127, a fragment consisting of amino acids 40-149 of SEQ ID NO. 58.

[82] Amino acids 40-149 of PilA from *H. influenzae* strain 86-028NP - SEQ ID NO. 127.

T KKAHVSELLQ  
ASAPYKADVE LCVYSTNETT NCTGGKNGIA ADITTAKGYV KSVTTSNGAI  
TVKGDGTLAN MEYILQATGN AATGVTWTTT CKGTDASLFP ANFCGSVTQ

[83] In another embodiment, A is an immunogenic fragment of PilA consisting of amino acids 40-149 from any of SEQ ID NO. 58 – SEQ ID NO. 121. In an additional embodiment, A is an immunogenic fragment at least 95% identical to amino acids 40-149 from any of SEQ ID NO. 58 – SEQ ID NO. 121.

[84] In one embodiment, the fusion proteins of formula (I) are defined wherein Y is selected from the group consisting of GG, SG and SS. In another embodiment, the fusion proteins of formula (I) are defined wherein Y is GG or SG. In one particular embodiment, Y is GG.

[85] In one embodiment, the fusion proteins of formula (I) are defined wherein o is 1. In another embodiment, o is 0.

[86] In one embodiment, the fusion proteins of formula (I) are defined wherein B is PilA from *H. influenzae* or an immunogenic fragment of PilA from *H. influenzae* when A is Protein E from *H. influenzae* or an immunogenic fragment of Protein E from *H. influenzae*. For example, B is PilA from *H. influenzae* strain 86-028NP. In another embodiment, B is PilA from *H. influenzae* having an amino acid sequence selected from the group consisting of SEQ ID NO. 58, SEQ ID NO. 59, SEQ ID NO. 60, SEQ ID NO. 61, SEQ ID NO. 62, SEQ ID NO. 63, SEQ ID NO. 64, SEQ ID NO. 65, SEQ ID NO. 66, SEQ ID NO. 67, SEQ ID NO. 68, SEQ ID NO. 69, SEQ ID NO. 70, SEQ ID NO. 71, SEQ ID NO. 72, SEQ ID NO. 73, SEQ ID NO. 74, SEQ ID NO. 75, SEQ ID NO. 76, SEQ ID NO. 77, SEQ ID NO. 78, SEQ ID NO. 79, SEQ ID NO. 80, SEQ ID NO. 81, SEQ ID NO. 82, SEQ ID NO. 83, SEQ ID NO. 84, SEQ ID NO. 85, SEQ ID NO. 86, SEQ ID NO. 87, SEQ ID NO. 88, SEQ ID NO. 89, SEQ ID NO. 90, SEQ ID NO. 91, SEQ ID NO. 92, SEQ ID NO. 93, SEQ ID NO. 94, SEQ ID NO. 95, SEQ ID NO. 96, SEQ ID NO. 97, SEQ ID NO. 98, SEQ ID NO. 99, SEQ ID NO. 100, SEQ ID NO. 101, SEQ ID NO. 102, SEQ ID NO. 103, SEQ ID NO. 104, SEQ ID NO. 105, SEQ ID NO. 106, SEQ ID NO. 107, SEQ ID NO. 108, SEQ ID NO. 109, SEQ ID NO. 110, SEQ ID NO. 111, SEQ ID NO. 112, SEQ ID NO. 113, SEQ ID NO. 114, SEQ ID NO. 115, SEQ ID NO. 116, SEQ ID NO. 117, SEQ ID NO. 118, SEQ ID NO. 119, SEQ ID NO. 120 and SEQ ID NO. 121; or any subset of SEQ ID NO. 58 through SEQ ID NO. 121. In another embodiment, B is PilA wherein PilA is approximately 80% to 100% identical to SEQ ID NO. 58. In another embodiment, B is PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121. In a particular embodiment, B is PilA of SEQ ID NO. 58.

[87] In another embodiment, B is PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121 and A is PE wherein PE is at least 95% identical to any of SEQ ID NO. 4 – SEQ ID NO. 57.

[88] In another embodiment, the fusion proteins of formula (I) are defined wherein B is an immunogenic fragment of PilA from *H. influenzae* when A is an immunogenic fragment of



Protein E from *H. influenzae*. For example, B is an immunogenic fragment of the PilA from *H. influenzae* strain 86-028NP. In another embodiment, B is an immunogenic fragment of PilA wherein PilA is approximately 80% to 100% identical to SEQ ID NO: 58. In another embodiment, B is an immunogenic fragment of PilA wherein PilA has an amino acid selected from the group consisting of SEQ ID NO. 58, SEQ ID NO. 59, SEQ ID NO. 60, SEQ ID NO. 61, SEQ ID NO. 62, SEQ ID NO. 63, SEQ ID NO. 64, SEQ ID NO. 65, SEQ ID NO. 66, SEQ ID NO. 67, SEQ ID NO. 68, SEQ ID NO. 69, SEQ ID NO. 70, SEQ ID NO. 71, SEQ ID NO. 72, SEQ ID NO. 73, SEQ ID NO. 74, SEQ ID NO. 75, SEQ ID NO. 76, SEQ ID NO. 77, SEQ ID NO. 78, SEQ ID NO. 79, SEQ ID NO. 80, SEQ ID NO. 81, SEQ ID NO. 82, SEQ ID NO. 83, SEQ ID NO. 84, SEQ ID NO. 85, SEQ ID NO. 86, SEQ ID NO. 87, SEQ ID NO. 88, SEQ ID NO. 89, SEQ ID NO. 90, SEQ ID NO. 91, SEQ ID NO. 92, SEQ ID NO. 93, SEQ ID NO. 94, SEQ ID NO. 95, SEQ ID NO. 96, SEQ ID NO. 97, SEQ ID NO. 98, SEQ ID NO. 99, SEQ ID NO. 100, SEQ ID NO. 101, SEQ ID NO. 102, SEQ ID NO. 103, SEQ ID NO. 104, SEQ ID NO. 105, SEQ ID NO. 106, SEQ ID NO. 107, SEQ ID NO. 108, SEQ ID NO. 109, SEQ ID NO. 110, SEQ ID NO. 111, SEQ ID NO. 112, SEQ ID NO. 113, SEQ ID NO. 114, SEQ ID NO. 115, SEQ ID NO. 116, SEQ ID NO. 117, SEQ ID NO. 118, SEQ ID NO. 119, SEQ ID NO. 120 and SEQ ID NO. 121; or any subset of SEQ ID NO. 58 through SEQ ID NO. 121. In another embodiment, B is an immunogenic fragment of PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121. In a particular embodiment, B is an immunogenic fragment of PilA from *H. influenzae* wherein PilA has the amino acid sequence set forth in SEQ ID NO. 58. In another embodiment, B is an immunogenic fragment of PilA consisting of amino acids 40-149 from any of SEQ ID NO. 58 – SEQ ID NO. 121. More specifically, in one embodiment B is the fragment of PilA as set forth in SEQ ID NO. 127. In an additional embodiment, B is an immunogenic fragment at least 95% identical to amino acids 40-149 of any of SEQ ID NO. 58 – SEQ ID NO. 121.

[89] In one particular embodiment, B is the fragment of PilA as set forth in SEQ ID NO. 127 and A is an immunogenic fragment of Protein E selected from the group consisting of SEQ ID NO. 122, SEQ ID NO. 124, SEQ ID NO. 125 and SEQ ID NO. 126. More particularly, B is the fragment of PilA as set forth in SEQ ID NO. 127 and A is the fragment of Protein E as set forth in SEQ ID NO. 124, amino acids 19-160 of Protein E from SEQ ID NO. 4. In another embodiment, B is the fragment of PilA as set forth in SEQ ID NO. 127 and A is the fragment of Protein E as set forth in SEQ ID NO. 125.

[90] In another embodiment, B is an immunogenic fragment of PilA wherein PilA is at least 95% identical to any of SEQ ID NO. 58 – SEQ ID NO. 121 and A is an immunogenic fragment of PE wherein PE is at least 95% identical to any of SEQ ID NO. 4 – SEQ ID NO. 57.

[91] In another embodiment, the fusion proteins of formula (I) are defined wherein B is Protein E from *H. influenzae* when A is PilA from *H. influenzae*. For example, B is Protein E having an amino acid sequence selected from the group consisting of SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8, SEQ ID NO. 9, SEQ ID NO. 10, SEQ ID NO. 11, SEQ ID NO. 12, SEQ ID NO. 13, SEQ ID NO. 14, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 18, SEQ ID NO. 19, SEQ ID NO. 20, SEQ ID NO. 21, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 25, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 29, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32, SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 36, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO. 39, SEQ ID NO. 40, SEQ ID NO. 41, SEQ ID NO. 42, SEQ ID NO. 43, SEQ ID NO. 44, SEQ ID NO. 45, SEQ ID NO. 46, SEQ ID NO. 47, SEQ ID NO. 48, SEQ ID NO. 49, SEQ ID NO. 50, SEQ ID NO. 51, SEQ ID NO. 52, SEQ ID NO. 53, SEQ ID NO. 54, SEQ ID NO. 55, SEQ ID NO. 56 and SEQ ID NO. 57; or any subset of SEQ ID NO. 4 through SEQ ID NO. 57. In another embodiment, the fusion proteins of formula (I) are defined wherein B is Protein E wherein Protein E is approximately 75% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO. 4. In another embodiment, B is Protein E wherein Protein E is approximately 90% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO. 4. For example, B is Protein E wherein Protein E is at least 95% identical to Protein E as set forth in SEQ ID NO. 4. In another embodiment, B is Protein E wherein Protein E is at least 95% identical to any of SEQ ID NO. 4 – SEQ ID NO. 57. In a particular embodiment, B is Protein E having the amino acid sequence set forth in SEQ ID NO. 4.

[92] In another embodiment, the fusion proteins of formula (I) are defined wherein B is an immunogenic fragment of Protein E from *H. influenzae* when A is an immunogenic fragment of PilA from *H. influenzae*. For example, B is an immunogenic fragment of Protein E wherein Protein E has an amino acid sequence selected from the group consisting of SEQ ID NO. 4, SEQ ID NO. 5, SEQ ID NO. 6, SEQ ID NO. 7, SEQ ID NO. 8, SEQ ID NO. 9, SEQ ID NO. 10, SEQ ID NO. 11, SEQ ID NO. 12, SEQ ID NO. 13, SEQ ID NO. 14, SEQ ID NO. 15, SEQ ID NO. 16, SEQ ID NO. 17, SEQ ID NO. 18, SEQ ID NO. 19, SEQ ID NO. 20, SEQ ID NO. 21, SEQ ID NO. 22, SEQ ID NO. 23, SEQ ID NO. 24, SEQ ID NO. 25, SEQ ID NO. 26, SEQ ID NO. 27, SEQ ID NO. 28, SEQ ID NO. 29, SEQ ID NO. 30, SEQ ID NO. 31, SEQ ID NO. 32,

SEQ ID NO. 33, SEQ ID NO. 34, SEQ ID NO. 35, SEQ ID NO. 36, SEQ ID NO. 37, SEQ ID NO. 38, SEQ ID NO. 39, SEQ ID NO. 40, SEQ ID NO. 41, SEQ ID NO. 42, SEQ ID NO. 43, SEQ ID NO. 44, SEQ ID NO. 45, SEQ ID NO. 46, SEQ ID NO. 47, SEQ ID NO. 48, SEQ ID NO. 49, SEQ ID NO. 50, SEQ ID NO. 51, SEQ ID NO. 52, SEQ ID NO. 53, SEQ ID NO. 54, SEQ ID NO. 55, SEQ ID NO. 56 and SEQ ID NO. 57; or any subset of SEQ ID NO. 4 through SEQ ID NO. 57. In another embodiment, the fusion proteins of formula (I) are defined wherein B is an immunogenic fragment of Protein E wherein Protein E is approximately 75% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO. 4. In another embodiment, B is an immunogenic fragment of Protein E wherein Protein E is approximately 90% to 100% identical to the Protein E amino acid sequence set forth in SEQ ID NO. 4. In a particular embodiment, B is an immunogenic fragment of Protein E having the amino acid sequence set forth in SEQ ID NO. 4. In an additional embodiment, B is an immunogenic fragment of Protein E, wherein Protein E is at least 95% identical to any of SEQ ID NO. 4 – SEQ ID NO. 57.

[93] In another embodiment, B is a fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 17-160 of SEQ ID NO. 4 (SEQ ID NO. 122), amino acids 18-160 of SEQ ID NO. 4 (SEQ ID NO. 123), amino acids 19-160 of SEQ ID NO. 4 (SEQ ID NO. 124), amino acids 20-160 of SEQ ID NO. 4 (SEQ ID NO. 125) and amino acids 22-160 of SEQ ID NO. 4 (SEQ ID NO. 126). In another embodiment, B is an immunogenic fragment of Protein E from *H. influenzae* selected from the group consisting of amino acids 17-160 of SEQ ID NO. 4 (SEQ ID NO. 122), amino acids 18-160 of SEQ ID NO. 4 (SEQ ID NO. 123), amino acids 19-160 of SEQ ID NO. 4 (SEQ ID NO. 124), amino acids 20-160 of SEQ ID NO. 4 (SEQ ID NO. 125), amino acids 22-160 of SEQ ID NO. 4 (SEQ ID NO. 126), amino acids 23-160 of SEQ ID NO. 4 (SEQ ID NO. 179) and amino acids 24-160 of SEQ ID NO. 4 (SEQ ID NO. 180). More specifically, in one embodiment, B is the fragment of Protein E as set forth in SEQ ID NO. 123, amino acids 18-160 of SEQ ID NO. 4.

[94] In one particular embodiment B is an immunogenic fragment of Protein E as set forth in SEQ ID NO. 123, amino acids 18-160 of SEQ ID NO. 4 when A is an immunogenic fragment of PIIA as set forth in SEQ ID NO. 127.

[95] In one embodiment, the fusion proteins of formula (I) are defined wherein p is 0. In another embodiment, the fusion proteins of formula (I) are defined wherein p is 1.

[96] In one embodiment, the fusion protein of formula (I) is selected from the group consisting of SEQ ID NO. 136, SEQ ID NO. 138, SEQ ID NO. 140, SEQ ID NO. 142, SEQ ID NO. 144, SEQ ID NO. 146, SEQ ID NO. 148, SEQ ID NO. 150, SEQ ID NO. 182, SEQ ID NO. 184, SEQ ID NO. 186, SEQ ID NO. 188, SEQ ID NO. 190, SEQ ID NO. 192, SEQ ID NO. 194, SEQ ID NO. 196, SEQ ID NO. 198, SEQ ID NO. 200, SEQ ID NO. 202 and SEQ ID NO. 204; or any subset thereof. In another embodiment, the fusion protein of formula (I) is approximately 95% identical to any of SEQ ID NO. 136, SEQ ID NO. 138, SEQ ID NO. 140, SEQ ID NO. 142, SEQ ID NO. 144, SEQ ID NO. 146, SEQ ID NO. 148, SEQ ID NO. 150, SEQ ID NO. 182, SEQ ID NO. 184, SEQ ID NO. 186, SEQ ID NO. 188, SEQ ID NO. 190, SEQ ID NO. 192, SEQ ID NO. 194, SEQ ID NO. 196, SEQ ID NO. 198, SEQ ID NO. 200, SEQ ID NO. 202 or SEQ ID NO. 204.

[97] Fusion proteins of formula (I) are useful as immunogens in subjects such as mammals, particularly humans. In particular, the fusion proteins of formula (I) are useful in inducing an immune response against *H. influenzae* in subjects, particularly humans. More specifically, the fusion proteins of formula (I) are useful in the treatment or prevention of otitis media and/or AECOPD and/or pneumonia.

[98] The present invention relates to immunogenic compositions comprising Protein E from *H. influenzae* (or an immunogenic fragment thereof) and PilA from *H. influenzae* (or an immunogenic fragment thereof), and immunogenic compositions comprising fusion proteins of Protein E from *H. influenzae* (or an immunogenic fragment thereof) and PilA from *H. influenzae* (or an immunogenic fragment thereof). The present invention also relates to vaccines comprising such immunogenic compositions and therapeutic uses of the same.

[99] In one embodiment, the immunogenic compositions comprise Protein E from *H. influenzae* (or an immunogenic fragment thereof) and PilA from *H. influenzae* (or an immunogenic fragment thereof). Protein E may be SEQ ID NO. 4 or a Protein E sequence at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO. 4.

[100] The immunogenic fragment of Protein E may be SEQ ID NO. 122, SEQ ID NO. 123, SEQ ID NO. 124, SEQ ID NO. 125 or SEQ ID NO. 126, or a sequence having at least 90%, 95%, 96%, 97%, 98%, 99% sequence identity to any one of SEQ ID NO. 122, SEQ ID NO. 123, SEQ ID NO. 124, SEQ ID NO. 125 or SEQ ID NO. 126. The immunogenic fragment of Protein E may be SEQ ID NO. 122, SEQ ID NO. 123, SEQ ID NO. 124, SEQ ID NO. 125,

SEQ ID NO. 126, SEQ ID NO. 179 or SEQ ID NO. 180 or a sequence having at least 90%, 95%, 96%, 97%, 98%, 99% sequence identity to any one of SEQ ID NO. 122, SEQ ID NO. 123, SEQ ID NO. 124, SEQ ID NO. 125, SEQ ID NO. 126, SEQ ID NO. 179 or SEQ ID NO. 180. Amino acid differences have been described in Protein E from various *Haemophilus* species when compared to Protein E from *Haemophilus influenzae* Rd as a reference strain. Microbes & Infection (Corrigendum to "Identification of a novel *Haemophilus influenzae* protein important for adhesion to epithelia cells" [Microbes Infect. 10 (2008) 87-97], available online July 6, 2010, "Article in Press") provides a sequence for Protein E from *H. influenzae* strain 772. WO2002/28889 provides a sequence for Protein E from *H. influenzae* strain 12085.

[101] Protein E contains an epithelial cell binding region (PKRYARSVRQ YKILNCANYH LTQVR, SEQ ID NO. 128) that has been reported to be conserved among more than 100 clinical NTHi isolates, encapsulated *H. influenzae*, and culture collection strains analyzed (Singh et al, J. Infect. Dis. 201(3):414-9 (2010)). Singh et al. reported that Protein E was highly conserved in both NTHi and encapsulated *H. influenzae* (96.9% – 100% identity without the signal peptide). In one embodiment, the fragment of Protein E comprises the binding region of SEQ ID NO. 128 (PKRYARSVRQ YKILNCANYH LTQVR).

[102] PilA is a conserved adhesin expressed *in vivo*. Full length comparison of 64 sequences of PilA from *Haemophilus influenzae* demonstrated approximately 80% to 100% identity.

[103] In another embodiment, the immunogenic composition comprises a fusion protein as defined by formula (I).

[104] In one embodiment, the present immunogenic compositions may be administered with other antigens from *H. influenzae*. For example, the PE and PilA or the fusion protein of formula (I) may be administered with Protein D from *H. influenzae*. Protein D may be as described in WO91/18926. In another embodiment, the immunogenic composition may include the fusion protein of formula (I) and Protein D from *H. influenzae*.

[105] In another embodiment, the immunogenic compositions of the invention may be administered with additional antigens from other bacterial species also known to cause otitis media, AECOPD or pneumonia.

[106] The amount of the immunogenic composition which is required to achieve the desired therapeutic or biological effect will depend on a number of factors such as the use for which it

is intended, the means of administration, the recipient and the type and severity of the condition being treated, and will be ultimately at the discretion of the attendant physician or veterinarian. In general, a typical dose for the treatment of a condition caused in whole or in part by *H. influenzae* in a human, for instance, may be expected to lie in the range of from about 0.003 mg to about 0.090 mg. More specifically, a typical dose for the treatment of a condition caused wholly or in part by *H. influenzae* in a human may lie in the range of from about 0.01 mg to about 0.03 mg of fusion protein. The immunogenic composition may contain additional antigens; a typical dose for the treatment of a condition caused wholly or in part by *H. influenzae* in a human may lie in the range of from about 0.01 mg to about 0.03 mg for each additional antigen. This dose may be administered as a single unit dose. Several separate unit doses may also be administered. For example, separate unit doses may be administered as separate priming doses within the first year of life or as separate booster doses given at regular intervals (for example, every 1, 5 or 10 years).

[107] Formulations comprising the immunogenic compositions of the invention may be adapted for administration by an appropriate route, for example, by the intramuscular, sublingual, transcutaneous, intradermal or intranasal route. Such formulations may be prepared by any method known in the art.

[108] The immunogenic compositions of the present invention may additionally comprise an adjuvant. When the term "adjuvant" is used in this specification, it refers to a substance that is administered in conjunction with the immunogenic composition to boost the patient's immune response to the immunogenic component of the composition.

[109] Suitable adjuvants include an aluminum salt such as aluminum hydroxide gel or aluminum phosphate or alum, but may also be a salt of calcium, magnesium, iron or zinc, or may be an insoluble suspension of acylated tyrosine, or acylated sugars, cationically or anionically derivatized saccharides, or polyphosphazenes. In one embodiment, the fusion protein, PE or PiA may be adsorbed onto aluminium phosphate. In another embodiment, the fusion protein, PE or PiA may be adsorbed onto aluminium hydroxide. In a third embodiment, alum may be used as an adjuvant.

[110] Suitable adjuvant systems which promote a predominantly Th1 response include: non-toxic derivatives of lipid A, Monophosphoryl lipid A (MPL) or a derivative thereof, particularly 3-de-O-acylated monophosphoryl lipid A (3D-MPL) (for its preparation see GB 2220211 A); and

a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with either an aluminum salt (for instance aluminum phosphate or aluminum hydroxide) or an oil-in-water emulsion. In such combinations, antigen and 3D-MPL are contained in the same particulate structures, allowing for more efficient delivery of antigenic and immunostimulatory signals. Studies have shown that 3D-MPL is able to further enhance the immunogenicity of an alum-adsorbed antigen (Thoelen *et al.* Vaccine (1998) 16:708-14; EP 689454-B1).

[111] AS01 is an Adjuvant System containing MPL (3-O-desacyl-4'-monophosphoryl lipid A), QS21 (*Quillaja saponaria* Molina, fraction 21) Antigenics, New York, NY, USA) and liposomes. AS01B is an Adjuvant System containing MPL, QS21 and liposomes (50 µg MPL and 50 µg QS21). AS01E is an Adjuvant System containing MPL, QS21 and liposomes (25 µg MPL and 25 µg QS21). In one embodiment, the immunogenic composition or vaccine comprises AS01. In another embodiment, the immunogenic composition or vaccine comprises AS01B or AS01E. In a particular embodiment, the immunogenic composition or vaccine comprises AS01E.

[112] AS03 is an Adjuvant System containing α-Tocopherol and squalene in an oil/water (o/w) emulsion. AS03<sub>A</sub> is an Adjuvant System containing α-Tocopherol and squalene in an o/w emulsion (11.86 mg tocopherol). AS03<sub>B</sub> is an Adjuvant System containing α-Tocopherol and squalene in an o/w emulsion (5.93 mg tocopherol). AS03<sub>C</sub> is an Adjuvant System containing α-Tocopherol and squalene in an o/w emulsion (2.97 mg tocopherol). In one embodiment, the immunogenic composition or vaccine comprises AS03.

[113] AS04 is an Adjuvant System containing MPL (50 µg MPL) adsorbed on an aluminum salt (500 µg Al<sup>3+</sup>). In one embodiment, the immunogenic composition or vaccine comprises AS04.

[114] A system involving the use of QS21 and 3D-MPL is disclosed in WO 94/00153. A composition wherein the QS21 is quenched with cholesterol is disclosed in WO 96/33739. An additional adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil in water emulsion is described in WO 95/17210. In one embodiment the immunogenic composition additionally comprises a saponin, which may be QS21. The formulation may also comprise an oil in water emulsion and tocopherol (WO 95/17210). Unmethylated CpG containing oligonucleotides (WO 96/02555) and other immunomodulatory oligonucleotides (WO 0226757

and WO 03507822) are also preferential inducers of a TH1 response and are suitable for use in the present invention.

[115] Additional adjuvants are those selected from the group of metal salts, oil in water emulsions, Toll like receptor agonists, (in particular Toll like receptor 2 agonist, Toll like receptor 3 agonist, Toll like receptor 4 agonist, Toll like receptor 7 agonist, Toll like receptor 8 agonist and Toll like receptor 9 agonist), saponins or combinations thereof.

[116] The present invention provides a process for preparing an immunogenic composition comprising combining a fusion protein of formula (I) with an adjuvant.

[117] The present invention further provides a vaccine containing an immunogenic composition of the invention and a pharmaceutically acceptable excipient.

[118] Possible excipients include arginine, pluronic acid and/or polysorbate. In a preferred embodiment, polysorbate 80 (for example, TWEEN<sup>®</sup> 80) is used. In a further embodiment, a final concentration of about 0.03% to about 0.06% is used. Specifically, a final concentration of about 0.03%, 0.04%, 0.05% or 0.06% polysorbate 80 (w/v) may be used.

[119] The present invention provides a process for preparing an immunogenic composition or vaccine comprising combining a fusion protein of formula (I) with a pharmaceutically acceptable excipient.

[120] The present invention also provides nucleic acids encoding the proteins of the invention. The term "nucleic acid" refers to a polymeric form of nucleotides. Nucleotides can be ribonucleotides, deoxyribonucleotides, or modified forms of either ribonucleotides or deoxyribonucleotides. The term includes single and double forms of DNA. The nucleic acids are preferably substantially free from other nucleic acids.

[121] The present invention provides a process of producing nucleic acids of the invention. Nucleic acids of the invention may be prepared by methods known by those skilled in the art. For example, the nucleic acids of the invention may be synthesized in part or in whole. The nucleic acids may be prepared by digesting longer amino acids or joining shorter amino acids.

[122] The following examples are intended for illustration only and are not intended to limit the scope of the invention in any way.



[123] In the examples, the following terms have the designated meaning:

6xhis = six histidines;

xg = centrifugal force (number gravities)

ATP = adenosine triphosphate;

BCA = bicinchoninic acid;

BSA = bovine serum albumin;

°C = degrees Celsius;

CaCl<sub>2</sub> = calcium chloride;

CV = column volume;

DNA = deoxyribonucleic acid;

DSC = differential scanning calorimetry;

DTT = dithiothreitol;

dNTP = deoxynucleoside triphosphate;

EDTA = ethylenediaminetetraacetic acid;

FT = flow through;

HCl = hydrogen chloride;

His = his = histidine;

HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid;

IMAC = immobilized metal affinity chromatography;

IPTG = isopropyl β-D-1-thiogalactopyranoside;

KCl = potassium chloride;

K<sub>2</sub>HPO<sub>4</sub> = dibasic potassium phosphate;

KH<sub>2</sub>PO<sub>4</sub> = monobasic potassium phosphate;

LDS = lithium dodecyl sulfate;

L = liter;

MES = 2-(*N*-morpholino)ethanesulfonic acid;

MgCl<sub>2</sub> = magnesium chloride;

ml = milliliter;

RPM = revolutions per minute;

min = minute;

mM = millimolar;

μL = microliter;

NaCl = sodium chloride;

Na<sub>2</sub>HPO<sub>4</sub> = dibasic sodium phosphate;

$\text{NaH}_2\text{PO}_4$  = monobasic sodium phosphate;

ng = nanogram;

nm = nanometer;

O/N = overnight;

PBS = phosphate buffered saline;

PCR = polymerase chain reaction;

SB = sample buffer;

sec = second;

w/v = weight/volume.

## [124] EXAMPLES

### Example 1: Fusion proteins

Fusion proteins were produced with different signal peptides and amino acid linker sequences. These fusion proteins allowed for secretion of both Protein E and PilA (or fragments thereof) without being restricted to a single bacterial strain. The fusion protein is released into the periplasm after removal of the heterologous signal peptide by a signal peptide peptidase. Fusion protein purified from the bacteria does not contain the heterologous signal peptide. "Purified" proteins are removed from the bacteria and lack the signal peptide.

The following table describes fusion protein constructs made.

Table 3: Fusion Protein Constructs containing PilA and Protein E.

| Construct ID | N-terminal -----C-Terminal |                                                                     |                                                                     |                                                                     |        |                                                                    |                                                                  |     |                      |         |
|--------------|----------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|--------|--------------------------------------------------------------------|------------------------------------------------------------------|-----|----------------------|---------|
| LVL312       | flgI<br>sp                 | E                                                                   | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127)    |                                                                     | G<br>G | ProtE fragment<br>(A.A.: 18 to 160 of SEQ ID NO. 4, SEQ ID NO.123) |                                                                  |     | GGHHHHHHH            |         |
| A.A.         | 1                          | 19                                                                  | 21                                                                  |                                                                     |        | 130                                                                | 133                                                              |     | 275                  | 276 283 |
| LVL291       | pelB sp                    |                                                                     | ProtE fragment<br>(A.A.: 19 to 160 of SEQ ID NO. 4, SEQ ID NO. 124) |                                                                     |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GGHHHHHHH            |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 23                                                                  |        |                                                                    | 164                                                              | 167 | 276                  | 277 284 |
| LVL268       | pelB sp                    |                                                                     | D                                                                   | ProtE fragment<br>(A.A.: 20 to 160 of SEQ ID NO. 4, SEQ ID NO. 125) |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GGHHHHHHH            |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 24                                                                  |        |                                                                    | 164                                                              | 167 | 276                  | 277 284 |
| LVL269       | nadA sp                    |                                                                     | AT<br>ND<br>DD                                                      | ProtE fragment<br>(A.A.: 22 to 160 of SEQ ID NO. 4, SEQ ID NO. 126) |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GGHHHHHH<br>H        |         |
| A.A.         | 1                          |                                                                     | 23                                                                  | 24-29                                                               | 30     |                                                                    | 168                                                              | 171 | 280                  | 281 288 |
| LVL270       | MH<br>HH<br>HH<br>H        | ProtE fragment<br>(A.A.: 17 to 160 of SEQ ID NO. 4, SEQ ID NO. 122) |                                                                     |                                                                     | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127)   |                                                                  |     |                      |         |
| A.A.         | 1                          | 7                                                                   | 8                                                                   |                                                                     |        | 151                                                                | 154                                                              |     | 263                  |         |
| LVL315       | pelB<br>sp                 |                                                                     | M<br>D                                                              | ProtE fragment<br>(A.A.: 22 to 160 of SEQ ID NO. 4, SEQ ID NO. 126) |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GGHHHHHHH            |         |
|              |                            |                                                                     | 1                                                                   | 22                                                                  | 25     |                                                                    | 163                                                              | 166 | 275                  | 276 283 |
| LVL317       | pelB sp                    |                                                                     | ProtE fragment<br>(A.A.: 19 to 160 of SEQ ID NO. 4, SEQ ID NO. 124) |                                                                     |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     |                      |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 23                                                                  |        |                                                                    | 164                                                              | 167 | 276                  |         |
| LVL318       | pelB sp                    |                                                                     | M<br>D                                                              | ProtE fragment<br>(A.A.: 22 to 160 of SEQ ID NO. 4, SEQ ID NO. 126) |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     |                      |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 25                                                                  |        |                                                                    | 163                                                              | 166 | 275                  |         |
| LVL702       | pelB sp                    |                                                                     | ProtE fragment<br>(A.A.: 20 to 160 of SEQ ID NO. 4, SEQ ID NO. 125) |                                                                     |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GGHHHHHHH            |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 23                                                                  |        |                                                                    | 163                                                              | 166 | 275                  | 283     |
| LVL736       | pelB sp                    |                                                                     | ProtE fragment<br>(A.A.: 17 to 160 of SEQ ID NO. 4, SEQ ID NO. 122) |                                                                     |        | G<br>G                                                             | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) |     | GG<br>HH<br>HH<br>HH |         |
| A.A.         | 1                          |                                                                     | 22                                                                  | 23                                                                  |        |                                                                    | 166                                                              | 169 | 278                  | 286     |
| LVL737       | pelB sp                    |                                                                     | ProtE fragment<br>(A.A.: 18 to 160 of SEQ ID NO. 4, SEQ ID          |                                                                     |        | G                                                                  | PilA fragment<br>(A.A.: 40-149 of SEQ ID NO. 58,                 |     | GGH<br>HHH           |         |

|               |         |                                                                        |        |                                                                        |           |
|---------------|---------|------------------------------------------------------------------------|--------|------------------------------------------------------------------------|-----------|
|               |         | NO.123)                                                                | G      | SEQ ID NO. 127)                                                        | HH        |
| A.A.          | 1       | 22 23                                                                  | 165    | 168                                                                    | 277 285   |
| <b>LVL738</b> | pelB sp | ProtE fragment<br>(A.A.: 22 to 160 of SEQ ID NO. 4,<br>SEQ ID NO. 126) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO.<br>127) | GGHHHHHHH |
| A.A.          | 1       | 22 23                                                                  | 161    | 164                                                                    | 273 281   |
| <b>LVL739</b> | pelB sp | ProtE fragment<br>(A.A.: 23 to 160 of SEQ ID NO. 4,<br>SEQ ID NO. 179) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ<br>ID NO. 58, SEQ ID<br>NO. 127) | GGHHHHHHH |
| A.A.          | 1       | 22 23                                                                  | 160    | 163                                                                    | 272 280   |
| <b>LVL740</b> | pelB sp | ProtE fragment<br>(A.A.: 24 to 160 of SEQ ID NO.<br>4, SEQ ID NO. 180) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ<br>ID NO. 58, SEQ ID<br>NO. 127) | GGHHHHHHH |
| A.A.          | 1       | 22 23                                                                  | 159    | 162                                                                    | 271 279   |
| <b>LVL735</b> | pelB sp | ProtE fragment<br>(A.A.: 20 to 160 of SEQ ID NO. 4, SEQ ID<br>NO. 125) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 163    | 166                                                                    | 275       |
| <b>LVL778</b> | pelB sp | ProtE fragment<br>(A.A.: 17 to 160 of SEQ ID NO. 4, SEQ ID<br>NO. 122) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 166    | 169                                                                    | 278       |
| <b>LVL779</b> | pelB sp | ProtE fragment<br>(A.A.: 18 to 160 of SEQ ID NO. 4, SEQ ID<br>NO.123)  | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 165    | 168                                                                    | 277       |
| <b>LVL780</b> | pelB sp | ProtE fragment<br>(A.A.: 22 to 160 of SEQ ID NO. 4, SEQ ID<br>NO. 126) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 161    | 164                                                                    | 273       |
| <b>LVL781</b> | pelB sp | ProtE fragment<br>(A.A.: 23 to 160 of SEQ ID NO. 4, SEQ ID<br>NO. 179) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 160    | 163                                                                    | 272       |
| <b>LVL782</b> | pelB sp | ProtE fragment<br>(A.A.: 24 to 160 of SEQ ID NO. 4, SEQ ID<br>NO. 180) | G<br>G | PilA fragment<br>(A.A.: 40-149 of SEQ ID<br>NO. 58, SEQ ID NO. 127)    |           |
| A.A.          | 1       | 22 23                                                                  | 159    | 162                                                                    | 271       |

sp = signal peptide; A.A. = amino acid

The DNA and amino acid sequences for each of the signal peptides and plasmids listed in Table 3 are set forth below.

SIGNAL SEQUENCES:

*pelB* signal peptide (DNA) – SEQ ID NO. 129:

atgaaatacctgctgccgaccgctgctgctggctgctgctcctcgctgccagccggcgatggcc

*pelB* signal peptide (Amino Acid) - SEQ ID NO. 130:

MKYLLPTAAA GLLLLAAQPA MA

*FlgI* signal peptide (DNA) - SEQ ID NO. 131:

atgattaaatttctctctgcattaattcttctactgggtcacgacggcggtcaggct

*FlgI* signal peptide (Amino Acid) - SEQ ID NO. 132:

MIKFLSALIL LLVTTAAQA

*NadA* signal peptide (DNA) - SEQ ID NO. 133:

atgaaacactttccatccaaagtactgaccacagccatccttgccactttctgtagcggcgactggca

*NadA* signal peptide (Amino Acid) - SEQ ID NO. 134:

MKHFPKVL TAILATFCSG ALA

FUSION PROTEIN CONSTRUCT SEQUENCES:

The single underlined portion of the amino acid sequences is from PilA from *Haemophilus influenzae* strain 86-028NP. The emboldened underlined portion of the amino acid sequences was derived from Protein E from *Haemophilus influenza* strain 772.

LVL312 (DNA) - SEQ ID NO. 135:

atgattaaatttctctctgcattaattcttctactgggtcacgacggcggtcaggctgagactaaaaaagcagcggtatctgaattactg  
caagcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatg  
gtattgcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggat  
ggcacattggcaaatatggaatatatttgcaagctacaggtaatgctgcaacaggtgtaacttggaacaacaactgcaaaggaac  
ggatgcctcttatttccagcaaaattttgcggaagtgtcacacaaggcggcgcgcagattcagaaggctgaacaaaatgatgtgaa  
gctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgaaatcgatctgggtg

gataaccaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacggttatgca  
 cgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggggacagggg  
 ttgcgggcagcacctaaaaagcaaaagaaacatacgttaagttaacacctgatacaacgctttataatgctgctcagattattgtg  
 cgaactatggtgaagcattttcagttgataaaaaaggcgccaccaccaccaccactaa

*LVL312 (protein):* (flgI sp)(E)(PilA aa 40-149)(GG)(ProtE aa 18-160)(GGHHHHHH) - SEQ ID NO. 136

MIKFLSALIL LLVTTAAQAE TKKAAVSELL QASAPYKADV ELCVYSTNET  
TNCTGGKNGI AADITTAKGY VKSVTTSNGA ITVKGDGTLA NMEYILQATG  
NAATGVTWTT TCKGTDASLF PANFCGSVTQ GGAQIQKAEQ NDVKLAPPTD  
VRSGYIRLVK NVNYYIDSES IWVDNQEPQI VHFDAVVNLD KGLYVYPEPK  
RYARSVRQYK ILCANYHLT QVRTDFYDEF WGOGLRAAPK QKKHTLSLT  
PDTTLYNAAQ IICANYGEAF SVDKKGHHH HHH

*LVL291 (DNA) - SEQ ID NO. 137:*

atgaaatacctgctgccgaccgctgctggtctgctgctcctgctgccagccggcgatggcccagattcagaaggctgaaca  
 aatgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagta  
 atcgatctgggtggataaccaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcc  
 taaacggtatgcacgttctgtcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatt  
 tggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgttaagttaacacctgatacaacgctttataatgctgc  
 tcagattatttgcgaactatggtgaagcattttcagttgataaaaaaggcgccactaaaaagcagcggatctgaattactgcaa  
 gcgtcagcgccttataaggctgatgtgaattatgtgtatatagcacaatgaaacaacaaactgtacgggtggaaaaaatggtatt  
 gcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggc  
 acattggcaaatatggaatatatttgaagctacaggtaatgctgcaacaggtgtaacttggaacaactgcaaaggaacgga  
 tgccctcttattccagcaaattttgcggaagtgtcacacaaggcgccaccaccaccaccactaa

*LVL291 (Protein)(pelB sp)(ProtE aa 19-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 138*

MKYLLPTAAA GLLLLAAQPA MAQIQKAEQN DVKLAPPTDV RSGYIRLVKN  
VNYYIDSESI WVDNQEPQIV HFDAVVNLDK GLYVYPEPKR YARSVRQYKI  
ILNCANYHLTQ VRTDFYDEFW GOGLRAAPKK QKKHTLSLTP DTTLYNAAQI  
ICANYGEAFS VDKKGGTKKA AVSELLQASA PYKADVELCV YSTNETTNCT

GGKNGIAADI TTAKGYVKSV TTSNGAITVK GDGTLANMEY ILQATGNAAT  
GVTWTTTCKG TDASLFPANF CGSVTQGGHH HHHH

*LVL268 (DNA) - SEQ ID NO. 139:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccgatattcagaaggctgaaca  
aatgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtga  
atcgatctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcc  
taaacgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatt  
tggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagtttaacacctgatacaacgctttataatgctgc  
tcagattattgtgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaattactgcaa  
gcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtatt  
gcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgaataacagtaaaaggggatggc  
acattggcaaatatggaatatattttgcaagctacaggaatgtctgcaacaggtgtaacttggacaacaactgcaaaggaacgga  
tgctctttatttcagcaaattttgcggaagtgtcacacaaggcggccaccaccaccaccac

*LVL268 (protein): (pelB sp)(D)(ProtE aa 20-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 140:*

MKYLLPTAAA GLLLLAAQPA MADIQKAEQN DVKLAPPTDV RSGYIRLVKN  
VNYYIDSESI WVDNQEPQIV HFDVAVNLDK GLYVYPEPKR YARSVRQYKI  
LNCANYHLTQ VRTDFYDEFW GQGLRAAPKK QKKHTLSLTP DTTLYNAAQI  
ICANYGEAFS VDKKGGTKKA AVSELLQASA PYKADVELCV YSTNETTNCT  
GGKNGIAADI TTAKGYVKSV TTSNGAITVK GDGTLANMEY ILQATGNAAT  
GVTWTTTCKG TDASLFPANF CGSVTQGGHH HHHH

*LVL269 (DNA) - SEQ ID NO. 141:*

atgaaacactttccatccaaagtactgaccacagccatccttgccactttctgtagcggcgactggcagccacaaacgacgacg  
ataaggctgaacaaaatgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattatt  
acatcgatagtgaatcgatctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtat  
gtttatcctgagcctaaacgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgat  
ttctatgatgaattttggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagtttaacacctgatacaac  
gctttataatgctgctcagattattgtgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtat  
ctgaattactgcaagcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtg



gaaaaaatggtattgcagcagatataaccacagcaaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagta  
aaaggggatggcacattggcaaatatggaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggacaacaacttg  
caaaggaacggatgcctctttattccagcaaattttgcggaagtgtcacacaaggcggccaccaccaccaccactaa

LVL269 (protein): (nadA sp)(ATNDDD)(ProtE aa 22-160)(GG)(PilA aa 40-149)(GGHHHHHH)  
- SEQ ID NO.142

MKHFPSKVLT TAILATFCSG ALAATNDDDK AEQNDVKLAP PTDVRSYIR  
LVKNVNYID SESIWVDNQE PQIVHFDAVV NLDKGLYVYP EPKRYARSVR  
QYKILNCANY HLTQVRTDFY DEFWGOGLRA APKKQKKHTL SLTPDTTLYN  
AAQIICANYG EAFSVDKKG TKKAHVSELL QASAPYKADV ELCVYSTNET  
TNCTGGKNGI AADITTAKGY VKSVTTSNGA ITVKGDTLA NMEYILQATG  
NAATGVTWTT TCKGTDASLF PANFCGSVTQ GGHHHHHH

LVL270 (DNA) - SEQ ID NO. 143:

atgcaccaccaccaccacagcgcgagattcagaaggctgaacaaaatgatgtgaagctggcaccgccgactgatgtacg  
aagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgatctgggtggataaccaagagccacaaaattgta  
cattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacggtatgcacgttctgtctcagtataagatcttg  
aattgtgcaaattatcatttaactcaagtagaactgatttctatgatgaattttggggacagggttgcgggcagcacctaaaaagca  
aaagaaacatacgttaagttaacacctgatacaacgctttataatgctgctcagattatttgtgcgaactatggtgaagcattttcagtt  
gataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgtcagcgccttataaggctgatgtggaattatgtgt  
atatagcaciaaatgaaacaacaaactgtacgggtggaaaaaatggtattgcagcagatataaccacagcaaaaaggctatgtaa  
aatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggcacattggcaaatatggaatatatttgaagctacag  
gtaatgctgcaacaggtgtaacttggacaacaacttgcaaaggaacggatgcctctttattccagcaaattttgcggaagtgtcac  
acaataa

LVL270 (protein): (MHHHHHH)(ProtE aa 17-160)(GG)(PilA aa40-149) - SEQ ID NO. 144:

MHHHHHHSQA IQKAEQNDVK LAPPTDVRSY YIRLVKNVNY YIDSESIWVD  
NOEPQIVHFD AVVNLDKGLY VYPEPKRYAR SVROYKILNC ANYHLTQVRT  
DFYDEFWGOGLRAAPKKQKK HTLSLTPDTT LYNAAQIICA NYGEAFSVDK  
KGGTKKAHVS ELLQASAPYK ADVELCVYST NETTNCTGGK NGIAADITTA  
KGYVKSVTTS NGAITVKGDTLANMEYILQ ATGNAATGVT WTTTCKGTDA  
SLFPANFCGS VTQ

*LVL315 (DNA) - SEQ ID NO. 145:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccatggataaggctgaacaaaa  
tgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgaatcg  
atctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaa  
cgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatttggg  
gacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataatgctgctcag  
attatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggatctgaattactgcaagcgt  
cagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtattgca  
gcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggcacat  
tggcaaatatggaatatatttgcagctacaggtaatgctgcaacaggtgtaactggacaacaactgcaaaggaacggatgcc  
tctttattccagcaaattttgcggaagtgtcacacaaggcggccaccaccaccaccactaa

*LVL315 (protein): (pelB sp)(MD)(ProtE aa 22-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 146:*

MKYLLPTAAA GLLLLAAQPA MAMDKAEQND VKLAPPTDVR SGYIRLVKNV  
NYIDSESIW VDNQEPQIVH FDAVVNLDKG LYVYPEPKRY ARSVROYKIL  
NCANYHLTOV RTDFYDEFWG OGLRAAPKKQ KKHTLSLTPD TTYNAAQII  
CANYGEAFSV DKKGGTKKAA VSELLQASAP YKADVELCVY STNETTNCTG  
GKNGIAADIT TAKGYVKSVT TSNGAITVKG DGTLANMEYI LQATGNAATG  
VTWTTTCKGT DASLFPANFC GSVTQGGHHH HHH

*LVL317 (DNA) - SEQ ID NO. 147:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccagattcagaaggctgaaca  
aatgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtga  
atcgatctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcc  
taaacgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatt  
tggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataatgctgc  
tcagattatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggatctgaattactgcaa  
gcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtatt  
gcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggc

acattggcaaatatggaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggacaacaacttgcaaaggaacgga  
tgcctctttatttccagcaaatttttgcggaaggtcacacaataa

*LVL317 (protein):* (pelB sp)(ProtE aa 19-160)(GG)(PilA aa40-149) - SEQ ID NO. 148:

MKYLLPTAAA GLLLLAAQPA MAQIQKAEQN DVKLAPPTDV RSGYIRLVKN  
VNYYIDSESI WVDNQEPQIV HFDAVVNLDK GLYVYPEPKR YARSVROYKI  
LNCANYHLTQ VRTDFYDEFW GQGLRAAPKK OKKHTLSLTP DTTYNAAOI  
ICANYGEAFS VDKKGGTKKA AVSELLQASA PYKADVELCV YSTNETTNCT  
GGKNGIAADI TTAKGYVKSV TTSNGAITVK GDGTLANMEY ILQATGNAAT  
GVTWTTTCKG TDASLFPANF CGSVTQ

*LVL318 (DNA) - SEQ ID NO. 149:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccatggataaggctgaacaaaa  
tgatgtgaagctggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgatcg  
atctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaattagataagggattgtatgtttatcctgagcctaaa  
cgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggg  
gacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgttaagttaacacctgatacaacgctttataatgctgctcag  
attatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaattactgaagcgt  
cagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtattgca  
gcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgaataacagtaaaaggggatggcacat  
tggcaaatatggaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggacaacaacttgcaaaggaacgggatgcc  
tctttatttccagcaaatttttgcggaaggtcacacaataa

*LVL318 (protein):* (pelB sp)(MD)(ProtE aa 22-160)(GG)(PilA aa40-149) - SEQ ID NO. 150:

MKYLLPTAAA GLLLLAAQPA MAMDKAEQND VKLAPPTDVR SGYIRLVKNV  
NYIDSESIW VDNQEPQIVH FDAVVNLDKG LYVYPEPKRY ARSVROYKIL  
NCANYHLTQV RTDFYDEFWG QGLRAAPKKQ KKHTLSLTPD TTYNAAOII  
CANYGEAFSV DKKGGTKKAA VSELLQASAP YKADVELCVY STNETTNCTG  
GKNGIAADIT TAKGYVKSVT TSNNGAITVKG DGTLANMEYI LQATGNAATG  
VTWTTTCKGT DASLFPANFC GSVTQ

*LVL702 (DNA) - SEQ ID NO. 181:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccattcagaaggctgaacaaaa  
tgatgtgaagctggcaccgcccactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgaatcg  
atctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaa  
cgttatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatttggg  
gacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataatgctgctcag  
attatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaattactgcaagcgt  
cagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtattgca  
gcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggcacat  
tggcaaatatggaatatatttgcagctacaggtaatgctgcaacaggtgtaactggacaacaactgcaaaggaacggatgcc  
tctttattccagcaaattttgcggaagtgtcacacaaggcggccaccaccaccaccaccac

*LVL702 (protein): (pelB sp)(ProtE aa 20-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 182:*

MKYLLPTAAA GLLLLAAQPA MAIQKAEQND VKLAPPTDVR SGYIRLVKNV  
NYIIDSESIW VDNQEPQIVH FLAVVNLDKG LYVYPEPKRY AREVROYKIL  
NCANYHLLTQV RTDFYDEFWG QGLRAAPKKQ KKHTLSLTPD TTYNAAQII  
CANYGEAFSV DKKGGTKKAA VSELLQASAP YKADVELCVY STNETTNCTG  
GKNGIAADIT TAKGYVKSVT TSNGAITVKG DGTLANMEYI LQATGNAATG  
VTWTTTCKGT DASLFPANFC GSVTQGGHHH HHH

*LVL736 (DNA) - SEQ ID NO. 183:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccagcggccagattcagaaggc  
tgaacaaaatgatgtgaagctggcaccgcccactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcga  
tagtgaatcgatctgggtggataaccaagagccacaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcct  
gagcctaaacggtatgcacgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgat  
gaatttggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataa  
tgctgctcagattatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaatta  
ctgcaagcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaa  
atggtattgcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaagg

gatggcacattggcaaatatggaatatatttgaagctacaggaatgctgcaacaggtgaactggacaacaactgcaaagg  
aacggatgcctctttattccagcaaatatttgcggaagtgtcacacaaggcgccaccaccaccaccac

*LVL736 (protein): (pelB sp)(ProtE aa 17-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 184:*

MKYLLPTAAA GLLLLAAQPA MASAQIQKAE QNDVKLAPPT DVRSGYIRLV  
KNVNYIIDSE SIWVDNQEPQ IVHFDAVVNL DKGLYVYPEP KRYARSVROY  
KILNCANYHL TQVRTDFYDE FWGOGLRAAP KKQKKHTLSL TPDITLYNAA  
QIICANYGEA FSVDKKGGTK KAAVSELLQA SAPYKADVEL CVYSTNETTN  
CTGGKNGIAA DITTAKGYVK SVTTSNGAIT VKGDGTLANM EYILQATGNA  
ATGVTWTTTC KGTDASLFPA NFCGSVTQGG HHHHHH

*LVL737 (DNA) - SEQ ID NO. 185:*

atgaaatacctgctgccgaccgctgctggtctgctgctcctcgctgccagccggcgatggccgcccagattcagaaggctga  
acaaaatgatgtgaagctggcaccgcccactgatgtacgaagcgatatatcgtttgtaagaatgtgaattattacatcgatg  
tgaatcgatctgggtggataaccaagagccacaaaattgtacattttgatgcagtggatgaattagataagggattgtatgtttatcctga  
gcctaaacggtatgcacgttctgttcgtcagtataagatctgaattgtgcaattatcatttaactcaagtacgaactgatttctatgatga  
attttggggacaggggttgcgggcagcacctaaaaagcaaaagaaacatacgttaagttaaacctgatacaacgctttataatgc  
tgctcagattatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaattactg  
caagcgtcagcgcttataaggctgatgtggaattatgtgtatatagcacaatgaaacaacaaactgtacgggtggaaaaaatg  
gtattgcagcagatataaccacagcaaaaggctatgtaaaatcagtgaacaacgaacggtgcaataacagtaaaaggggat  
ggcacattggcaaatatggaatatatttgaagctacaggaatgctgcaacaggtgaactggacaacaacttgcaaaggaac  
ggatgcctctttattccagcaaatatttgcggaagtgtcacacaaggcgccaccaccaccaccac

*LVL737 (protein): (pelB sp)(ProtE aa 18-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 186:*

MKYLLPTAAA GLLLLAAQPA MAAQIQKAEQ NDVKLAPPTD VRSGYIRLVK  
NVNYIIDSES IWVDNQEQI VHFDAVVNLD KGLYVYPEPK RYARSVROYK  
ILNCANYHLT QVRTDFYDEF WGOGLRAAPK KQKKHTLSLT PDIITLYNAAQ  
IICANYGEAF SVDKKGGTKK AAVSELLQAS APYKADVELC VYSTNETTNC  
TGGKNGIAAD ITTAKGYVKS VTTSNGAITV KGDGTLANME YILQATGNAA  
TGVTWTTTCK GTDASLFPAN FCGSVTQGGH HHHHH

LVL738 (DNA) - SEQ ID NO. 187:

atgaaatacctgctgccgaccgctgctggtctgctgctcctcgctgccagccggcgatggccaaggctgaacaaaatgatgt  
gaagctggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgatcgatctg  
ggtggataaccaagagccacaaaattgtacatttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacgta  
tgcacgttctgtcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatttggggaca  
gggttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataatgctgctcagattatt  
tgtgcgaactatggtgaagcatttgcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgtcagc  
gccttataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtattgcagcag  
atataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggatggcacattggca  
aatatggaatatatttgcagctacaggaatgtctgcaacaggtgaacttggaacaacaactgcaaaggaacgggatgcctcttatt  
tccagcaaattttgcggaagtgtcacacaaggcggccaccaccaccaccac

LVL738 (protein): (pelB sp)(ProtE aa 22-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 188:

MKYLLPTAAA GLLLLAAQPA MAKAEQNDVK LAPPTDVRSG YIRLVKNVNY  
YIDSESIWVD NOEPQIVHFD AVVNLDKGLY VYPEPKRYAR SVROYKILNC  
ANYHLTQVRT DFYDEFWGQG LRAAPKKQKK HTLSLTPDPTT LYNAAQIICA  
NYGEAFSVDK KGGTKKAASV ELLQASAPYK ADVELCVYST NETTNCTGGK  
NGIAADITTA KGYVKSVTTS NGAITVKGDG TLANMEYILQ ATGNAATGVT  
WTTTCKGTDA SLFPANFCGS VTQGGHHHHH H

LVL739 (DNA) - SEQ ID NO. 189:

atgaaatacctgctgccgaccgctgctggtctgctgctcctcgctgccagccggcgatggccgctgaacaaaatgatgtgaa  
gctggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgatcgatctgggtg  
gataaccaagagccacaaaattgtacatttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacgttatgca  
cgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaatttggggacagggt  
ttgcgggcagcacctaaaaagcaaaagaaacatacgtaagttaacacctgatacaacgctttataatgctgctcagattatttg  
cgaactatggtgaagcatttgcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgtcagcgcct  
tataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtattgcagcagatat  
aaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggatggcacattggcaaat

atggaatatattttgcaagctacaggtaatgctgcaacaggtgtaactggacaacaacttgcaaaggaacggatgcctctttatttcc  
agcaaattttgcggaagtgtcacacaaggcgccaccaccaccaccac

*LVL739 (protein): (pelB sp)(ProtE aa 23-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 190:*

MKYLLPTAAA GLLLLAQPA MAAEQNDVKL APPTDVRSY IRLVKNVNY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGOGL RAAPKKQKKH TLSLTPDTTL YNAAQIICAN  
YGEAFSVDKK GGTKKAAVSE LLQASAPYKA DVELCVYSTN ETTNCCTGGKN  
GIAADITTAK GYVKSVTTSN GAITVKGDGT LANMEYILQA TGNAATGVTW  
TTTCKGTDAS LPANFCGSV TQGGHHHHHH

*LVL740 (DNA) - SEQ ID NO. 191:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccgaacaaaatgatgtgaagct  
ggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgatcgatctgggtggat  
aaccaagagccacaaattgtacatttgatgcagtggtgaatttagataagggtatgtttatcctgagcctaaacggtatgcacgt  
tctgttcgctcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttgggacagggttg  
cgggcagcacctaaaaagcaaaagaaacatacgttaagtttaacacctgatacaacgctttataatgctgctcagattatttgctg  
aactatggtgaagcatttcagttgataaaaaaggcgccactaaaaagcagcggtatctgaattactgcaagcgctcagcgccctat  
aaggctgatgtgaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtattgcagcagatataac  
cacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggcacattggcaaatatg  
gaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttgacaacaacttgcaaaggaacggatgcctctttatttccag  
caaattttgcggaagtgtcacacaaggcgccaccaccaccaccac

*LVL740 (protein): (pelB sp)(ProtE aa 24-160)(GG)(PilA aa40-149)(GGHHHHHH) - SEQ ID NO. 192:*

MKYLLPTAAA GLLLLAQPA MAEQNDVKLA PPTDVRSYI RLVKNVNYI  
DSESIWVDNQ EPQIVHFDAV VNLDKGLYVY PEPKRYARV ROYKILNCAN  
YHLTQVRTDF YDEFWGOGLR AAPKKQKKHT LSLTPDTTLY NAAQIICANY  
GEAFSVDKKG GTKKAAVSEL LQASAPYKAD VELCVYSTNE TTNCTGGKNG  
IAADITTAKG YVKSVTTSNG AITVKGDGTL ANMEYILQAT GNAATGVTWT  
TTCKGTDASL FPANFCGSVT QGGHHHHHH



*LVL735 (DNA) - SEQ ID NO. 193:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccattcagaaggctgaacaaaa  
tgatgtgaagctggcaccgcccactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgatcg  
atctgggtggataaccaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaa  
cgttatgcacgttctgttcgctcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggg  
gacaggggttgccgggcagcacctaaaaagcaaaagaaacatacgttaagtttaacacctgatacaacgctttataatgctgctcag  
attatttgtgcaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgt  
cagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtattgca  
gcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggatggcacatt  
ggcaaatatggaatatatttgaagctacaggtaatgctgcaacagggtgaacttggacaacaacttgcaaaggaacgggatgcct  
ctttattccagcaaattttgcggaagtgtcacacaa

*LVL735 (protein): (pelB sp)(ProtE aa 20-160)(GG)(PilA aa40-149) - SEQ ID NO. 194:*

MKYLLPTAAA GLLLLAAQPA MAIQKAEOND VKLAPPTDVR SGYIRLVKNV  
NYIIDSESIW VDNQEPQIVH FDAVVNLDKG LYVYPEPKRY ARSVRQYKIL  
NCANYHLTOV RTDFYDEFWG QGLRAAPKKQ KKHTLSLTPD TTYLNAAQII  
CANYGEAFSV DKKGGTKKAA VSELLQASAP YKADVELCVY STNETTNCTG  
GKNGIAADIT TAKGYVKSVT TSNGAITVKG DGTLANMEYI LQATGNAATG  
VTWTTTCKGT DASLFPANFC GSVTQ

*LVL778 (DNA) - SEQ ID NO. 195:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccagcggccagattcagaaggc  
tgaacaaaatgatgtgaagctggcaccgcccactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcga  
tagtgaatcgatctgggtggataaccaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcct  
gagcctaaacggtatgcacgttctgttcgctcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgat  
gaattttggggacaggggttgccgggcagcacctaaaaagcaaaagaaacatacgttaagtttaacacctgatacaacgctttataat  
gctgctcagattatttgtgcaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattac  
tgcaagcgtcagcgccttataaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaat  
ggtattgcagcagatataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaagggg  
atggcacattggcaaatatggaatatatttgaagctacaggtaatgctgcaacagggtgaacttggacaacaacttgcaaagga  
acgggatgcctctttattccagcaaattttgcggaagtgtcacacaa



*LVL778 (protein): (pelB sp)(ProtE aa 17-160)(GG)(PilA aa40-149) - SEQ ID NO. 196:*

MKYLLPTAAA GLLLLAAQPA MASAOIQKAE ONDVKLAPPT DVRSGYIRLV  
KNVNYIIDSE SIWVDNQEPQ IVHFDAVVNL DKGLYVYPEP KRYARSVROY  
KILNCANYHL TQVRTDFYDE FWGOGLRAAP KKOKKHTLSL TPDITLYNAA  
OIICANYGEA FSVDKKGGTK KAAVSELLQA SAPIKADVEL CVYSTNETTN  
CTGGKNGIAA DITTAAGYVK SVTTSNGAIT VKGDGTLANM EYILQATGNA  
ATGVTWTTTC KGTDASLFPA NFCGSVTQ

*LVL779 (DNA) - SEQ ID NO. 197:*

atgaaatacctgctgccgaccgctgctggtctgctgctcctcgctgccagccggcgatggccgcccagattcagaaggctga  
acaaaatgatgtgaagctggcaccgcccactgatgtacgaagcgatatatcgtttggtaaagaatgtgaattattacatcgatag  
tgaatcgatctgggtggataaccaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctga  
gcctaaacggtatgcacgttctgttcgtcagtataagatctgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatga  
atgttggggacagggtttgcgggcagcacctaaaaagcaaaagaaacatacgttaagttaacacctgatacaacgctttataatgc  
tgctcagattatttgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggtatctgaattactg  
caagcgtcagcgccctataaggctgatgtggaattatgtgtatatagcacaatgaaacaacaaactgtacgggtggaaaaaatg  
gtattgcagcagatataaccacagcaaaaggctatgtaaaatcagtgaacaacaagcaacggtgcaataacagtaaaaggggat  
ggcacattggcaaatatggaatatatttgcagctacaggaatgctgcaacagggtgaacttgacaacaacttgcaaaaggaac  
ggatgcctctttattccagcaaattttgcggaagtgtcacacaa

*LVL779 (protein): (pelB sp)(ProtE aa 18-160)(GG)(PilA aa40-149) - SEQ ID NO. 198:*

MKYLLPTAAA GLLLLAAQPA MAAQIQKAEQ NDVKLAPPTD VRSGYIRLVK  
NVNYIIDSES IWVDNQEPQI VHFDAVVNLD KGLYVYPEPK RYARSVROYK  
ILNCANYHLT QVRTDFYDEF WGOGLRAAPK KOKKHTLSLT PDIITLYNAAQ  
IICANYGEAF SVDKKGGTKK AAVSELLQAS APYKADVELC VYSTNETTNC  
TGGKNGIAAD ITTAKGYVKS VTTTSNGAITV KGDGTLANME YILQATGNAA  
TGVTWTTTCK GTDASLFPA FCGSVTQ

*LVL780 (DNA) - SEQ ID NO. 199:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccaaggctgaacaaaatgatgt  
gaagctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgatcgatctg  
ggtggataaccaagagccacaaaattgtacatttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacgtta  
tgcacgttctgtcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggggaca  
gggtttgcgggcagcacctaaaaagcaaaagaaacatacgtaaagtttaacacctgatacaacgctttataatgctgctcagattatt  
tgtgcgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgtcagc  
gccttataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtattgcagcag  
atataaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggatggcacattggca  
aatatggaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggaacaacaactgcaaaggaacgggatgcctctttatt  
tccagcaaattttgcggaagtgtcacacaa

*LVL780 (protein): (pelB sp)(ProtE aa 22-160)(GG)(PilA aa40-149) - SEQ ID NO. 200:*

MKYLLPTAAA GLLLLAAQPA MAKAEONDVK LAPPTDVRSG YIRLVKVN  
NY YIDSESIWVD NOEPQIVHFD AVVNLDKGLY VYPEPKRYAR SVROYKILNC  
ANYHLTQVRT DFYDEFWGOG LRAAPKQKK UTLSLTEDTT LYNAAQILCA  
NYGEAFSVDK KGGTKKAASV ELLQASAPYK ADVELCVYST NETTNCTGGK  
NGIAADITTA KGYVKSVTTS NGAITVKGDG TLANMEYILQ ATGNAATGVT  
WTTTCKGTDA SLFPANFCGS VTQ

*LVL781 (DNA) - SEQ ID NO. 201:*

atgaaatacctgctgccgaccgctgctgctggtctgctgctcctcgctgccagccggcgatggccgctgaacaaaatgatgtgaa  
gctggcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgatcgatctgggtg  
gataaccaagagccacaaaattgtacatttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacgttatgca  
cgttctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggggacaggg  
ttgcgggcagcacctaaaaagcaaaagaaacatacgtaaagtttaacacctgatacaacgctttataatgctgctcagattatttg  
cgaactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcgggtatctgaattactgcaagcgtcagcgct  
tataaggctgatgtggaattatgtgtatatagcacaaatgaacaacaaactgtacgggtggaaaaaatggtattgcagcagatat  
aaccacagcaaaaggctatgtaaaatcagtgacaacaagcaacgggtgcaataacagtaaaaggggatggcacattggcaaat  
atggaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggaacaacaactgcaaaggaacgggatgcctctttatttc  
agcaaattttgcggaagtgtcacacaa

*LVL781 (protein): (pelB sp)(ProtE aa 23-160)(GG)(PilA aa40-149) - SEQ ID NO. 202:*

MKYLLPTAAA GLLLLAAQPA MAEONDVKL APPTDVRSGY IRLVKNVNY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGOGL RAAPKKQKKH TSLTPDCTL YNAAQIICAN  
YGEAFSVDKK GGTKAAVSE LLQASAPYKA DVELCVYSTN ETTNCTGGKN  
GIAADITTAK GYVKSVTTSN GAITVKGDGT LANMEYILQA TGNAATGVTW  
TTCKGTDAS LPANFCGSV TQ

*LVL782 (DNA) - SEQ ID NO. 203:*

atgaaatacctgctgccgaccgctgctggtctgctgctcctcgctgccagccggcgatggccgaacaaaatgatgtgaagct  
ggcaccgccgactgatgtacgaagcggatatatacgtttgtaaagaatgtgaattattacatcgatagtgaatcgatctgggtggat  
aaccaagagccacaaattgtacatttgatgcagtggtgaatttagataagggattgatgtttatcctgagcctaaacgttatgcacgt  
tctgttcgctcagtataagatcttgaattgtgcaaattatcatttaactcaagtacgaactgatttctatgatgaattttggggacagggttg  
cgggcagcacctaaaaagcaaaagaaacatacgttaagtttaacacctgatacaacgctttataatgctgctcagattattgtgcg  
aactatggtgaagcattttcagttgataaaaaaggcggcactaaaaaagcagcggatctgaattactgcaagcgtcagcgcttat  
aaggctgatgtggaattatgtgtatatagcacaaatgaaacaacaaactgtacgggtggaaaaaatggtattgcagcagatataac  
cacagcaaaaggctatgtaaaatcagtgacaacaagcaacggtgcaataacagtaaaaggggatggcacattggcaaatatg  
gaatatattttgcaagctacaggtaatgctgcaacaggtgtaacttggaacaacactgcaaaggaacggatgcctctttattccag  
caaatttttcggaagtgtcacacaa

*LVL782 (protein): (pelB sp)(ProtE aa 24-160)(GG)(PilA aa40-149) - SEQ ID NO. 204:*

MKYLLPTAAA GLLLLAAQPA MAEONDVKLA PPTDVRSGYI RLVKNVNYII  
DSESIWVDNQ EPQIVHFDAV VNLDKGLYVY PEPKRYARSV RQYKILNCAN  
YHLTQVRTDF YDEFWGOGLR AAPKKQKKHT LSLTPDCTL YNAAQIICANY  
GEAFSVDKKG GGTKAAVSEL LQASAPYKAD VELCVYSTNE TTNCTGGKNG  
IAADITTAKG YVKSVTTSNG AITVKGDGTL ANMEYILQAT GNAATGVTWT  
TTCKGTDASL FPANFCGSVT Q

The full length sequence for PE and PilA from which the above sequences were obtained are set forth in SEQ ID NO. 4 (PE) and SEQ ID NO. 58 (PilA), respectively.

### **Example 2: Vector Construction and Transformation**

Primers for amplifying PE from *H. influenzae* strain 772 were designed based on the sequence of *H. influenzae* strain Hi Rd. The 5' primer sequence contains one nucleotide difference compared to the NTHi 772 sequence, introducing an amino acid difference at position 24 when compared with the currently reported NTHi 772 genome sequence. Amino acid #24 in the fusion protein constructs is E (glutamic acid) instead of K (lysine) as found in NTHi 772.

#### *DNA Sequence for PE from H. influenzae strain Rd. - SEQ ID NO. 151*

```
atgaaaaaaattatTTAACATTATCacttgggttacttaccgcttgttctgctcaaataccaaaaggctgaacaaaatgatgtgaagctg
gcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgatcgatctgggtggata
accaagagccacaaaattgtacattttgatgctgtggtgaatttagataggggattgtatgtttatcctgagcctaaacgttatgcacgttc
tgttcgtcagtataagattttgaattgtgcaaattatcatttaactcaaatacgaactgatttctatgatgaattttggggacagggtttgcg
ggcagcacctaaaaagcaaaagaaacatacgttaagtttaacacctgatacaacgctttataatgctgctcagattatttgcgaaa
tatggtaaagcattttcagttgataaaaaataa
```

#### *Protein Sequence for PE from H. influenzae strain Rd. - SEQ ID NO. 152*

```
MKKIILTSL GLLTACSAQI QKAEQNDVKL APPTDVRSYG IRLVKNVNY
IDSESIWVDN QEPQIVHFDA VVNLDRGLYV YPEPKRYARS VRQYKILNCA
NYHLTQIRTD FYDEFWGQGL RAAPKKQKKH TSLTPDTTL YNAAQIICAN
YGKAFSVDKK
```

#### *DNA Sequence for PE from H. influenzae strain 772 (as set forth in: Microbes & Infection, Corrigendum to "Identification of a novel Haemophilus influenzae protein important for adhesion to epithelia cells" [Microbes Infect. 10 (2008) 87-97], available online July 6, 2010, "Article in Press")) - SEQ ID NO. 153*

```
atgaaaaaaattatTTAACATTATCacttgggttacttactgcctgttctgctcaaataccaaaaggctaaacaaaatgatgtgaagctg
gcaccgccgactgatgtacgaagcggatatatacgtttggtaaagaatgtgaattattacatcgatagtgatcgatctgggtggata
accaagagccacaaaattgtacattttgatgcagtggtgaatttagataagggattgtatgtttatcctgagcctaaacgttatgcacgtt
ctgttcgtcagtataagatcttgaattgtgcaaattatcatttaactcaagtagcgaactgatttctatgatgaattttggggacagggtttgc
```

gggcagcacctaaaaagcaaaagaaacatacgtaagtttaacacctgatacaacgctttataatgctgctcagattattgtgcga  
actatggtgaagcattttcagttgataaaaaa

*Protein Sequence for PE from H. influenzae strain 772 (as set forth in: Microbes & Infection, Corrigendum to "Identification of a novel Haemophilus influenzae protein important for adhesion to epithelia cells" [Microbes Infect. 10 (2008) 87-97], available online July 6, 2010, "Article in Press")) - SEQ ID NO. 154*

MKKIILTLSL GLLTACSAQI QKAKQNDVKL APPTDVRSYG IRLVKNVNYY  
IDSESIWVDN QEPQIVHFDA VVNLDKGLYV YPEPKRYARS VRQYKILNCA  
NYHLTQVRTD FYDEFWGQGL RAAPKKQKKH TSLSTPDTTL YNAAQIICAN  
YGEAFSVDKK

#### Vector construction:

To generate LVL312, LVL291, LVL268, LVL269, LVL270, LVL702, LVL735, LVL778, LVL779, LVL780, LVL781 and LVL782, a polymerase chain reaction (PCR) preparation of the following components was prepared (specific components are subsequently exemplified): 36.6 µl of deionized water, 5 µl of buffer #1 10X, 5 µl of dNTPs 2mM, 2 µl MgCl<sub>2</sub> 25 mM, 0.4 µl of primer #1 (50 µM), 0.4 µl of primer #2 (50 µM), 0.5 µl of template (100 ng/µl) and 0.4 µl of KOD HiFi DNA polymerase 2.5 units/µl (NOVAGEN®) was formulated. Polymerase chain reaction involved 25 cycles of 15 seconds of denaturation at 98°C, 2 seconds for annealing at 55°C and 20 seconds of primer extension at 72°C. The PCR products were purified using QIAQUICK® PCR purification kit (QIAGEN®). This product was used under conditions recommended by the supplier which were: the addition of 5 volumes Buffer PB, provided in the QIAQUICK® PCR purification kit, to 1 volume of the PCR preparation. The PCR preparation with Buffer PB was subsequently mixed by vortex. A QIAQUICK® column was placed into a 2 ml collection tube. To bind DNA in the PCR preparation to the column, the mixed sample was applied to the QIAQUICK® column and centrifuged for 30–60 seconds at 14 000 RPM. The flow-through was discarded and the QIAQUICK® column was placed back in the same tube. To wash the bound DNA 0.75 ml Buffer PE, provided in the QIAQUICK® PCR purification kit, was added to the QIAQUICK® column, and the column was centrifuged for 30–60 seconds at 14 000 RPM. The flow-through was discarded and the QIAQUICK® column was placed back in the same tube. The QIAQUICK® column was centrifuged once more in the 2 ml collection tube for 1 minute to remove residual wash buffer. Each QIAQUICK® column was placed in a clean 1.5 ml microcentrifuge tube. To elute the DNA, 33 µl water was added to the center of the QIAQUICK®

membrane and the column was centrifuged for 1 minute at 14 000 RPM. Restriction enzymes and buffer related were obtained from New England BioLabs. For example, approximately 5 µl of pET26b vector (100 ng/µl), 2 µl of NEBuffer 2 (New England Biolabs, 1X NEBuffer 2: 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol, pH 7.9 at 25°C), 1 µl of *NdeI* (20 000 units/ml), 1 µl of *HindIII* (20 000 units/ml) and 11 µl of deionized water were mixed and incubated for two hours at 37°C for DNA digestion. Thereafter, a second step of purification was performed using the QIAQUICK<sup>®</sup> PCR purification kit (QIAGEN<sup>®</sup>) with the procedure described above.

Ligation was performed using Quick T4 DNA ligase and Quick Ligation Reaction Buffer from New England BioLabs. For example, around 10 ng of vector and 30 ng of insert in 10 µl of deionized water were mixed with 10 µl of 2X Quick Ligation Reaction Buffer (New England Biolabs, 132 mM Tris-HCl, 20 mM MgCl<sub>2</sub>, 2mM dithiothreitol, 2 mM ATP, 15% polyethylene glycol, pH 7.6 at 25°C) and 1 µl of Quick T4 DNA ligase (New England Biolabs). The enzymatic reaction was incubated for 5 minutes at room temperature before transformation.

To generate LVL315, LVL317, LVL318, LVL736, LVL737, LVL738, LVL739 and LVL740, a PCR preparation of the following components was prepared: 40 µl of deionized water, 5 µl of reaction buffer 10X, 1 µl of dNTPs mix, 1 µl of primer #1 (10 µM), 1 µl of primer #2 (10 µM), 1 µl of template (25 ng/µl) and 1 µl of *PfuUltra* High-Fidelity DNA polymerase 2.5 units/µl (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies, Stratagene Division) was formulated. Polymerase chain reaction involved one cycle of denaturation at 95°C for 30 sec, 18 cycles of 30 sec of denaturation at 95°C, 1 min for annealing at 55°C and 5 min 30 sec of primer extension at 68°C. The PCR products were digested using 1 µl of DpnI restriction enzyme at 37°C for one hour before transformation.

A detailed list of PCR primer sequences used for amplifications is illustrated in Table 4.

To generate pRIT16711, the PE gene fragment coding for amino acids 22 to 160 of SEQ ID NO. 4, which excludes the sequence coding for its corresponding secretion signal, was amplified by PCR from genomic DNA of NTHi strain 772. The amplification primers were designed based on the available strain Hi Rd sequence (at that time, the 772 sequence was not known). The 5' primer sequence contains one mutation compared to the NTHi 772 sequence (sequence as now available), introducing one amino acid difference in PE coding sequence at position 24, glutamic

acid (E) instead of lysine (K). After PCR amplification, the insert was cloned in the pET-26(+) expression vector (NOVAGEN®) using *Bam*HI and *Xho*I restriction sites.

To generate pRIT16671, a DNA fragment coding for a *PilA* gene fragment (amino acids 40 to 149 of SEQ ID NO. 58, SEQ ID NO. 127), which excludes its leader peptide as well as a portion of the predicted hydrophobic alpha helix, was amplified from genomic DNA of NTHi strain 86-028NP and cloned into the pET15 expression vector. The vector pRIT16790 (containing amino acids 40 to 149 from NTHi strain 86-028NP) was used as a template to generate the vector pRIT16671. The *PilA* gene fragment was amplified by PCR using the vector pRIT16790 and primers MDES PILA-3 and MDES PILA-4. The *PilA* fragment was cloned into the pET-26 expression vector using *Nde*I / *Xho*I restriction sites. The DNA sequence encoding six histidine (his) amino acids was incorporated into the 5' primer to add six histidines (6xhis) at the N-terminal end of the *PilA* sequence (MDES PILA-3).

To generate LVL312 (*FlgI* signal peptide-E-*PilA* fragment-GG-PE fragment-GGHHHHHH), a polymerase chain reaction was performed to amplify the *PilA* gene (amino acids 40-149 / strain 86-028NP) using the pRIT16671 vector as a template and primers CAN534 and CAN537. DNA sequence corresponding to *FlgI* signal peptide (sp) and glutamic acid (E) amino acid was incorporated into the 5' primer (CAN534). To link the *PilA* sequence to *PE* sequence, DNA sequence corresponding to the two glycine (GG) amino acids linker and the N-terminal *PE* amino acids were incorporated into the 3' primer (CAN537). Another polymerase chain reaction was performed to amplify the *PE* gene (amino acids 18-160) using pRIT16711 vector as a template and primers CAN536 and CAN538. DNA sequence corresponding to the C-terminal *PilA* amino acids and GG amino acids were incorporated into the 5' primer to link *pilA* to *PE* sequence (CAN536). DNA sequence corresponding to the GG amino acids linker and 6xhis amino acids were incorporated into the 3' primer (CAN538). Finally, to generate LVL312, a third polymerase chain reaction was performed to amplify the *PilA* and *PE* genes in fusion with the *FlgI* signal peptide at the N-terminus, a glutamic acid (E) amino acid between *FlgI* and *pilA*, a GG linker between *PilA* and *PE* sequences and a GG linker between *PE* and the 6xhis amino acids at the C-terminus. To achieve this amplification, the products of the two polymerase chain reactions described above were used as a template with primers CAN534 and CAN538. DNA sequence corresponding to *Nde*I restriction site was incorporated into the 5' primer and *Hind*III restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL291 (pelB signal peptide-PE fragment-GG-PilA fragment-GG-6xhis), a polymerase chain reaction was performed to amplify the *PE* gene (amino acids 19-160) using the pRIT16711 vector as a template and primers CAN544 and CAN546. DNA sequence corresponding to pelB signal peptide (sp) amino acids was incorporated into the 5' primer (CAN544). To link the *PilA* sequence to the *PE* sequence, DNA sequence corresponding to GG amino acids linker and the N-terminal PilA amino acids were incorporated into the 3' primer (CAN546). Another polymerase chain reaction was performed to amplify the *PilA* gene (amino acids 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) using the pRIT16671 vector as a template with primers CAN545 and CAN535. DNA sequence corresponding to the C-terminal PE amino acids and GG amino acids were incorporated into the 5' primer (CAN545) to link the PilA sequence to the PE sequence. DNA sequence corresponding to linker GG amino acids and 6xhis amino acids were incorporated into the 3' primer (CAN535). Finally, to generate LVL291, a third polymerase chain reaction was performed to amplify the PE and PilA genes in fusion with the pelB signal peptide at the N-terminus, a GG linker between the PE and PilA sequences and a GG linker between PilA and 6xhis amino acids at the C-terminus. To achieve this amplification, the products of two polymerase chain reactions described above were used as a template with primers CAN544 and CAN535. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *HindIII* restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL268 (pelB signal peptide-D-PE fragment-GG-PilA fragment-GG-6xhis), a polymerase chain reaction was performed to amplify the *PE* gene (amino acids 20-160) using the pRIT16711 vector as a template with primers CAN547 and CAN546. DNA sequence corresponding to the pelB signal peptide (sp) amino acids and aspartic acid (D) amino acid were incorporated into the 5' primer (CAN547). To link the *PilA* sequence to the *PE* sequence, DNA sequence corresponding to GG amino acids linker and the N-terminal PilA amino acids were incorporated into the 3' primer (CAN546). Another polymerase chain reaction was performed to amplify the *PilA* gene (amino acids 40-149 / NTHi strain 86-028NP) using the pRIT16671 vector as a template with CAN545 and CAN535. DNA sequence corresponding to the C-terminal PE amino acids and GG amino acids were incorporated into the 5' primer (CAN545) to link the PilA sequence to the PE sequence. DNA sequence corresponding to linker GG amino acids and 6xhis amino acids were incorporated into the 3' primer (CAN535). Finally, to generate LVL268,



a third polymerase chain reaction was performed to amplify the PE and PilA genes in fusion with the pelB signal peptide at the N-terminus, a D amino acid between pelB signal peptide and PE, a GG linker between PE and pilA sequences and a GG linker between PilA and 6xhis amino acids in C-term. To achieve this amplification, the products of the two polymerase chain reactions described above were used as a template with primers CAN547 and CAN535. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *HindIII* restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL269 (NadA signal peptide-ATNDDD-PE fragment-GG-PilA fragment-GG-6xhis), a polymerase chain reaction was performed to amplify the *PE* gene (amino acids 22-160 of SEQ ID NO. 4) using the pRIT16711 vector as a template with primers CAN548 and CAN546. DNA sequence corresponding to pelB signal peptide (sp) amino acids and ATNDDD amino acids were incorporated into the 5' primer (CAN548). To link the *PilA* sequence to the *PE* sequence, DNA sequence corresponding to the GG amino acids linker and the N-terminal PilA amino acids were incorporated into the 3' primer (CAN546). Another polymerase chain reaction was performed to amplify the *PilA* gene (amino acids 40-149 of SEQ ID NO. 58, SEQ ID NO. 127) using the pRIT16671 vector as a template with primers CAN545 and CAN535. DNA sequence corresponding to the C-terminal PE amino acids and GG amino acids were incorporated into the 5' primer to link the PilA sequence to the PE sequence (CAN545). DNA sequence corresponding to linker GG amino acids and 6xhis amino acids were incorporated into the 3' primer (CAN535). Finally, to generate LVL269, a third polymerase chain reaction was performed to amplify the PE and PilA gene in fusion with the NadA signal peptide at the N-terminus, ATNDDD amino acids between the pelB signal peptide and PE, a GG linker between the PE and pilA sequences and a GG linker between PilA and 6xhis amino acids at the C-terminus. To achieve this amplification, the products of the two polymerase chain reactions describe above were used as a template with primers CAN548 and CAN535. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *HindIII* restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL270 (M-6xHis-PE fragment-GG-PilA fragment), a polymerase chain reaction was performed to amplify the *PE* gene (amino acids 17-160) using the pRIT16711 vector as a template with primers CAN540 and CAN542. DNA sequence corresponding to 6xhis amino

acids were incorporated into the 5' primer (CAN540). To link the *PilA* sequence to the *PE* sequence, DNA sequence corresponding to the GG amino acids linker and the N-terminal *PilA* amino acids were incorporated into the 3' primer (CAN542). Another polymerase chain reaction was performed to amplify the *PilA* gene (amino acids 40-149 / NTHi strain 86-028NP) using pRIT16671 vector as a template with primers CAN541 and CAN543. DNA sequence corresponding to the C-terminal *PE* amino acids and GG amino acids were incorporated into the 5' primer (CAN541) to link the *PilA* to the *PE* sequence. Finally, to generate LVL270, a third polymerase chain reaction was performed to amplify the 6-his-*PE*-GG-*PilA* gene in fusion. To achieve this amplification, the products of the two polymerase chain reactions describe above were used as a template with primers CAN540 and CAN543. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *HindIII* restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL315 (pelB signal peptide-MD-*PE* fragment-GG-*PilA* fragment-GG-6xhis), a site-directed mutagenesis was performed to change the N-terminal *PE* amino acid sequence from QIQ to MD using LVL291 as a template with primers CAN670 and CAN671 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL317 (pelB signal peptide-*PE* fragment-GG-*pilA* fragment), a site-directed mutagenesis was performed to incorporate a stop codon between the *PilA* gene and the DNA sequence corresponding to GGHHHHHH amino acid residues (SEQ ID NO: 3) using LVL291 as a template with primers CAN678 and CAN679 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL318 (pelB signal peptide-MD-*PE*-GG-*PilA*), a site-directed mutagenesis was performed to incorporate a stop codon between the *PilA* gene and the DNA sequence corresponding to GGHHHHHH amino acid residues (SEQ ID NO: 3) using LVL315 as a template with primers CAN678 and CAN679 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL702 (LVL291  $\Delta$ Q), a polymerase chain reaction was performed using the LVL291 vector as template and primers CAN1517 and CAN1518. Deletion of three nucleotides corresponding to the amino acid Q at the position 23 on LVL291 sequence was incorporated to

the 5' primer. The only difference between LVL702 and LVL291 is the deletion of amino acid Q at the position 23 on LVL291 sequence. *NdeI* and *HindIII* restriction sites were incorporated into the 5' and 3' primers respectively. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL735 (LVL317  $\Delta$ Q), a polymerase chain reaction was performed using the LVL317 vector as template and primers CAN1517 and CAN1519. Deletion of three nucleotides corresponding to the amino acid Q at the position 23 on LVL317 sequence was incorporated to the 5' primer. The only difference between LVL735 and LVL317 is the deletion of amino acid Q at the position 23 on LVL317 sequence. *NdeI* and *HindIII* restriction sites were incorporated into the 5' and 3' primers respectively. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

To generate LVL736 (LVL291 + SA), a site-directed mutagenesis was performed to add amino acids S and A between amino acid 22 and 23 on LVL291 sequence. LVL291 was used as template with primers CAN1531 and CAN1532 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL737 (LVL291 + A), a site-directed mutagenesis was performed to add amino acid A between amino acid 22 and 23 on LVL291 sequence. LVL291 was used as template with primers CAN1529 and CAN1530 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL738 (LVL291  $\Delta$ QIQ), a site-directed mutagenesis was performed to delete amino acids Q, I and Q at positions 23 to 25 on LVL291 sequence. LVL291 was used as template with primers CAN1523 and CAN1524 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL739 (LVL291  $\Delta$ QIQK), a site-directed mutagenesis was performed to delete amino acids Q, I, Q and K at positions 23 to 26 on LVL291 sequence. LVL291 was used as template with primers CAN1525 and CAN1526 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL740 (LVL291  $\Delta$ QIQKA), a site-directed mutagenesis was performed to delete amino acids Q, I, Q, K and A at positions 23 to 27 on LVL291 sequence. LVL291 was used as template with primers CAN1527 and CAN1528 and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

To generate LVL778 (LVL736  $\Delta$ 6xHis tag), LVL779 (LVL737  $\Delta$ 6xHis tag), LVL780 (LVL738  $\Delta$ 6xHis tag), LVL781 (LVL739  $\Delta$ 6xHis tag) and LVL782 (LVL740  $\Delta$ 6xHis tag) a polymerase chain reaction was performed using the LVL736, LVL737, LVL738, LVL739 and LVL740 vectors as template, respectively, with primers CAN1669 and CAN543. Deletion of 6xHis tag corresponds to the amino acid sequence GGHHHHHH (SEQ ID NO. 3) at the C-terminal sequences. This deletion was incorporated to the 3' primer. *NdeI* and *HindIII* restriction sites were incorporated into the 5' and 3' primers respectively. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN®).

Table 4: PCR primer sequences used for PE, PilA and PE-PilA amplifications

| Primer ID | DNA Sequence<br>5' – 3'                                                                                         |
|-----------|-----------------------------------------------------------------------------------------------------------------|
| CAN534    | CACACACATATGATTAAATTTCTCTCTGCATTAATTCTTCTACTGGTCACGACGG<br>CGGCTCAGGCTGAGACTAAAAAGCAGCGGTATCTG (SEQ ID NO. 155) |
| CAN535    | TGTGTGAAGCTTTTAGTGTTGGTGGTGGTGGTGGCCGCCTTGTGTGACACTTCCG<br>CAAAAATTGTC (SEQ ID NO. 156)                         |
| CAN536    | TTTGCGGAAGTGTCACACAAGGCGGCGCGCAGATTCAGAAGGCTGAACAAAATGA<br>TGT (SEQ ID NO. 157)                                 |
| CAN537    | ACATCATTTTGTTCAGCCTTCTGAATCTGCGCGCCGCCTTGTGTGACACTTCCGC<br>AAA (SEQ ID NO. 158)                                 |
| CAN538    | TGTGTGAAGCTTTTAGTGTTGGTGGTGGTGGTGGTGGCCGCCTTTTTTATCAACTGAA<br>AATG (SEQ ID NO. 159)                             |
| CAN540    | CACACACATATGCACCACCACCACCACCACAGCGCGCAGATTCAGAAGGCTGAAC<br>AAAATGATGT (SEQ ID NO. 160)                          |
| CAN541    | CATTTTCAGTTGATAAAAAAGGCGGCACTAAAAAAGCAGCGGTATC (SEQ ID<br>NO. 161)                                              |

|                |                                                                                                                                              |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>CAN542</b>  | GATACCGCTGCTTTTTTAGTGCCGCCTTTTTTATCAACTGAAAATG (SEQ ID NO. 162)                                                                              |
| <b>CAN543</b>  | TGTGTGAAGCTTTTATTGTGTGACACTTCCGCAA (SEQ ID NO. 163)                                                                                          |
| <b>CAN544</b>  | CACACACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCG<br>CTGCCCAGCCGGCGATGGCCCAGATTTCAGAAGGCTGAACAAAATGATGT (SEQ<br>ID NO. 164)            |
| <b>CAN545</b>  | GCATTTTCAGTTGATAAAAAAGGCGGCACTAAAAAAGCAGCGGTATCTG (SEQ<br>ID NO. 165)                                                                        |
| <b>CAN546</b>  | CAGATACCGCTGCTTTTTTAGTGCCGCCTTTTTTATCAACTGAAAATGC (SEQ<br>ID NO. 166)                                                                        |
| <b>CAN547</b>  | CACACACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCG<br>CTGCCCAGCCGGCGATGGCCGATATTTCAGAAGGCTGAACAAAATGATGT (SEQ<br>ID NO. 167)            |
| <b>CAN548</b>  | CACACACATATGAAACACTTTCCATCCAAAGTACTGACCACAGCCATCCTTGCCA<br>CTTTCTGTAGCGGCGCACTGGCAGCCACAAACGACGACGATAAGGCTGAACAAAA<br>TGATG (SEQ ID NO. 168) |
| <b>CAN670</b>  | GCCGGCGATGGCCATGGATAAGGCTGAACAAAATG (SEQ ID NO. 169)                                                                                         |
| <b>CAN671</b>  | CATTTTGTTTCAGCCTTATCCATGGCCATCGCCGGC (SEQ ID NO. 170)                                                                                        |
| <b>CAN678</b>  | GGAAGTGTACACAATAAGGCGGCCACCACCACC (SEQ ID NO. 171)                                                                                           |
| <b>CAN679</b>  | GGTGGTGGTGGCCGCCTTATTGTGTGACACTTCC (SEQ ID NO. 172)                                                                                          |
| <b>CAN1517</b> | GATATACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCG<br>CTGCCCAGCCGGCGATGGCCATTTCAGAAGGCTGAACAAAA (SEQ ID NO. 205)                        |
| <b>CAN1518</b> | GGCCGCAAGCTTTTAGTGGTGGTGGTGGTGGTGGCCGCC (SEQ ID NO. 206)                                                                                     |
| <b>CAN1519</b> | GGCCGCAAGCTTTTATTGTGTGACACTTCC (SEQ ID NO. 207)                                                                                              |
| <b>CAN1523</b> | GCTGCCCAGCCGGCGATGGCCAAGGCTGAACAAAATGATGTG (SEQ ID NO.<br>208)                                                                               |
| <b>CAN1524</b> | CACATCATTTTGTTTCAGCCTTGGCCATCGCCGGCTGGGCAGC (SEQ ID NO.<br>209)                                                                              |
| <b>CAN1525</b> | GCTGCCCAGCCGGCGATGGCCGCTGAACAAAATGATGTGAAGC (SEQ ID NO.<br>210)                                                                              |

|                   |                                                                         |
|-------------------|-------------------------------------------------------------------------|
| <b>CAN1526</b>    | GCTTCACATCATTTTGTTCAGCGGCCATCGCCGGCTGGGCAGC (SEQ ID NO. 211)            |
| <b>CAN1527</b>    | GCTGCCCAGCCGGCGATGGCCGAACAAAATGATGTGAAGCTGG (SEQ ID NO. 212)            |
| <b>CAN1528</b>    | CCAGCTTCACATCATTTTGTTCGGCCATCGCCGGCTGGGCAGC (SEQ ID NO. 213)            |
| <b>CAN1529</b>    | GCTGCCCAGCCGGCGATGGCCGCCAGATTCAGAAGGCTGAAC (SEQ ID NO. 214)             |
| <b>CAN1530</b>    | GTTTCAGCCTTCTGAATCTGGGCGGCCATCGCCGGCTGGGCAGC (SEQ ID NO. 215)           |
| <b>CAN1531</b>    | GCTGCCCAGCCGGCGATGGCCAGCGCCCAGATTCAGAAGGCTGAAC (SEQ ID NO. 216)         |
| <b>CAN1532</b>    | GTTTCAGCCTTCTGAATCTGGGCGCTGGCCATCGCCGGCTGGGCAGC (SEQ ID NO. 217)        |
| <b>CAN1669</b>    | CACACACATATGAAATACCTGCTGCCGACC (SEQ ID NO. 218)                         |
| <b>MDesPILA-3</b> | GAATTCCATATGCACCATCACCATCACCATACTAAAAAAGCAGCGGTATCTGAA (SEQ ID NO. 173) |
| <b>MDesPILA-4</b> | GCGCCGCTCGAGTCATTGTGTGACACTTCCGC (SEQ ID NO. 174)                       |
| <b>MnoNTHi-44</b> | GCCCAGCCGGCGATGGCCAGATCCAGAAGGCTGAACAAAATG (SEQ ID NO. 175)             |
| <b>MnoNTHi-45</b> | CATTTTGTTCAGCCTTCTGGATCTGGGCCATCGCCGGCTGGGC (SEQ ID NO. 176)            |

### Transformation

*Escherichia coli* BLR (DE3) or *E. coli* HMS (DE3) cells were transformed with plasmid DNA according to standard methods with CaCl<sub>2</sub>-treated cells. (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.). Briefly, BLR (DE3) or HMS174(DE3) competent cells were gently thawed on ice. Approximately 4µl of plasmid (10-100 ng) were mixed using 50-100 µl competent cells. Thereafter, this

formulation was incubated on ice for 30 min. To perform the transformation reaction, the formulation was heat pulsed at 42°C for 45 seconds then incubated on ice for 2 minutes. Approximately 0.5 ml of SOC medium (Super Optimal broth with Catabolite repression) was added to the transformed cells and the cell culture was incubated at 37°C for one hour before plating on Luria-Bertani (LB) agar with 50 ug/ml kanamycin. Around 100 µl of transformed cell culture was plated and incubated overnight at 37°C.

BLR (DE3): BLR is a *recA*<sup>-</sup> derivative of BL21 (F<sup>-</sup> *ompT hsdSB*(rB<sup>-</sup> mB<sup>-</sup>) *gal dcm* (DE3). This *E. coli* strain used for expression of recombinant proteins improves plasmid monomer yields and may help stabilize target plasmids containing repetitive sequences or whose products may cause the loss of the DE3 prophage. (Studier, F.W. (1991) J. Mol. Biol. 219: 37–44). The detailed genotype of *E.coli* BLR (DE3) has been published by NOVAGEN®. (F<sup>-</sup> *ompT hsdSB* (rB<sup>-</sup> mB<sup>-</sup>) *gal dcm* Δ(*srl-recA*)306::Tn10 (TetR) (DE3).

HMS174 (DE3): HMS174 strains provide the *recA* mutation in a K-12 background. Like BLR, these strains may stabilize certain target genes whose products may cause the loss of the DE3 prophage. The detailed genotype of *E.coli* HMS174 (DE3) has been published by NOVAGEN®. (F<sup>-</sup> *recA1 hsdR*(rK12<sup>-</sup> mK12<sup>+</sup>) (DE3) (Rif R ).

**Production using BLR (DE3) and Characterization of His tagged constructs are described in Example 3 through Example 6**

**Example 3: Protein expression using shake flask**

Generally, one confluent agar plate inoculated with *Escherichia coli* BLR (DE3) transformed with recombinant plasmid was stripped, resuspended in culture media and used to inoculate 800 ml of LB broth (Becton, Dickinson and Company) ± 1% (weight/volume, w/v) glucose (Laboratoire MAT, catalogue number: GR-0101) and 50µg/ml kanamycin (Sigma) to obtain O.D.<sub>600nm</sub> between 0.1 and 0.2. Cultures were incubated at 37 °C with agitation of 250 RPM to reach an O.D.<sub>600nm</sub> of ~0.8.

One ml of each culture was then collected, centrifuged at 14 000 RPM for 5 minutes and supernatants and pellets were frozen at -20°C separately.

At an O.D.<sub>600nm</sub> ~0.8, the BLR (DE3) cultures were cooled down (-20°C, 20 minutes or 4°C, 1 hour, preferably at 4°C for 1 hour) before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubation overnight at 16, 22 and 30°C, or 3 hours at 37°C with agitation of 250 RPM, preferably overnight at 22°C. After the induction period the cultures were centrifuged at 14 000 RPM for 5 minutes or 6 000 RPM for 15 minutes and supernatant (media fraction sample) and pellets (containing soluble and insoluble fractions) were frozen at -20°C separately.

These conditions are used for periplasmic protein expression.

#### **Example 4: Protein purification using shake flask, cell pastes, His tagged constructs**

Each bacterial pellet obtained after induction was resuspended in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 8.0) containing 500 mM NaCl, 10 mM imidazole and Roche COMPLETE® Protease Inhibitor Cocktail (1 tablet/50 ml of HEPES buffer containing 500 mM NaCl, Roche COMPLETE® ULTRA tablets, Roche Diagnostics Corporation).

Alternatively, 20 to 50 mM bicine buffer may be used instead of HEPES buffer containing NaCl. For example, 20 mM bicine buffer may be used. Bacteria were lysed using a Constant System 1.1 KW 2 X 30 000 PSI (pounds per square inch). Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 20 min at 4°C.

6-His tagged-proteins were purified under native conditions on immobilized metal affinity chromatography (IMAC) using PROFINIA™ protein purification protocol (Bio-Rad Laboratories, Inc.). The soluble components were loaded on a 5ml His Trap column (Bio-Rad Laboratories, Inc.) preequilibrated with the same buffer used for bacterial resuspension; the soluble components were added at up to 5 ml/min (producing a “flow through fraction”) After loading on the column, the column was washed with 10 column volumes of the same buffer at a rate of 10 ml/min (producing a “wash fraction #1). A second wash using 20 mM bicine buffer or 20 mM HEPES buffer (pH 8.0) containing 500 mM NaCl and 20 mM imidazole was performed, producing a “wash fraction #2). Elution was performed using 2 column volumes of 20mM HEPES buffer or 50mM bicine buffer (pH 8.0) containing 500 mM NaCl and 250 mM imidazole at a rate of 10 ml/min, producing an “elution fraction”.



To improve the purity of the protein, positive elution fractions from IMAC were pooled and loaded on a size exclusion chromatography (SEC) column (HILOAD™ SUPERDEX™ 200 26/60 from GE Healthcare) preequilibrated in phosphate buffered saline without calcium or magnesium (NaCl 137 mM, KCl 2.7 mM, Na<sub>2</sub>HPO<sub>4</sub> 8.1 mM, KH<sub>2</sub>PO<sub>4</sub> 1.47 mM, pH 7.4). Samples from elution fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were concentrated using Centricon 10 000 MW (Millipore).

Protein concentration was determined using spectrometer

**Example 5: SDS-PAGE and Western Blot Analysis of His tagged constructs & SDS-PAGE Analysis of non-his tagged LVL317 & LVL318 constructs**

**Soluble and insoluble fraction preparation**

For example, 1 ml of culture after induction (see, for example, Example 3 above) was centrifuged at 14 000 RPM for 2 min. The pellet was resolubilized using 40 µl of BUGBUSTER® Protein Extraction Reagent (NOVAGEN®, EMD4 Biosciences, Merck), creating a cell suspension. The cell suspension was incubated on a rotating platform for 10 min at room temperature. The cell suspension was then centrifuged at 14 000 RPM for 2 min to separate the soluble fraction. The resulting pellet (insoluble fraction) was resolubilized using 70 µl of deionized water, 5 µl of dithiothreitol (DTT) 1M and 25 µl of NUPAGE® LDS (Lithium Dodecyl Sulphate) Sample Buffer 4X (INVITROGEN™). The soluble fraction (supernatant from the cell suspension of the resolubilized pellet) was added to 30 µl of deionized water, 5 µl of DTT 1M and 25 µl of LDS Sample Buffer 4X.

**Media fraction preparation**

For example, to prepare the media fraction, 100 µl of the supernatant from the induced whole cell culture following centrifugation (see, for example, Example 3 above) was concentrated by adding 500 µl of RC reagent I (Bio-Rad Laboratories, Inc.); the sample was mixed and incubated for 1 min at room temperature. Then, 500 µl of Reagent II (Bio-Rad Laboratories, Inc.) was added to the sample and mixed. This formulation was centrifuged at 14 000 RPM for 10 min. The pellet was resolubilized using 28 µl of deionized water, 2 µl of DTT 1M and 10 µl of LDS SB 4X.

#### Purification fraction preparation

For example, purified proteins (for example, obtained as described in Example 4) were prepared for SDS-PAGE analysis by adding 70 µl of sample, 5 µl of DTT 1M and 25 µl of LDS Sample Buffer 4X.

#### SDS-PAGE analysis and transfer to nitrocellulose membrane

SDS-PAGE analysis and transfer to nitrocellulose membrane were performed according to manufacturer's recommendations (Invitrogen) using NUPAGE® Bis-Tris 4-12% gels. Preparations of samples, buffers and migration conditions were done under conditions recommended by the suppliers.

In one example, the gel was loaded with a 20 µl sample from a master mix comprising 70 µl of a purified protein fraction, 5 µl of DTT 1M and 25 µl of LDS SB 4X.

After samples were run on NUPAGE® Bis-Tris 4-12% gels, the proteins were transferred to nitrocellulose membranes.

Nitrocellulose membranes were blocked for 30 minutes at 37°C, 60 RPM using 3 % milk / PBS 1X fresh solution. After the blocking incubation, Primary Antibodies were added (6X His Tag® antibody, Abcam PLC, catalogue number: ab9108) at a dilution of: 1:1000 in 3 % milk / PBS 1X fresh solution for 1 hour at 37°C, 60 RPM. After that, membranes were washed three times, for 5 minutes each, at room temperature using 0.02% polysorbate 20 (for example, TWEEN™ 20) / PBS 1X. Secondary Antibodies (alkaline phosphatase (AP) Rabbit anti-IgG (H+L) rabbit, Jackson ImmunoResearch Laboratories, Inc.) were added at dilution 1:14 000 using 3 % milk / PBS 1X fresh solution. Membranes were incubated for 1 hour at 37°C, 60 RPM. After that, membranes were washed three times for 5 minutes at room temperature using 0.02% polysorbate 20 (for example, TWEEN™ 20) / PBS 1X before the membrane expositions to 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (for example, BCIP®/NBT from Sigma-Aldrich®, 1 tablet / 10 ml water).

See Figure 1 for SDS-PAGE of induced bacterial extracts for fusion protein constructs LVL291, LVL268 and LVL269. Insoluble fraction (I), Soluble fraction (S) and Culture Media fraction (M) were loaded for LVL291, LVL268 and LVL269 before and after induction (ind).

See Figure 2 for SDS-PAGE and Western blot related to purification extracts for fusion protein constructs LVL291, LVL268 and LVL269. Flow through fraction (Ft), Wash fraction (W) and Elution fraction (E) were loaded for purification of LVL291, LVL268 and LVL269. Anti-his tag was used to probe extracts.

See Figure 3 for SDS-PAGE of induced bacterial and purification extracts for fusion protein constructs LVL291 and LVL315. Culture Media fraction (M), Soluble fraction (Sol), Insoluble fraction (Ins), Flow through fraction (Ft), Wash fraction #1 (W1), Wash fraction #2 (W2) and Elution fraction (E) were loaded for LVL291 and LVL315.

See Figure 4 for SDS-PAGE of induced bacterial and purification extracts for fusion protein construct LVL312. Culture Media fraction (M), Soluble fraction (Sol), Insoluble fraction (Ins), Flow Through fraction (Ft), Wash fraction #1 (W1), Wash fraction #2 (W2) and Elution fraction (E) were loaded for LVL312.

See Figure 25 for SDS-PAGE of soluble fractions from induced bacterial extracts for fusion protein constructs LVL291, LVL702, LVL736, LVL737, LVL738, LVL739, LVL740 and pET26b vector (negative control). (a) Experiment 1 (b) Experiment 2 (c) Experiment 3. PE-PilA fusion protein indicated by arrow.

See Figure 26 for the average band percentage of fusion protein in the soluble fraction from Experiments 1, 2 and 3.

LVL317 and LVL318 bacterial extracts used in the SDS-PAGE analysis in Figure 5 and Figure 6, respectively, were prepared generally as described above.

Figure 5. SDS-PAGE of induced (1mM and 10 $\mu$ M IPTG) bacterial extracts for fusion protein construct LVL317. Extracts from before (NI) and after induction (In), Soluble fraction (S), Insoluble fraction (I).

Figure 6. SDS-PAGE of induced (1mM and 10 $\mu$ M IPTG) bacterial extracts for fusion protein construct LVL318. Extracts from before (NI) and after induction (In), Culture Media fraction (M), Soluble fraction (S), Insoluble fraction (I).

Proteins separate by SDS-PAGE were transferred to an Immobilon-P membrane. The Coomassie Blue stained protein bands were cut and placed in a sequenator reactor. Sequencing was carried out according to manufacturer's protocol using an Applied Biosystems PROCISE® Protein Sequencer, model 494-cLC.

Table 5: Shake flask protein expression profiles and signal peptide cleavage for fusion protein constructs.

| Fusion Protein Construct ID | Description<br>N-term → C-term                                 | Protein Expression profile   | Signal peptide cleavage |
|-----------------------------|----------------------------------------------------------------|------------------------------|-------------------------|
| LVL312                      | FlgI sp – E – PilA fragment – GG – PE fragment – GGHHHHHH      | In: +++<br>So: +<br>Se: +    | Confirmed               |
| LVL291                      | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH          | In : +++<br>So : ++<br>Se: + | Confirmed               |
| LVL268                      | PelB sp – D – PE fragment – GG – PilA fragment – GGHHHHHH      | In : +++<br>So : ++<br>Se: + | Confirmed               |
| LVL269                      | NadA sp – ATNDDD – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: + | Confirmed               |
| LVL270                      | MHHHHHH – PE fragment – GG – PilA fragment                     | In : +<br>So : –<br>Se: –    | Not tested              |
| LVL315                      | PelB sp – MD – PE fragment – GG – PilA fragment – GGHHHHHH     | In : +++<br>So : ++<br>Se: + | Confirmed               |
| LVL317                      | PelB – PE fragment – GG – PilA fragment                        | In : +++<br>So : +<br>Se: Nt | Confirmed               |

|        |                                                       |                               |           |
|--------|-------------------------------------------------------|-------------------------------|-----------|
| LVL318 | PelB sp – MD – PE fragment – GG – PilA fragment       | In : +++<br>So : +<br>Se: -   |           |
| LVL702 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |
| LVL736 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |
| LVL737 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |
| LVL738 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |
| LVL739 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |
| LVL740 | PelB sp – PE fragment – GG – PilA fragment – GGHHHHHH | In : +++<br>So : ++<br>Se: Nt | Confirmed |

So = Soluble fraction. In = Insoluble fraction. Se = Protein Secreted in the media fraction. Nt = Not tested. The following rating were based on a visual inspection (coomassie) + : low expression; ++ : medium expression; +++ : high expression; - : no expression

#### **Example 6: LVL291 Fusion protein characterization**

##### **PHYSICAL PROPERTIES OF LVL291: Folding of PE and PilA in LVL291 & Melting Point**

##### **Circular Dichromism :**

##### **Analysis of Secondary Structure**

Circular dichroism (CD) is used to determine the secondary structure composition of a protein by measuring the difference in the absorption of left-handed polarized light versus right-handed polarized light which is due to structural asymmetry. The shape and the magnitude of the CD spectra in the far-UV region (190-250nm) are different whether a protein exhibits a beta-sheet,

alpha-helix or random coil structure. The relative abundance of each secondary structure type in a given protein sample can be calculated by comparison to reference spectra.

Far UV spectra are measured using an optical path of 0,01cm from 178 to 250nm, with a 1nm resolution and bandwidth on a Jasco J-720 spectropolarimeter. Temperature of the cell is maintained at 23°C by a Peltier thermostated RTE-111 cell block. A nitrogen flow of 10L/min is maintained during the measurements.

#### *Results:*

The far-UV CD spectra obtained for PE (from construct pRIT16762), PilA (from construct pRIT 16790) and PE-PilA proteins are characteristic of folded proteins containing a mix of alpha and beta structures, but PE is significantly richer in alpha helix than PilA and PE-PilA (Figure 7, CD spectra of PE, PilA and PE-PilA fusion proteins).

In order to evaluate the integrity of the folding of PE and PilA individual proteins once bound together in a chimeric protein and then verify a possible interaction between both, difference spectra were calculated.

- When the PE and PilA far-UV spectra are combined, the resulting spectrum superposes to the spectrum of PE-PilA chimera (Figure 8, Combination of PE and PilA CD spectrum). This result suggests that the PE-PilA chimera contains all the secondary structures that are detected in the individual components. It also suggests that the fusion of the proteins has no major impact on the secondary structures of the individual components and consequently that the folding of PE and PilA is not significantly different whether the proteins are separate or in fusion.

#### Melting Point Evaluation:

In order to evaluate if the expression in fusion has an impact on the thermodynamic properties of the individual proteins, the melting points of PE, PilA and PE-PilA have been evaluated by monitoring the defolding of the alpha helix with temperature by circular dichroism.

The presence of alpha helix is characterized by a minimum in the Circular dichroism signal at 222nm, so a significant increase in CD signal at 222nm during temperature increase is an indication of protein denaturation. The determination of the temperature at which the protein

undergoes loss in secondary structure allows the determination of the melting point ( $T_m$ ), which corresponds to the temperature at which half of the proteins have lost their structure.

Melting point can be determined by identification of the inflexion point on the thermal denaturation curve obtained from a temperature versus CD 222nm plot.

- Melting point of PilA and PE as determined by far-UV CD are respectively of 52°C and 68°C (Figure 9, PilA thermal denaturation curve; Figure 10, PE thermal denaturation curve).
- The PE-PilA fusion protein exhibits two distinct  $T_m$ 's at 48°C and 71°C (Figure 11, PE-PilA fusion protein thermal denaturation curve). Those values indicate that the PE and PilA proteins are still independently folded when bound into a chimera and that they defold at a similar temperature whether they are separate or in fusion. The observation that the defolding of the PilA portion at 48°C doesn't cause precipitation or impact the  $T_m$  of the PE portion at 71°C is a strong indication that the interaction between PE and PilA within the fusion is minimal and that they don't have a major observable impact on each other. The melting points of proteins are sensitive to various external conditions, including buffer composition or presence of interacting molecules; that no major variation is observed upon fusion of PE and PilA is a strong indication of the preservation of most of the structure and of the properties of both PE and PilA when they are bound together.

#### **Example 7: Fermentation process**

Fusion proteins of the invention may be prepared by methods known by those skilled in the art.

#### **Example 8: Protein Purification of PE, PilA, and LVL317**

##### **PE protein purification from pRIT16762:**

To generate the pRIT16762 expression vector, the pRIT16711 vector was digested using BamHI and NcoI restriction enzymes in order to delete 6 amino acid residues between the signal sequence (pelB) and PE. The vector obtained was named pRIT16712. In this vector, there are 3 amino acids between the signal sequence pelB and PE: MDP. In a second step, a site directed mutagenesis was performed to change amino acid sequence from MDP to QIQ using pRIT16712 as template with primers MnoNTHi-44 and MnoNTHi-45 (described in Table 4) and the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene Division).

Working seed of *E. coli* BLR(DE3) containing PE QIQ (from the pRIT16762 construct) was thawed from -80°C and used to prepare 100 ml of pre-culture in LB broth by overnight incubation at 37°C under agitation at 215 RPM. After overnight incubation, eight flasks containing 800 ml of LB APS were inoculated with 12.5 ml of pre-culture and OD<sub>600</sub> measured at around 0.06. The cultures were incubated 3h at 37°C with shaking. At a OD<sub>600</sub> of around 0.9, 1mM IPTG was added to start the induction. During the induction, the cultures were incubated 19h at 22°C with shaking. After induction, OD<sub>600</sub> was at around 2.2. The cell cultures were transferred into 1L centrifuge bags placed inside 1L bottles and centrifuged at 4°C for 30 minutes at 6,000xg and supernatant discarded. 1ml aliquots of culture pre- and post-induction and supernatant were kept for future analysis.

#### *Lysis of the BLR(DE3) induced with PE QIQ*

The centrifuge bags were removed from the centrifugation bottles, opened and the pellet was expelled from the bag into a beaker. The eight pellets were pulled together and resuspended in 100ml of binding buffer (20mM Hepes, 10mM imidazole, 500mM NaCl, pH 8.01). The *E.coli* BLR (DE3) containing the PE QIQ construct were disrupted with the TS Series Bench Top cell disrupter from Constant Systems Ltd. (1x30 kPsi; 1x15kPsi). The lysate was centrifuged 30 minutes, 6000RPM, 4°C. The supernatant was kept and loaded on an IMAC column.

#### *IMAC purification of PE QIQ*

IMAC column (BioRad, Bio-Scale Mini Profinity IMAC cartridge 5ml) was equilibrated with 5CV of Binding buffer (20mM HEPES, 10mM imidazole, 500mM NaCl, pH 8.01) at 5ml/min. 100ml of lysate supernatant was loaded on the IMAC at 2.5mL/min. Flow-through was collected in 50ml fractions for future analysis. The column was washed with 3CV of Binding buffer to remove unbound protein. Sample containing unbound proteins was collected in one aliquot of 15 ml in a 50 ml tube. The column was washed with 2CV of Wash buffer (20mM HEPES, 20mM imidazole, 500mM NaCl, pH 8.01) collected in 2 ml fractions in a 96 well plate. The bound protein was then eluted with 6CV of 100% Elution buffer (20mM HEPES, 250mM imidazole, 500mM NaCl, pH 8.01). The eluted protein was collected in 2 ml fractions in 96-well plates. Wash and elution were performed at 5ml/min.

#### *Size exclusion chromatography (SEC) on the IMAC pool of PE QIQ*



SEC column (GE healthcare, HILOAD™ 26/60 SUPERDEX™ 75 prep grade, 60cm height approx 319ml volume) was equilibrated with 3CV of SEC buffer (20mM HEPES, 150mM NaCl, pH8.49). 11 ml of IMAC eluate was loaded onto the column at a flow rate of 2.5 ml/min. 2ml fractions were collected from 0.3CV to 0.9CV. Two runs were performed then fractions were analyzed by SDS-PAGE. Fractions from the two runs containing Prot E protein were pooled together ("SEC pool", 48ml approx total volume). 500mM of Arginine was added to the SEC pool.

*Dosage of the PE QIQ pooled samples generated in the above SEC protocol*

The SEC pool was dosed with the RCDC (Reducing Agent and Detergent Compatible) method from the Bio-Rad RC DC™ kit following manufacturer's protocol:

For each tested sample and standard, 25µL was distributed in microfuge tubes in duplicate. 125µL of Bio-Rad RC Reagent I was added into each tube; each tube was vortexed and incubate for 1 minute at room temperature. 125µL of Bio-Rad RC Reagent II is added into each tube; each tube is vortexed and then centrifuged at 14,000xg for 5 minutes.

Supernatants are discarded by inverting the tubes on clean, adsorbent tissue paper allowing the liquid to drain completely from the tubes. 25.4µL of Reagent A (already prepared by mixing 20µL of Reagent S per 1ml of Reagent A) is added to each tube; each tube is vortexed and incubated at room temperature for 5 minutes, or until precipitate is completely dissolved. Vortex before proceeding to next step. Add 200µL of DC reagent B to each tube and vortex immediately. Incubate at room temperature for 15 minutes. Transfer all samples to a 96-well plate and read the adsorbance at 750nm to determine the protein concentration for each unknown protein sample.

The ProtE concentration was 1.069 mg/ml

*PilA His-tagged protein purification:*

PilA was purified following the general procedure below:

*E. coli* cells containing a construct encoding PilA or a fragment thereof are suspended in BUGBUSTER® and BENZONASE® nuclease (NOVAGEN®), for example 10 ml BUGBUSTER® and 10 ul BENZONASE® nuclease. The cell lysate is mixed at room temperature on a rotating platform, for example, for 15 minutes. The cell lysate is centrifuged at 4°C, for example at 16,000g for 20 minutes. The supernatant containing the protein is

added to a Ni NTA column containing Ni NTA HIS ·BIND® resin and mixed at 4°C, for example for 1 hour. The column may consist of 2 ml of Ni NTA HIS ·BIND® resin (NOVAGEN®) and 10 ml 1X Binding Buffer (from NOVAGEN®'s Ni-NTA Buffer Kit). The column flow through is then collected. The resin is washed two times with 1X wash buffer, for example, containing 300 mM NaCl, 50mM NaH<sub>2</sub>PO<sub>4</sub>, 25 mM imidazole, pH 8.0). The wash is collected by gravity flow. The protein is eluted from the column with 1X elution buffer, for example, 300 mM NaCl, 50mM NaH<sub>2</sub>PO<sub>4</sub>, 250 mM imidazole, pH 8.0. The protein may be further purified by dialysis with the Binding Buffer and rerun over a Ni NTA column as described above.

#### *Thrombin cleavage of PilA.*

PilA is then incubated with thrombin (diluted 1/50) at room temperature for 16h, to remove the histidine tag.

#### *Size exclusion chromatography (SEC) on PilA cleaved with thrombin.*

SEC column (GE healthcare, HILOAD™ 26/60 SUPERDEX™ 75 prep grade, 60cm height approx 319ml volume) was equilibrated with 5CV of SEC buffer (20mM HEPES, 150mM NaCl, pH8.52). Approximately 10 ml of cleaved PilA was loaded onto the column at a flow rate of 2.5 ml/min. 2ml fractions collected from 0.3CV to 0.9CV. Two runs were performed then fractions were analyzed by SDS-PAGE. Fractions from the two runs containing cleaved PilA protein were pooled together ("SEC pool", 52ml approx total volume).

#### *Dosage of PilA, SEC pool.*

The SEC pool was dosed with the RCDC method as described above. The cleaved PilA concentration was at 5.37 mg/ml.

#### *Dialysis of the PilA SEC pool with PBS 1x pH 7.4 (dialysis factor = 1600) and dosage by RCDC*

The concentration post-dialysis determined by RCDC was at 3.0 mg/ml.

#### *Purification of LVL317*

##### *Osmotic shock*

Since LVL317 fusion protein is expressed and processed in bacterial periplasm, the protein was extracted by osmotic shock.

Frozen (-20°C) harvested *E. coli* B2448 cell paste containing LVL317 from 4 L of fermentor culture were pooled and resuspended in a hypertonic buffer consisting of 24 mM Tris-HCl, 16% (w/v) sucrose, 9.9% (w/v) glucose, 10 mM EDTA, pH 8.0 up to a final volume of 4L. The suspension was mixed gently for 30 min at room temperature using a 3-blade propeller installed on RW 16 basic stirrer, at medium speed. The suspension was centrifuged at 15,900 x g for 30 minutes at room temperature. Supernatant (SN1) was kept for gel analysis.

The resulting pellet was resuspended in a hypotonic solution; 38 mM MgCl<sub>2</sub>, and mixed for 30 min at room temperature. The mixture was centrifuged at 15,900 x g for 30 minutes at room temperature and the antigen recovered in the supernatant (SN2).

A clarification of the SN2 was performed by filtration through a 0.45/0.2 µm polyethersulfone Sartorius Sartopore 2 MidiCap filter, at 600ml/min of flow rate.

The SN2 was diluted 1:3 with 20 mM NaH<sub>2</sub>PO<sub>4</sub>-Na<sub>2</sub>HPO<sub>4</sub>, pH 7.0, the pH adjusted to 7.0 if necessary and another clarification by filtration through a 0.45/0.2 µm polyethersulfone Sartorius Sartopore 2 MidiCap filter, at 600ml/min was performed.

#### *SP SEPHAROSE™ Fast Flow (SP FF) chromatography*

The diluted/filtered SN2 was loaded and captured on a strong cationic exchanger resin (SP SEPHAROSE™ FF - GE Healthcare) in a 14 cm ID (internal diameter) x 20 cm length column (column volume 3100ml) equilibrated with 2CV of 20 mM NaH<sub>2</sub>PO<sub>4</sub> / Na<sub>2</sub>HPO<sub>4</sub> buffer pH 7.0. After washing the column with 5CV of 20 mM NaH<sub>2</sub>PO<sub>4</sub> / Na<sub>2</sub>HPO<sub>4</sub> buffer pH 7.0, the antigen (contained within LVL317) was eluted by increasing the concentration of NaCl up to 100 mM in the same washing buffer.

See Figure 12 for a typical SP SEPHAROSE™ Fast Flow chromatogram.

#### *Q SEPHAROSE™ Fast Flow (Q FF) chromatography*

The antigen present in the SP FF Eluate was diluted 1:4 with a 20 mM Tris pH 8.5, pH adjusted to 8.5 if necessary and passed through a strong anionic exchanger resin (Q SEPHAROSE™ FF - GE Healthcare) in a 14 cm ID x 11.8 cm length column (column volume

1800ml) equilibrated with 2CV of 20 mM Tris buffer pH 8.5. The antigen was recovered in the flow-through fraction.

See Figure 13 for a typical Q SEPHAROSE™ Fast Flow chromatogram.

#### *Concentration, diafiltration, polysorbate 80 addition and sterile filtration*

The Q FF flow-through containing the antigen was concentrated up to 0.7-0.8mg/ml based on chromatogram UV and diafiltered with 5DV of 10 mM KH<sub>2</sub>PO<sub>4</sub> / K<sub>2</sub>HPO<sub>4</sub> buffer pH 6.5 using a Pellicon-2™ 10 kDa cutoff membrane (Millipore).

Using a 5% stock solution, polysorbate 80 (for example, TWEEN™ 80) was added to the ultrafiltration retentate and agitated for 30 minutes with magnetic stirrer at 130rpm at 4°C. The final concentration of polysorbate 80 was 0.04%. Ultrafiltration retentate was sterilized by filtration through a 0.45/0.2 µm Cellulose Acetate membrane (Sartobran 300, Sartorius). The purified bulk was stored at -20°C or -80°C. Absolute protein concentration was measured by AAA (Amino Acid Analysis) at 0.737mg/ml.

#### **Example 9: Use of Polysorbate 80**

A titration experiment indicated that the addition of polysorbate 80, specifically, TWEEN™ 80 to a final concentration of 0.04% (w/v) to the purified bulk prior to sterile filtration reduced filamentous particle formation and aggregation.

According to DSC analysis, TWEEN™ 80 reduced the degree of structural change (30-45°C) seen after freeze/thaw cycles after storage at -20°C and after storage 4 days at 4°C, -20°C and -80°C and 37°C.

#### **Example 10: SDS-PAGE and Western Blot Analysis of LVL317**

##### **SDS-PAGE and Western Blot analysis:**

NUPAGE®, Bis-Tris 4-12% gel was loaded as described below with 10µg of sample in NUPAGE® LD sample buffer containing 50mM DTT heated 5min at 95°C (20µL of sample was loaded for samples having low concentration). Migration: 35 minutes at 200Volts at room temperature (RT) in NUPAGE® MES Running Buffer. Gel Stained 2 hours in Instant blue (Novexin cat.: ISB01L) and destained

overnight in water.

Lane contents:

|                                                                                                                                 |                                                                           |                                  |
|---------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------|
| 1: MW standard (10 $\mu$ L)                                                                                                     | 2: Start (total fraction) (10 $\mu$ g)                                    | 3: SN1 non filtered (10 $\mu$ g) |
| 4: SN2 not filtered (10 $\mu$ g)                                                                                                | 5: Not extracted (10 $\mu$ g)                                             | 6: Load SP FF (10 $\mu$ g)       |
| 7: Flow through SP FF (6.9 $\mu$ g)                                                                                             | 8: Wash SP FF (20 $\mu$ L)                                                | 9: Elution SP FF (10 $\mu$ g)    |
| 10: Strip SP FF (10 $\mu$ g)                                                                                                    | 11: Load Q FF (8.9 $\mu$ g)                                               | 12: Elution Q FF (9.8 $\mu$ g)   |
| 13: Strip Q FF (4.8 $\mu$ g)                                                                                                    | 14: TFF retentate before 0.04% TWEEN <sup>TM</sup> 80 spiked (10 $\mu$ g) |                                  |
| 15: Purified bulk Not filtered 0.04% TWEEN <sup>TM</sup> 80 spiked (10 $\mu$ g)                                                 |                                                                           |                                  |
| 16: Purified bulk Sterile Filtered 0.04% TWEEN <sup>TM</sup> 80 spiked (10 $\mu$ g)                                             |                                                                           |                                  |
| 17: Purified bulk Sterile Filtered 0.04% TWEEN <sup>TM</sup> 80 spiked (20 $\mu$ g + spiked E. Coli Cell lysate Rix (1 $\mu$ g) |                                                                           |                                  |
| 18: E. Coli Cell lysate Rix (2 $\mu$ g)                                                                                         |                                                                           |                                  |
| 19: E. Coli Cell lysate Rix (1 $\mu$ g)                                                                                         |                                                                           |                                  |
| 20: E. Coli Cell lysate Rix (0.5 $\mu$ g)                                                                                       |                                                                           |                                  |

See Figure 14 for a SDS-PAGE of In-process samples from purification process of PE-PilA fusion protein.

For Western Blot, proteins were transferred at 4°C overnight at 30Volts in NUPAGE<sup>®</sup> transfer buffer + 20% Methanol, 0.1% SDS on nitrocellulose membrane. Membranes were blocked 1 hour with 50mM Tris, 150mM NaCl pH 7.4 + 5% non-fat dry milk, incubated 2 hours in rabbit polyclonal primary antibody diluted in blocking buffer (anti-Prot-E 1/50 000 and anti-Ecoli (BLR) 1/1 000), washed 3x5minutes in 50mM Tris pH 7.4 + 0.05% Tween 20, incubated 1 hour in secondary antibody (goat anti-rabbit conjugated to alkaline phosphatase diluted 1/5000 in blocking buffer), washed 3x5minutes in wash buffer and developed in BCIP/NBT substrate (1 tablet per 10ml). All incubations performed in 25ml per membrane.

See Figure 15 for a Western Blot of In-process samples of purification process from PE-PilA fusion protein. Blot using rabbit polyclonal anti-PE.

Lane contents:

|                                     |                                                                           |                                  |
|-------------------------------------|---------------------------------------------------------------------------|----------------------------------|
| 1: MW standard (10 $\mu$ L)         | 2: Start (total fraction) (10 $\mu$ g)                                    | 3: SN1 non filtered (10 $\mu$ g) |
| 4: SN2 not filtered (10 $\mu$ g)    | 5: Not extracted (10 $\mu$ g)                                             | 6: Load SP FF (10 $\mu$ g)       |
| 7: Flow through SP FF (6.9 $\mu$ g) | 8: Wash SP FF (20 $\mu$ L)                                                | 9: Elution SP FF (10 $\mu$ g)    |
| 10: Strip SP FF (10 $\mu$ g)        | 11: Load Q FF (8.9 $\mu$ g)                                               | 12: Elution Q FF (9.8 $\mu$ g)   |
| 13: Strip Q FF (4.8 $\mu$ g)        | 14: TFF retentate before 0.04% TWEEN <sup>TM</sup> 80 spiked (10 $\mu$ g) |                                  |

- 15: Purified bulk Not filtered 0.04% TWEEN™ 80 spiked (10µg)
- 16: Purified bulk Sterile Filtered 0.04% TWEEN™ 80 spiked (10µg)
- 17: Purified bulk Sterile Filtered 0.04% TWEEN™ 80 spiked (20µg + spiked E. Coli Cell lysate Rix (1µg)
- 18: E. Coli Cell lysate Rix (2µg)
- 19: E. Coli Cell lysate Rix (1µg)
- 20: E. Coli Cell lysate Rix (0.5µg)

See Figure 16 for a Western Blot of In-process samples of purification process from PE-PilA fusion protein. Blot using rabbit polyclonal anti-*E.coli* (BLR).

Lane contents:

- |                                                                                                        |                                                        |                            |
|--------------------------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------------|
| 1: MW standard (10µL)                                                                                  | 2: Start (total fraction) (10µg)                       | 3: SN1 non filtered (10µg) |
| 4: SN2 not filtered (10µg)                                                                             | 5: Not extracted (10µg)                                | 6: Load SP FF (10µg)       |
| 7: Flow through SP FF (6.9µg)                                                                          | 8: Wash SP FF (20µL)                                   | 9: Elution SP FF (10µg)    |
| 10: Strip SP FF (10µg)                                                                                 | 11: Load Q FF (8.9µg)                                  | 12: Elution Q FF (9.8µg)   |
| 13: Strip Q FF (4.8µg)                                                                                 | 14: TFF retentate before 0.04% TWEEN™ 80 spiked (10µg) |                            |
| 15: Purified bulk Not filtered 0.04% TWEEN™ 80 spiked (10µg)                                           |                                                        |                            |
| 16: Purified bulk Sterile Filtered 0.04% TWEEN™ 80 spiked (10µg)                                       |                                                        |                            |
| 17: Purified bulk Sterile Filtered 0.04% TWEEN™ 80 spiked (20µg + spiked E. Coli Cell lysate Rix (1µg) |                                                        |                            |
| 18: E. Coli Cell lysate Rix (2µg)                                                                      |                                                        |                            |
| 19: E. Coli Cell lysate Rix (1µg)                                                                      |                                                        |                            |
| 20: E. Coli Cell lysate Rix (0.5µg)                                                                    |                                                        |                            |

**SDS-PAGE and Western Blot figures comments:** The PE-PilA fusion protein migrates at 30kDa. The extraction by osmotic shock extracts the fusion protein expressed and processed in bacteria periplasm and reduced contamination from bacteria. Small loss of fusion protein during hypertonic treatment (lane 3). A small proportion is not extracted by hypotonic treatment and remains associated with cells (lane 5). Small loss in SP FF Flow through (lane 7) and in strip fraction of both columns (lanes 10 and 13). Since the total volume of strip fraction is low the loss of fusion protein is not significant. Degraded bands are visible in strip fractions but not in final product. No significant contamination from *E. coli* host cell proteins in purified bulk (lane 16).

Analysis of LVL735 and LVL778 yielded similar profiles as LVL317.

### **Example 11: Melting Point Data for PE, PilA and LVL317**

Thermal transition of PE-PilA fusion non His-tagged protein (LVL317) was compared with the thermal transition of both PE his-tagged (as described in Example 8) and cleaved PilA (as described in Example 8) proteins, purified as described above.

Before DSC, PE and PilA were dialyzed overnight in 10mM K<sub>2</sub>HPO<sub>4</sub>/KH<sub>2</sub>PO<sub>4</sub> pH 6.5 + 0.04% Tween 80 (1:250 sample:buffer volume ratio) to have them in the same buffer as the fusion protein. After dialysis, proteins concentration was measured by BCA and adjusted to 300µg/ml (PE) and 500µg/ml (PilA).

Analysis done on VP<sup>TM</sup>-DSC from MicroCal, LLC (part of GE Healthcare). The final dialysis buffer was used as reference and subtracted from the scans. DSC scan rate 90°C/hr. In order to evaluate the capacity to measure the thermal transition in the Final Container (FC) after formulation, the fusion protein was diluted to the FC concentration (60µg/ml). Final container data not shown.

#### *Results:*

See Figure 17 for Thermal transition of PE-PilA fusion protein and PE and PilA proteins.

Curves: PilA (1), Protein E (Prot E, PE) (2), PE-PilA PB not diluted 737µg/ml (3), and PE-PilA PB diluted at FC concentration 60µg/ml (4).

|                                                     |                                                       |
|-----------------------------------------------------|-------------------------------------------------------|
| 1 – PilA                                            | T <sub>m</sub> : 53°C                                 |
| 2 – Protein E                                       | T <sub>m</sub> : 63                                   |
| 3 – PE-PilA PB (Purified Bulk) not diluted 737µg/ml | T <sub>m1</sub> : 53.7°C and T <sub>m2</sub> : 66.1°C |
| 4 – PE-PilA PB diluted at FC concentration 60µg/ml  | T <sub>m1</sub> : 53.2°C and T <sub>m2</sub> : 67.6°C |

Two transitions were detected in the purified fusion protein (LVL317) (curves 3 and 4).

The T<sub>m1</sub> (53.7°C) of the PE-PilA fusion protein is similar to PilA transition (53°C).

Significant shift of T<sub>m2</sub> in PE-PilA (66.1°C) as compared to PE transition (63°C). The fusion of both domains seems to stabilize the PE fragment.

The shift of  $Tm_2$  in the diluted fusion protein as compared to undiluted is a concentration artifact arising from the steep decreasing slope typical of aggregation which is concentration dependant.

Antigen folding analysis of LVL735 and LVL778 were similar to that of LVL317.

**Example 12: PE-PilA fusion protein construct LVL291 anti-PilA immunogenicity response in Balb/c mice.**

The immune response directed against purified LVL291 PE-PilA fusion protein (the LVL291 fusion protein without the heterologous signal peptide) formulated in AS03<sub>A</sub> was evaluated in Balb/c mice. Animals (20 mice/group) were immunized by the intramuscular route at days 0, 14 and 28 with 10 µg of PE (from vector pRIT16762), PilA (from vector pRIT16790) or PE-PilA, each formulated in AS03<sub>A</sub>. The control group was vaccinated with AS03<sub>A</sub> alone. Antibody response directed against each antigen was determined in individual sera collected at day 42. No antibody response was obtained with the negative control. As shown in Figure 18, the antibody response directed against PilA was higher in mice immunized with the PE-PilA fusion compared to antibody response in mice immunized with monovalent PilA. The antibody responses directed against PE were similar in mice immunized with the fusion protein and mice immunized with monovalent PE. GMT = geometric means titer. Data were captured and analyzed with the SOFTMAX<sup>®</sup> Pro Software (Molecular Devices) running under WINDOWS<sup>®</sup> (Microsoft); the four parameters logistic log function was used to calculate the standard curve. The four-parameter logistic-log function describes, with a high degree of accuracy, the curve of the reference serum displaying a pronounced sigmoidal shape when plotted on an optical density-versus-concentration (log) scale. Antibody concentrations were calculated at each dilution of mice serum samples by interpolation of the standard curve. The antibody in quality control sera and in unknown serum samples is obtained by averaging the values from all dilutions that fall within the working range (10-80 %) of the dilution curve of the reference.

Results are shown in Figure 18, which graphs the antibody responses against LVL291 PE-PilA fusion protein and against monovalent PE and PilA in the Balb/c mouse model.



**Example 13: Murine nasopharyngeal colonization model. Immunization with PE-PilA.**  
**Challenge with NTHi strain 86-028NP and NTHi strain 3224A.**

Balb/c female mice (20/group) were immunized intranasally at days 0 and 14 with 6µg of a purified PE-PilA fusion protein (LVL291 for challenge with 86-028NP; LVL317 for challenge with strain 3224A) formulated with LT (heat labile toxin of *Escheria coli*) and on day 28 with 6 µg of a purified PE-PilA fusion protein in phosphate buffered saline (PBS). Control mice (20/group) were vaccinated with LT alone. Mice were subsequently challenged intranasally with  $5 \times 10^6$  CFU (colony forming units) of homologous NTHi strain 86-028NP and heterologous NTHi strain 3224A. Homology and heterology are determined by reference to the NTHi strain with which the mice were immunized. Bacterial colonies were counted in nasal cavities removed 1 and 2 days after the challenge. D1 = day 1. D2 = day 2.

PE-PilA vaccination increased the clearance of NTHi strain 86-028NP and strain 3224A in the nasopharynx at day 1 and day 2 post challenge.

For the experiment performed with NTHi strain 86-028NP: A 2-way fixed ANOVA was performed using the log10 values of the counts as response, the fixed factors being the group (4 levels) and the day (2 levels). The assumption of variance heterogeneity was rejected and a model with heterogeneous variances was fitted to the data. No significant interaction was detected between the 2 factors. The group fusion PE-PilA (6 µg per mouse) significantly reduced CFU compared with the control group (LT); the geometric mean ratio being equal to 0.06 with a 95% confidence interval of 0.01, 0.25.

For the experiment conducted with NTHi strain 3224A: A 3-way fixed ANOVA was performed using the log10 values as response, the fixed factors being the group, the day, and the experiment. The Shapiro-Wilk and Levene's test did not reject the assumptions of normality and of homogeneity of variances. No significant interaction between any of the 2 factors or between the 3 factors was detected and only main factors were kept in the analysis. PE-PilA / LT significantly reduced CFU compared with the control group; the geometric mean ratio being equal to 0.11 with a 95% confidence interval of 0.02, 0.61.

See Figure 19 for effect of PE-PilA fusion protein vaccination on NTHi strain 86-028NP bacterial clearance in mouse nasopharynx.

See Figure 20 for effect of PE-PilA fusion protein vaccination on NTHi strain 3224A bacterial clearance in mouse nasopharynx.

**Example 14: Murine nasopharyngeal colonization model. Immunization with PilA. Challenge with NTHi strain 3219C.**

Female OF1 mice (20 mice/group) were immunized intranasally at days 0 and 14 with 3µg PilA (from vector 16790) formulated with LT and at day 28 with 3 µg PilA in PBS. Control mice were vaccinated with LT alone. Mice were subsequently challenged intranasally with  $5 \times 10^6$  CFU of NTHi strain 3219C. Bacterial colonies were counted in nasal cavities removed 3 and 4 days after the challenge. D3 = day 3. D4 = day 4.

See Figure 21 for effect of PilA vaccination on bacterial clearance in mouse nasopharynx.

**Example 15: Murine nasopharyngeal colonization model. Immunization with PE. Challenge with NTHi strain 3224A.**

Balb/c female mice (20 mice/group) were immunized intranasally at days 0 and 14 with 3µg PE (from vector pRIT16762) formulated with LT and at day 28 with 3 µg PE in PBS. Control mice were vaccinated with LT alone. Mice were subsequently challenged intranasally with  $5 \times 10^6$  CFU of NTHi strain 3224A. Bacterial colonies were counted in nasal cavities removed 3 and 4 days after the challenge. 10 mice were examined on day 3 (D3). 10 mice were examined on day 4 (D4). PE vaccination increased significantly the clearance of NTHi in the naso-pharynx at day 4 post challenge (Figure 22), using on the Dunn test for statistical analysis.

See Figure 22 for effect of PE vaccination on bacterial clearance in the nasopharynx of mice.

**Example 16: Vibronectin binding. Inhibition of vibronectin binding by LVL317 & LVL735 PE-PilA fusion protein.**

The ability of PE in the purified LVL317 PE-PilA fusion protein construct to bind to vitronectin was evaluated. Microtiter plates (POLYSORP™, Nunc, Thermo Fisher Scientific) were coated

with PE (from vector pRIT16762) or with purified LVL317 PE-PilA fusion protein (10 µg/ml). Plates were washed four times with NaCl 150mM-polysorbate 20, 0.05% (for example, TWEEN™ 20) and blocked for one to two hours with PBS-BSA 1%. After four washings, vitronectin (Vitronectin from human plasma, SIGMA-ALDRICH®) was added (10 µg/ml), two fold diluted (12 dilutions), and the plates were incubated for 1h at room temperature. The plates were then washed 4 times with NaCl 150mM-polysorbate 20, 0.05% (for example TWEEN™ 20) After washings, the bound vitronectin was detected using peroxidase sheep anti-human vitronectin (US Biological) followed by the addition of ortho-phenylene diamine/H<sub>2</sub>O<sub>2</sub> substrate. The color developed is directly proportional to the amount of antibody fixed to the vitronectin.

See Figure 23 for (a) LVL317 PE-PilA fusion protein bound to vitronectin. PilA = PilA from NTHi strain 86-028NP (as described for pRIT16790); PE = Protein E (as described for pRIT16762) and (b) LVL317 and LVL735 PE-PilA fusion protein bound to vitronectin.

**Example 17: Vitronectin binding. Inhibition of vitronectin binding by antibodies directed against the LVL291 PE-PilA fusion protein.**

Microtiter plates (POLYSORP™, Nunc, Thermo Fisher Scientific) were coated with PE (from vector pRIT16762) or with purified PE-PilA fusion protein (10 µg/ml). Plates were washed four times with NaCl 150mM-polysorbate 20, 0.05% (for example, TWEEN™ 20) and blocked for two hours with PBS-BSA 1%. After washings, vitronectin (Vitronectin from human plasma, SIGMA-ALDRICH®) was added at 50µg/ml and purified antibodies anti-PE-PilA (produced and purified in house) were two-fold serially diluted and incubated for 1h at room temperature. The plates were then washed 4 times with NaCl 150mM-polysorbate 20, 0.05% (for example, TWEEN™ 20). After four washings, the bound vitronectin was detected using peroxidase sheep anti-Vitronectin (US Biological) followed by the addition of ortho-phenylene diamine/H<sub>2</sub>O<sub>2</sub> substrate. The color developed is directly proportional to the amount of antibody fixed to the vitronectin.

Inhibition of vitronectin binding to PE by polyclonal antibodies directed against PE-PilA was observed.

See Figure 24 for inhibition of vitronectin binding by polyclonal antibodies against PE-PilA fusion protein.

**Example 18: Antigenicity of LVL291 PE-PilA fusion protein. ELISA.**

Purified LVL291 PE-PilA fusion protein was validated in an antigenicity test with monovalent proteins as control. The fusion protein was tested in a sandwich ELISA developed with polyclonal antibodies (rabbit and guinea pig) generated against the PE gene fragment coding for amino acids 22 to 160 of SEQ ID NO: 4 (as described for pRIT16711) or against PilA from NTHi strain 86-028NP (from vector pRIT16790).

PilA or PE was added at 100 ng/ml and serially two fold diluted. After 30 minutes incubation and after washing, the bound antigen was detected by a rabbit polyclonal serum obtained after immunisation with PE or PilA. The bound antibodies were detected using a peroxidase anti-rabbit Ig (Jackson ImmunoResearch Laboratories, Inc.) followed by the addition of ortho-phenylene-diamine/H<sub>2</sub>O<sub>2</sub> substrate. The color developed is directly proportional to the amount of antigen present. Absorbance readings were measured using a spectrophotometer for microtiter plates. The antigenicity of the samples was determined by comparison to the curve of the full length PE or full length PilA reference antigen and is expressed in ug/ml. The reference represented 100% of antigenicity.

As observed in the Table 6: Antigenicity was observed with the purified LVL291 PE-PilA fusion protein compared to the monovalent PE and PilA antigens.

Table 6 : Relative antigenicity obtained with purified LVL291 PE-PilA fusion protein in the antigenicity test.

|                        | PE relative antigenicity (%) |
|------------------------|------------------------------|
| Protein E as Reference | 100                          |
| PE-PilA                | 130-148                      |

|                   | PE relative antigenicity (%) |
|-------------------|------------------------------|
| PilA as Reference | 100                          |
| PE-PilA           | 120-152                      |

**Example 19: Immunogenicity of LVL735 PE-PilA fusion protein.**

Female Balb/c mice (n = 34) were immunized by the intramuscular route at days 0, 14 and 28 with 50 µl of vaccine formulation containing 1, 0.2 or 0.04 µg of PE-PilA fusion protein LVL317 or LVL735 formulated within AS01<sub>E</sub> or AlPO<sub>4</sub> (aluminium phosphate). The antibody responses to

PE and PilA were determined in individual sera collected at day 42 and the IgG level against PE and PilA was measured and expressed in µg /ml.

See Figure 27 for PE and PilA antibody response to LVL317 and LVL735. GMC= geometric mean concentration . GMT = geometric means titer. IC = confidence intervals.

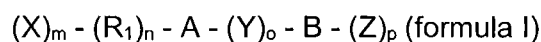
**Example 20: Protective efficacy of the LVL735 and LVL317 fusion proteins in a mouse model of Non-typeable *Haemophilus influenzae* nasopharyngeal colonization.**

Female Balb/c mice were intranasally immunized at days 0 and 14 with 10 µl of vaccine formulation containing 5.8 µg of LVL735 or LVL317 admixed with 0.5 µg of *E. coli* labile toxin (LT). A booster dose of 5.8 µg of non-adjuvanted LVL735 or LVL317 was administered at day 28. Control mice were vaccinated with LT alone at days 0 and 14, and PBS at day 28. Animals were intranasally challenged with  $5 \times 10^6$  cfu of NTHi 3224A strain at day 42. Bacterial colonies were counted in nasal cavities removed 1 and 2 days after the challenge (n = 10/time-point). Nasal cavities are homogenized in medium and a bacterial quantification is performed. Results are well expressed in CFU/ml.

See Figure 28 for the effect of LVL735 and LVL317 vaccination on bacterial clearance in a mouse model of non-typeable *Haemophilus influenzae* nasopharyngeal colonization.

**Claims:**

1. A fusion protein of formula 1:



wherein:

X is a signal peptide or MHHHHHH (SEQ ID NO. 2);

m is 0 or 1;

R<sub>1</sub> is an amino acid;

n is 0, 1, 2, 3, 4, 5 or 6;

A is Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof, or

PilA from *Haemophilus influenzae* or an immunogenic fragment thereof;

Y is selected from the group consisting of GG, SG, SS, GGG and (G)<sub>h</sub> wherein h is 4, 5, 6, 7,

8, 9, or 10;

o is 0 or 1;

B is PilA from *Haemophilus influenzae* or an immunogenic fragment thereof, or

Protein E from *Haemophilus influenzae* or an immunogenic fragment thereof;

Z is GGHHHHHH (SEQ ID NO. 3); and

p is 0 or 1,

wherein when A is Protein E from *Haemophilus influenzae* or an immunogenic fragment

thereof, B is not Protein E from *Haemophilus influenzae* or an immunogenic fragment

thereof, and wherein when A is PilA from *Haemophilus influenzae* or an immunogenic fragment

thereof, B is not PilA from *Haemophilus influenzae* or an immunogenic fragment thereof.

2. A fusion protein according to claim 1 wherein X is selected from the group consisting of FlgI, NadA and pelB.

3. A fusion protein according to claim 1 wherein m is 0.

4. A fusion protein according to any one of claims 1-3 wherein n is 0.

5. A fusion protein according to any one of claims 1-4 wherein A is an immunogenic fragment of Protein E, wherein Protein E is selected from any one of SEQ ID NO. 4 -SEQ ID NO. 57.

6. A fusion protein according to any one of claims 1-5 wherein A is the immunogenic fragment of Protein E from *H. influenzae* as set forth in SEQ ID NO: 124.

7. A fusion protein according to any one of claims 1-6 wherein Y is GG.

8. A fusion protein according to any one of claims 1-7 wherein B is an immunogenic

5 fragment of PilA, wherein PilA is selected from any one of SEQ ID NO. 58 – SEQ ID NO. 121.

9. A fusion protein according to any one of claims 1-8 wherein B is the immunogenic fragment of PilA from *H. influenzae* as set forth in SEQ ID NO. 127.

10 10. A fusion protein selected from the group consisting of SEQ ID NO. 136, SEQ ID NO. 138, SEQ ID NO. 140, SEQ ID NO. 142, SEQ ID NO. 144, SEQ ID NO. 146, SEQ ID NO. 148, SEQ ID NO. 150, SEQ ID NO.182, SEQ ID NO.184, SEQ ID NO.186, SEQ ID NO.188, SEQ ID NO. 190, SEQ ID NO.192, SEQ ID NO.194, SEQ ID NO.196, SEQ ID NO.198, SEQ ID NO.200, SEQ ID NO.202 and SEQ ID NO.204.

15

11. A fusion protein approximately 95% identical to any of SEQ ID NO. 136, SEQ ID NO. 138, SEQ ID NO. 140, SEQ ID NO. 142, SEQ ID NO. 144, SEQ ID NO. 146, SEQ ID NO. 148, SEQ ID NO. 150, SEQ ID NO. 182, SEQ ID NO. 184, SEQ ID NO. 186, SEQ ID NO. 186, SEQ ID NO. 188, SEQ ID NO. 190, SEQ ID NO. 192, SEQ ID NO. 194, SEQ ID NO.

20 196, SEQ ID NO. 198, SEQ ID NO. 200, SEQ ID NO. 202 or SEQ ID NO. 204.

12. A fusion protein of claim 10 or claim 11 wherein the signal peptide has been removed.

13. The fusion protein of SEQ ID NO. 148, wherein the signal peptide has been removed,

25 SEQ ID NO. 177 (QIQKAEQN DVKLAPPTDV RSGYIRLVKN VNYIDSESI  
WVDNQEPQIV HFDVVNLDK GLYVYPEPKR YARSVRQYKI LNCANYHLTQ  
VRTDFYDEFW GQGLRAAPKK QKKHTLSLTP DTTLYNAAQI ICANYGEAFS  
VDKKGGTKKA AVSELLQASA PYKADVELCV YSTNETTNCT GKNGIAADI  
TTAKGYVKS V TTSNGAITVK GDGTLANMEY ILQATGNAAT GVTWTTTCKG  
30 TDASLFPANF CGSVTQ) .

14. The fusion protein of SEQ ID NO. 194, wherein the signal peptide has been removed,

SEQ ID NO. 219 (IQKAEQND VKLAPPTDVR SGYIRLVKNV NYIDSESIW  
VDNQEPQIVH FDAVVNLDKG LYVYPEPKRY ARSVRQYKIL NCANYHLTQV  
35 RTDFYDEFWG QGLRAAPKKQ KKHTLSLTPD TTYNAAQII CANYGEAFSV  
DKKGGTKKAA VSELLQASAP YKADVELCVY STNETTNCTG GKNGIAADIT

TAKGYVKSVT TSNGAITVKG DGTLANMEYI LQATGNAATG VTWTTTCKGT  
DASLFPANFC GSVTQ) .

15. An immunogenic composition comprising isolated Protein E from *H. influenzae* and  
5 isolated PilA from *H. influenzae*.

16. An immunogenic composition of claim 15 wherein Protein E is a polypeptide of SEQ  
ID NO. 4, a polypeptide comprising a sequence having at least 75%, 77%, 80%, 85%, 90%,  
95%, 97%, 99% or 100% identity, over the entire length, to SEQ ID NO. 4, or is a polypeptide  
10 comprising an immunogenic fragment of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino  
acids of SEQ ID NO. 4.

17. The immunogenic composition of claim 16, wherein the immunogenic fragment  
comprises a B and/or T cell epitope of SEQ ID NO: 4.

18. The immunogenic composition of any one of claims 15-17 wherein Protein E is  
capable of eliciting an immune response which recognizes SEQ ID NO. 4.

19. An immunogenic composition of any one of claims 15-18 wherein PilA is a polypeptide  
20 of SEQ ID NO. 58, a polypeptide comprising a sequence having at least 80%, 85%, 90%,  
95%, 97% or 100% identity, over the entire length, to SEQ ID NO. 58, or is a polypeptide  
comprising an immunogenic fragment of at least 7, 10, 15, 20, 25, 30 or 50 contiguous amino  
acids of SEQ ID NO. 58.

20. The immunogenic composition of claim 19, wherein the immunogenic fragment  
comprises a B and/or T cell epitope of SEQ ID NO: 58.

21. An immunogenic composition of any one of claims 15-20 wherein PilA is capable of  
eliciting an immune response which recognizes SEQ ID NO. 58.

22. The immunogenic composition of any one of claims 15-21, wherein the Protein E from  
*H. influenzae* and the PilA from *H. influenzae* are comprised as the fusion protein of any one  
of claims 1-14.

23. An immunogenic composition comprising the fusion protein of SEQ ID NO. 177  
(QIQKAEQN DVKLAPPTDV RSGYIRLVKN VNYIDSESI WVDNQEPQIV



HFDAVVNLDK GLYVYPEPKR YARSVRQYKI LNCANYHLTQ VRTDFYDEFW  
 GQGLRAAPKK QKKHTLSLTP DTTLYNAAQI ICANYGEAFS VDKKGGTKKA  
 AVSELLQASA PYKADVELCV YSTNETTNCT GGKNGIAADI TTAKGYVKSV  
 TTSNGAITVK GDGTLANMEY ILQATGNAAT GVTWTTTCKG TDASLFPANF CGSVTQ).

5

24. An immunogenic composition comprising the fusion protein of SEQ ID NO.  
 219 (IQKAEQND VKLAPPTDVR SGYIRLVKNV NYYIDSESIW VDNQEPQIVH  
 FDAVVNLDKG LYVYPEPKRY ARSVRQYKIL NCANYHLTQV RTDFYDEFWG  
 QGLRAAPKKQ KKHTLSLTPD TTYLYNAAQII CANYGEAFSV DKKGGTKKAA  
 VSELLQASAP YKADVELCVY STNETTNCTG GKNGIAADIT TAKGYVKSVT  
 TSNGAITVKG DGTLANMEYI LQATGNAATG VTWTTTCKGT DASLFPANFC  
 GSVTQ).

10

25. A vaccine comprising the fusion protein of any one of claims 1-14 or  
 immunogenic compositions of any one of claims 15-24.

15

26. The fusion protein of any one of claims 1-14, or the immunogenic composition  
 of any one of claims 15-24, or the vaccine of claim 25, for use in the treatment or  
 prevention of otitis media.

20

27. The fusion protein, the immunogenic composition, or the vaccine of any one  
 of claims 1-25, 26, for use in the treatment or prevention of acute exacerbations of  
 chronic obstructive pulmonary disease (AECOPD).

25

28. The fusion protein, the immunogenic composition, or the vaccine of any one  
 of claims 1-25, 26-27, for use in the treatment or prevention of pneumonia.

29. The fusion protein, the immunogenic composition, or the vaccine of any one  
 of claims 1-25, 26-28, for use in the treatment or prevention of *H. influenzae* infection  
 or disease.

30

30. The fusion protein, the immunogenic composition, or the vaccine of claim 29,  
 for use in the treatment or prevention of NTHi infection or disease.

35

31. A process for producing periplasmic expression of a fusion protein of claim 1  
 wherein the process comprises inducing expression of proteins containing a signal  
 peptide.

32. The process of claim 31 wherein the signal peptide is from FlgI.

33. The process of claim 31 wherein the signal peptide is from pelB.

5        34. The process of any one of claims 31-33 wherein the fusion protein is a fusion protein of formula (1) or the fusion protein of claims 1-14.

10       35. A fusion protein according to claim 1 or 10-14, or an immunogenic composition according to claim 15, 23 or 24, substantially as hereinbefore described, with reference to any of the Examples.

## **FUSION PROTEINS AND COMBINATION VACCINES COMPRISING HAEMOPHILUS INFLUENZAE PROTEIN E AND PILIN A**

### **Abstract**

The present invention relates to compositions comprising *Haemophilus influenzae* Protein E and Pilin A. More particularly, the present application relates to fusion proteins and immunogenic compositions comprising Protein E and PilA, vaccines comprising such immunogenic compositions and therapeutic uses of the same.

No suitable figure