

USPA2 PROTEIN CONSTRUCTS AND USES THEREOF

FIELD OF THE INVENTION

[01] The present invention relates to compositions comprising *Moraxella catarrhalis* (*M. catarrhalis*, *M. cat.*) Ubiquitous surface protein A2 (UspA2). More particularly, the present application relates to UspA2 protein constructs and immunogenic compositions comprising the constructs, vaccines comprising such immunogenic compositions and therapeutic uses of the same.

BACKGROUND OF THE INVENTION

[02] Ubiquitous surface protein A2 (UspA2) is a trimeric autotransporter that appears as a lollipop-shared structure in electron micrographs (Hoiczky et al. EMBO J. 19: 5989-5999 (2000)). It is composed of a N-terminal head, followed by a stalk which ends by an amphipathic helix and a C-terminal membrane domain. (Hoiczky et al. EMBO J. 19: 5989-5999 (2000)). UspA2 contains a very well conserved domain (Aebi et al., Infection & Immunity 65(11) 4367-4377 (1997)), which is recognized by a monoclonal antibody that was shown protective upon passive transfer in a mouse *Moraxella catarrhalis* challenge model (Helminen et al. J Infect Dis. 170(4): 867-72 (1994)).

[03] UspA2 has been shown to interact with host structures and extracellular matrix proteins like fibronectin (Tan et al., J Infect Dis. 192(6): 1029-38 (2005)) and laminin (Tan et al., J Infect Dis. 194(4): 493-7 (2006)), suggesting it can play a role at an early stage of *Moraxella catarrhalis* infection.

[04] UspA2 also seems to be involved in the ability of *Moraxella catarrhalis* to resist the bactericidal activity of normal human serum. (Attia AS et al. Infect Immun 73(4): 2400-2410 (2005)). It (i) binds the complement inhibitor C4bp, enabling *Moraxella catarrhalis* to inhibit the classical complement system, (ii) prevents activation of the alternative complement pathway by absorbing C3 from serum and (iii) interferes with the terminal stages of the complement system, the Membrane Attack Complex (MAC), by binding the complement regulator protein vitronectin. (de Vries et al., Microbiol Mol Biol Rev. 73(3): 389-406 (2009)).

[05] *Moraxella catarrhalis* is an important and common respiratory pathogen that has been associated with increased risk of exacerbations in chronic obstructive pulmonary disease (COPD) in adults. (Sateesh et al., Journal of Chronic Obstructive Pulmonary Disease 3:109-115 (2006)).

[06] A need for vaccines for *Moraxella catarrhalis* exists.

BRIEF SUMMARY OF THE INVENTION

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

[08] $A - (R_1)_m - (B)_n$ (formula I)

[09] wherein:

A is UspA2 from *Moraxella catarrhalis* or an immunogenic fragment thereof;

R_1 is an amino acid;

m is 0, 1 or 2;

B is histidine; and

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

[10] As a second aspect, the present invention provides immunogenic compositions comprising proteins of formula (I) and proteins of the invention. 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 *Moraxella catarrhalis*. The method comprises administering to a subject in need thereof a therapeutically effective amount of a protein of formula (I) or a protein of the invention.

[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 a protein of formula (I) or a protein of the invention.

[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 a protein of formula (I) or a protein of the invention.

[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 a protein of formula (I) or a protein of the invention.

[15] In a seventh aspect, the present invention provides a pharmaceutical composition comprising a protein of formula (I) or a protein of the invention for use in the treatment or prevention of a condition or disease caused wholly or in part by *Moraxella catarrhalis*. 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] In a tenth aspect, the present invention provides a composition comprising at least one antigen from *Moraxella catarrhalis* and at least one antigen from *Haemophilus influenzae*. The composition may further comprise a pharmaceutically acceptable adjuvant. The composition may comprise an excipient.

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

[20] In a further 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 a protein of formula (I) or a protein of the invention and a therapeutically effective amount of at least one antigen from *Haemophilus influenzae*.

[21] The present invention also provides a pharmaceutical composition comprising a protein of formula (I) or a protein of the invention for use in the treatment or prevention of a condition or disease caused wholly or in part by *Moraxella catarrhalis* in combination with at least one antigen from *Haemophilus influenzae*. Pharmaceutical compositions may further comprise a pharmaceutically acceptable adjuvant.

[22] 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: A typical fermentation profile with the High Cell Density Induction (HCDI) processes and the parameters monitored during 20L-scale fed-batch fermentation.

Figure 2: A typical fermentation profile with the Low Cell Density Induction (LCDI) processes and the parameters monitored during 20L-scale fed-batch fermentation.

Figure 3: UspA2 yield from protein constructs MC-001, MC-002, MC-004, MC-005, MC-006, MC-007, MC-008 and MC-010 evaluated in fermenter; data from Table 4.

Figure 4: Molecular weight distribution of purified MC-005 determined by sedimentation velocity analytical ultracentrifugation. The majority of protein is found as a trimer, with a small proportion of a higher molecular weight oligomer that may correspond to dimer of trimer. MW = molecular weight. kDa = kilodalton.

Figure 5: Molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. Majority of protein is found as a trimer.

Figure 6: Molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. The sample presents multiple species and is highly polydisperse. The sedimentation coefficient of the major species detected doesn't correspond to the one of the trimers normally detected in the other lots.

Figure 7: Molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. Majority of protein is found as a trimer.

Figure 8: Molecular weight distribution of purified MC-007 determined by sedimentation velocity analytical ultracentrifugation. Majority of protein is found as a trimer.

Figure 9: Far-UV circular dichroism (CD) spectra of UspA2 constructs giving an indication of protein secondary structures.

Figure 10: Secondary structures monitoring by circular dichroism (CD) during thermal unfolding of MC-005 (UspA2 Δ helix + 6His). Visual analysis of the spectra clearly shows that protein loses most of its secondary structures at 33°C.

Figure 11: Secondary structures monitoring by circular dichroism (CD) during thermal unfolding of MC-007 (UspA2 full helix + 6His). Visual analysis of the spectra shows that loss of secondary structure is slower compared to the construct without helix. Structural changes are detectable upon heating to 33°C, but complete unfolding seems to occur between 35°C and 37°C.

Figure 12: MALDI spectrum of MC-001 lot opt-01. The mass observed at 57427Da may be coherent with the demethionylated protein, while the peak at 57620Da could correspond to the complete protein.

Figure 13: MALDI spectrum of MC-011 lot BMP37. The mass observed may be coherent with the demethionylated protein. The two other peaks at +186Da and +366Da are unidentified.

Figure 14: Protective efficacy of MC-001 and MC-007 in a mouse model of lung colonization.

Figure 15: Antibody response directed against UspA2 induced after intramuscular administration in mice, where PII and PIII indicate, respectively, anti-IgG levels in sera collected at day 28 (post II) and day 42 (post III).

Figure 16: Bactericidal titers induced by UspA2 against a homologous strain formulated with different adjuvants (AS01_E, AS04_C and AIPO₄).

Figure 17: Antibody response directed against UspA2 induced after intramuscular administration in mice, using different formulations of antigens and adjuvants.

Figure 18: Bactericidal titers induced by UspA2 against a homologous strain, using different formulations of antigens and adjuvants.

Figure 19: IgG response induced against PD in mice by PD-PEPiiA-UspA2 vaccine (a trivalent NTHi-M. cat. vaccine), formulated with different adjuvants.

Figure 20: IgG response induced against PE in mice by PD-PEPiiA-UspA2 vaccine (a trivalent NTHi-M. cat. vaccine), formulated with different adjuvants.

Figure 21: IgG response induced against PiiA in mice by PD-PEPiiA-UspA2 vaccine (a trivalent NTHi-M. cat. vaccine), formulated with different adjuvants.

Figure 22: Immunogenicity of PE in the bivalent PD-PEPiiA and trivalent PD-PEPiiA-UspA2 formulations with AS01E.

Figure 23: Immunogenicity of PiiA in the bivalent PD-PEPiiA and trivalent PE-PiiA-UspA2 formulations with AS01E.

Figure 24: Immunogenicity of PD in the bivalent PD-PEPiiA and trivalent PE-PiiA-UspA2 formulations with AS01E.

Figure 25: Effect of the tetravalent PD/ PEPiiA/ UspA2/ AS01_E vaccine formulation on mouse lungs pre-sensitized with heat inactivated *M. cat.* - Perivascularitis and peribronchiolitis in PBS immunized mice.

Figure 26: Effect of the tetravalent PD/ PEPiiA/ UspA2/ AS01_E vaccine formulation on mouse lungs pre-sensitized with heat inactivated *M. cat.* - Day 2 post-immunization.

Figure 27: Effect of the tetravalent PD/ PEPiiA/ UspA2/ AS01_E vaccine formulation on mouse lungs pre-sensitized with heat inactivated *M. cat.* - Day 7 post-immunization.

Figure 28: Effect of the tetravalent PD/ PEPiIA/ UspA2/ AS01_E vaccine formulation on mouse lungs pre-sensitized with heat inactivated *M. cat.* - Day 14 post-immunization.

Figure 29: Effect of the tetravalent PD/ PEPiIA/ UspA2/ AS01_E vaccine formulation on mouse lungs pre-sensitized with heat inactivated *M. cat.* – Detailed results

Figure 30: Post-vaccination lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IL17. Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 31: Post-vaccination lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing TNF α . Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 32: Post-vaccination lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IFN γ . Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 33: Post-vaccination lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IL13. Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 34: Post-challenge lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IL17. Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 35: Post-challenge lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing TNF α . Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 36: Post-challenge lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IFN γ . Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

Figure 37: Post-challenge lung CD4 T cell responses upon *M. cat.* WC re-stimulation. Lung CD4 cells expressing IL13. Restimulated with heat-inactivated *M. cat* whole cells (WC) or medium.

DETAILED DESCRIPTION OF THE INVENTION

[23] 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).

[24] 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.

[25] 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.

[26] 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."

[27] 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.

[28] 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).

[29] As used herein "UspA2" means Ubiquitous surface protein A2 from *Moraxella catarrhalis*. UspA2 may consist of or comprise the amino acid sequence of SEQ ID NO: 1 from ATCC 25238.

MKTMKLLPLKIAVTSAMIIGLGAASTANAQAKNDITLEDLPYLIKKIDQNELEADIGDIT
ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
AIKEDLQGLADFVEGQEGKILQNETSTKKNTQORNLVNGFEIEKNKDATAKNNESIEDLYD
FGHEVAESIGEIIHAHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
NIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTKVDGFGDSRVTAIDTK
VNAFDGRITALDSKVENGMAAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRV
NPNLAFKAGAAINTSGNKKGSYNIGVNYEF (SEQ ID NO: 1)

as well as sequences with at least or exactly 63%, 66%, 70%, 72%, 74%, 75%, 77%, 80%, 84%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity, over the entire length, to SEQ ID NO: 1. Comparison of 38 sequences of UspA2 from *Moraxella catarrhalis* (Table 1, SEQ ID NO: 1 – SEQ ID NO: 38) demonstrated approximately 63% to approximately 100% identity to UspA2 as set forth in SEQ ID NO. 1.

[30] UspA2 as described in SEQ ID NO: 1 contains a signal peptide (for example, amino acids 1 to 29 of SEQ ID NO: 1), a laminin binding domain (for example, amino acids 30 to 177 of SEQ ID NO: 1), a fibronectin binding domain (for example, amino acids 165 to 318 of SEQ ID NO: 1) (Tan et al. JID 192: 1029-38 (2005)), a C3 binding domain (for example, amino acids 30 to 539 of SEQ ID NO: 1 (WO2007/018463), or a fragment of amino acids 30 to 539 of SEQ ID NO: 1, for example, amino acids 165 to 318 of SEQ ID NO: 1 (Hallström T et al. J. Immunol. 186: 3120-3129 (2011)), an amphipathic helix (for example, amino acids 519 to 564 of SEQ ID NO: 1 or amino acids 520-559 of SEQ ID NO: 1, identified using different prediction methods) and a C terminal anchor domain (for example, amino acids 576 to 630 amino acids of SEQ ID NO: 1 (Brooks et al., Infection & Immunity, 76(11), 5330-5340 (2008)).

[31] UspA2 amino acid differences have been described for various *Moraxella catarrhalis* species. See for example, J Bacteriology 181(13):4026-34 (1999), Infection and Immunity 76(11):5330-40 (2008) and PLoS One 7(9):e45452 (2012).

[32] UspA2 may consist of or comprise an amino acid sequence that differs from SEQ ID NO. 1 at any one or more amino acid selected from the group consisting of: AA (amino acid) 30 to 298, AA 299 to 302, AA 303 to 333, AA 334 to 339, AA 349, AA 352 to 354, AA 368 to 403, AA 441, AA 451 to 471, AA 472, AA474 to 483, AA 487, AA 490, AA 493, AA 529, AA 532 or AA 543. UspA2 may consist of or comprise an amino acid sequence that differs from SEQ ID NO: 1 in that it contains at least one amino acid insertion in comparison to SEQ ID NO. 1. UspA2 may consists of or comprise an amino acid sequence that differs from SEQ ID NO. 1 at any one of the amino acid differences in SEQ ID NO: 2 through SEQ ID NO: 38. For example, SEQ ID NO. 1 may contain K instead of Q at amino acid 70, Q instead of G at amino acid 135 and/or D instead of N at amino acid 216.

[33] Table 1: UspA2 amino acid sequences from 38 strains of *Moraxella catarrhalis* (SEQ ID NO: 1 - SEQ ID NO: 38).

Strain	UspA2 sequences
ATCC 25238 (SEQ ID NO: 1)	MKTMKLLPLKIAVTSAMIIGLGAASTANAQAKNDITLEDLPYLIKKIDQNELEADIGDITALEK YLALSQYGNILALEELNKALEELDEDVGNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQG LADFVEGQREGKILQNETSIKKNTQRNLVNGFIEKNKDAIAKNNESIEDLYDFGHEVAESIGEI HAHNEAQNETLKGLITNSLENTNNITKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNI YELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQANIQLLATYNELQDQYAKQT

	EAIDALNKASSENTQNIEDLAAYNELQDAYAKQOTEATDALNKASSENTQNIEDLAAYNELQDAYAKQOTEATDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (630 aa)
American 2933 (SEQ ID NO: 2)	MKTMKLLPLKIAVTSAMLIIGLGAASTANAQSRDRSLEDIQDSISKLVQDDINTILKQDQQKMNKYLLLNQLANTLITDELNNNVIKNTNSIEALGDEIGWLENDIADLEEGVEELTKNQNTLIEKDEEHDRLIAQNQADIQTLENNVVEELFNLGRLIDQEADIAKNNASIEELYDFDNEVAERIGEIHAYTEEVNKTLENLITNSVKNTDNIDKNKADIDNNINHIYELAQQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKADLTKDIAKALESNVEEGLLDLSGRLLDQKADLTKDIAKALESNVEEGLLDLSGRLLDQKADIAQNQTDIQDLAAYNELQDQYAKQOTEATDALNKASSENTQNIEDLAAYNELQDAYAKQOTEATDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTNRIATAELGIAENKKDAQIAKAQANANKTAIDENKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDCRVTAIDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (613 aa)
American 2912 (SEQ ID NO: 3)	MKTMKLLPLKIAVTSALIIIGLGAASTANAQQQLQTETFLPNFLSNDNYDLTDPFYHNMILGDTALLDKQDGSQPQLKFYSNDKDSVPDSLLFSKLLHEQQQLNGFKKGDTIIPLDKDGKPVYQVDYKLDGKGGKQKRRQVYSVTTKTATDDVDNSAYSREILGKVDDLDDEMNFNLHNDITSLYDVTANQQDAIKDLKKGKVLNKLKELDKVGVLSRDIGSLNDDVAQNNESIEDLYDFSQEVADSIGEIHANNAQNETLQDLITNSVENTNNITKNKADIOALENNVVEELFNLGRLIDQKADLTKDICTLESNVEEGLLELSGHLIDQKADIAKNQADIAQNQANIQDLAAYNELQDAYAKQOTEATDALNKASSENTQNIEDLAAYNELQDAYAKQOTEATDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDENKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (644 aa)
American 2908 (SEQ ID NO: 4)	MKTMKLLPLKIAVTSALIVGLGAASTANAQLVERFFPNIFLDKPLAKQHYHNVVVGDTISIVSDLQSNSDQLKFYSDDDEGLVPDSLLFNKMLHEQLLNGFKREGDTIIPLDENGKPVYKVDYKLDGKEPRKVYSVTTKIATAEDVATSSYANGIQKDIDDLDFDHQVTERLTQHGKTIYRNGERILANEEVSQYLNKEVQNNIEHIYELAQQQDQHSSDIKTLKLESNVEKGLLELSGHLIDQKADLTKDICTLESNVEEGLLDLSGRLLDQKADLTKDICTLESNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAYNELQDQYAKQOTEATDALNKASSENTQNIEDLAAYNELQDAYAKQOTEATDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTNRIATAELGIAENKKDAQIAKAQANANKTAIDENKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTS

	GNKKGSYNIGVNYEF (591 aa)
Finnish 307 (SEQ ID NO: 5)	MKTMKLLPLKIAVTSAMIIGLGAASTANAQQQQQQQQQQSRTEIFFFNIFFNENHDELDDAYH NIILGDTALLDKQDGSQPQLKFYSNDKDSVPDSLLFSKLLHEQQQLNGFKKGDTIIFLDKDGKPV YQVDYKLDGKGGKKQKRRQVYSVTTTKTATDDDVNSAYS SRGILGKVDDLDDEMNFNLNHDITSLYDV TANQQDAIKGLKKGVKGLNKEKELKEVGVLSRDIGSLNDDVAQNNESIEDLYDFSQEVADSI GEIHAHNKAQNETLQDLITNSVENTNNITKNKADIQALENNVVEELFNLSGRLLDQKADLTKDI KTLESNVEEGLLELSGHLIDQKADIAKNQADIAQNQANIQDLAAYNELQDAYAKQQTEAIDALN KASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTE AIDALNKASSENTQNIKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKAD ADASFETLTKNQNTLIEKDKEHDKLITANKTAIDENKASADTKFAATADAITKNGNAITKNAKS ITDLGTKVDAEDGRVTALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQFPYSVGKFNATAAL GGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (697 aa)
Finnish 353 (SEQ ID NO: 6)	MKTMKLLPLKIAVTSAMIVGLGMASTANAQQQKSPKTETFLPNIFFNEYADDLDTLYHNMIIGD TAITHDDQYKFYADDATEVPDSLFFNKILHUQLLYGFKEGDKIIFLDENGKPVYKLDKRLENG QKTVYSVTTTKTATADDVNSAYS SRGIQGGIDDLYEANKENVNRLIEHGDKIFANEESVQYLNREV QNNIENIHELAAQQDQHSSDIKTLKKNVEKDLLDLSGRLIAQKEDIAQNQTDIQDLATYNELOD QYAKQKQTEAIDALNKASSENTQNIKNNGNHITLLENNIEEGLLELSGHLIDQKADLTKDIAKALE SNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI DLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSE NTQNIKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKN QNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTKVDGF DGRVTALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQFPYSVGKFNATAALGGYG SKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (683 amino acids)
Finnish 358 (SEQ ID NO: 7)	MKTMKLLPLKIAVTSAMMVGLGMASTANAQQQKSPKTEIFLPNLFNDNDNTELTDPPLYHNMIILGN TALLTQENQYKFYADDGNGVDPDSLLFNKILHUQLLHGFKEGGTIIFLDENGKPVYKLDSEIVEQG KTKTVYSVTTTKTATADDVNSAYS SRGIQGGIDDLYEANKENVNRLIEHGDKIFANEESVQYLNRE VQNNIENIHELAAQQDQHSSDIKTLKKNVEKDLLDLSGRLIAQKEDIAQNQTDIQDLATYNELOD DQYAKQKQTEAIDALNKASSENTQNIKNNSNHITLLENNIEEGLLELSGHLIDQKADLTKDIAKAL ESNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI EDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASS ENTQNIKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKN NQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSTFDLGTKVDG FDGRVTALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQFPYSVGKFNATAALGGY GSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (684 amino acids)

Finnish 216 (SEQ ID NO: 8)	MKTMKLLPLKIAVTSAMITGLGAASTANAQQQQKTKTEVFLPNLFDNDYYDLTDPLYHSMILGD TATLFDQQDNSKSQLKFYSNDKDSVPDSLLFSKLLHEQQQLNGFKAGDTIIPLDKDGKPVYTDQDT RTKDGKGVETVYSVTTKIATQDDVEQSAYSIRGIQGDIDDLIDYNREVNEYLKATIDYNERQTEAI DALNKASSANTDRIDTAERIDKNEYDIKALESNVGKDLLDLSGRLIAQKEDIDNNINHIYELA QQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKADLTCKDIKTLENNIEEGLLELSGHLIDQKADL TKDIKTLENNIEEGLLELSGHLIDQKADIAQNQANIQDLAAYNELQDQYAKQTEAIDALNKAS SENTQNIEDLAAYNELQDAYAKQOTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQOTEAID ALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKVSAANTDRIAKNKADADA SFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITD LGTKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQFPYSVGKFNATAALGGY GSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (684 amino acids)
Dutch H2 (SEQ ID NO: 9)	MKTMKLLPLKIAVTSAMMVGLGMASTANAQQQKSPKTEIFLPLNLFNDNDNTELTDPHYHNMILGN TALLTQENQYKFYADDGNGVDPDSLLFNKILHDQQLLHGFKKGDTIIPLDENGKPVYKLDSDIVEQG KTKTVYSVTTKTATADDVNSAYSIRGIQGDIDDLIEANKENVNRLIEHGDKIFANEESVQYLNRE VQNNIENIYELVQQQDQHSSDIKTLKNNVEKDLLDLSGRLIAQKEDIAQNQNTDIQDLATYNELQ DQYAKQOTEAIDALNKASSENTQNIAKNSNHKTLLENNIEEGLLELSGHLIDQKADLTCKDIKAL ESNVEEGLLDLSGRLIDQKADIAQNQANIQDLAAYNELQDAYAKQOTEAIDALNKASSENTQNI EDLAAYNELQDAYAKQOTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQOTEAIDALNKASS ENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLT KNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGKTKVDG FDGRVTALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQFPYSVGKFNATAALGGY GSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (684 amino acids)
Dutch F10 (SEQ ID NO: 10)	MKTMKLLPLKIAVTSAMITGLGAASTANAQLAEQFFPNIFSNHAPVKQHYHNVVVGDTSTIVENL QDSDDTQLKFYSNDEYSVPDSLLFNKMLHEQQQLNGFKKGDTIIPLDENGKPVYKVDYKLDGQEP RRVYSVTTKIATQDDVDNSPYSIRGIQGDIDDLIEANKENVNRLIEHGDKIFANEESVQYLNKEV QNNIENIYELAQQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKADLTCKDIKTLESNVEEGLLEL SGHLIDQKADIAKNQADIAQNQANIQDLAAYNELQDAYAKQOTEAIDALNKASSENTQNIEDLA AYNELQDAYAKQOTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAK ASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATAD AITKNGNAITKNAKSITDLGKTKVDAFDGRVTALDTKVNAFDGRITALDSKVENGMAAQAALSGL FQFPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (574 amino acids)
Norwegian 1	MKTMKLLPLKIAVTSALIVGLGAASTANAQQQPQTFETFFPNIFFNENHDALDDVYHNMILGDTA ITQDNQYKFYADAISEVPDSLLFNKILHDQQLNGFKEGDTIIPLDENGKPVYKLDEKVENGVKK

(SEQ ID NO: 11)	SVYSVTTKTATRADVEQSAYSRSIGQDIDDLYEANKENVNRLIEHGDKIFANEESVQYLNKEVQ NNIENIHELAAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKADLTCKDIKTLNESNVEEGLLDLS GRLLDQKADIAQNQANIQDLAAYNELQDAYAKQQTETAI DALNKASSENTQNI EDLAAAYNELQDA YAKQQTETAI DALNKASSENTQNI EDLAAAYNELQDAYAKQQTETAI DALNKASSENTQNI EDLAAAY NELQDAYAKQQTETAI DALNKASSENTQNI EDLAAAYNELQDAYAKQQTETAI DALNKASSENTQNI AKNQADIANNINNIYELAAQQQDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQNTLI EKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVDAFDGRVT ALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAV AIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (678 amino acids)
Norwegian 13 (SEQ ID NO: 12)	MKTMKLLPLKIAVTSAMIVGLGAASTANAQQQQQPRTEFFFPNIFNENHDALDDVYHNMI LGD TAITQDNQYKFYADAI SEVPDSLLFNKILHDQQINGFKEGDTIIFLDENGKPVYKLDEKVENG KKS SVYSVTTKTATRADVEQSAYSRSIGQDIDDLYEANKENVNRLIEHGDKIFANEESVQYLNRE VQNNIENIHELAAQQQDQHSSDIKTLKKNVEKDLLDL SGRLIAQKEDIAQNQTDI QDLATYNELQ DQYAKQQTETAI DALNKASSENTQNI AKNSNHIKTL ENNIEEGLLELSGHLIDQKADLTCKDIKTL ENNIEEGLLELSGHLIDQKADLTCKDIKALESNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAY NELQDQYAKQQTETAI DALNKASSENTQNI EDLAAAYNELQDAYAKQQTETAI DALNKASSENTQNI AKNQADIANNINNIYELAAQQQDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQNTLI EKDKEHDKLITANKTAIDTNKASADTKFAATADAITKNGNAITKNAKSITDLGTVDGFDGRVT ALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAV AIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (678 amino acids)
Norwegian 20 (SEQ ID NO: 13)	MKTMKLLPLKIAVTSALIVGLGAASTANAQLVERFFFPNIFLDKPLAKQHYHNVVVGDTSI VSDI QSNSDQLKFYSDDGLVPDELLFNKMLHEQLLNGFKEGDTIIFLDENGKPVYKVDYKLDGKEPR KVYSVTTKIATAEDVATSSYANGIQKDIIDLDYDFUHQVTERLTQHGKTIYRNGERTLANEESVQ YLNKEVQNNIEHIYELAAQQQDQHSSDIKTLNESNVEKGLLELSGHLIDQKADLTCKDIKTL ENNVE EGLLDLSGRLLIDQKADIAQNQANIQDLAAYNELQDQYAKQQTETAI DALNKASSENTQNI EDLAA YNELQDAYAKQQTETAI DALNKASSENTQNI AKNQADIANNINNIYELAAQQQDQHSSDIKTLAKA SAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDTNKASADTKFAATADA ITKNGNAITKNAKSITDLGTVKDGFDGRVTALDTKVNALDTKVNAFDGRITALDSKV ENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKK GSYNIGVNYEF (587 amino acids)
Norwegian 25 (SEQ ID NO: 14)	MKTMKLLPLKIAVTSAMIVGLGAASTANAQQQQQPRTEFFFPNIFNENHDALDDVYHNMI LGD TAITQDNQYKFYADAI SEVPDSLLFNKILHDQQINGFKEGDTIIFLDENGKPVYKLDEKVENG KKS SVYSVTTKTATRADVEQSAYSRSIGQDIDDLYEANKENVNRLIEHGDKIFANEESVQYLNRE VQNNIENIHELAAQQQDQHSSDIKTLKKNVEKDLLDL SGRLIAQKEDIAQNQTDI QDLATYNELQ DQYAKQQTETAI DALNKASSENTQNI AKNSNHIKTL ENNIEEGLLELSGHLIDQKADLTCKDIKTL

	ENNIEEGLLELSGHLIDQKADLT KD IKALESNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAY NELQDQYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNI AKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIE EKDKEHDKLITANKTAIDTNKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGDFGRVT ALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAV AIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (678 amino acids)
Norwegian 27 (SEQ ID NO: 15)	MKTMKLLPLKIAVTSALIVGLGAASTANAQVRDKSLEDIEALLGKIDISKLEKEKKQQTTELQKY LLLSQYANVLTMEELNKNVEKNTNSIEALGYEIGWLENDIADLEEGVEELTKNQNTLIEKDEEH DRLIAQNQADIKTLENNVVEELFNLSDRLLDQFADIAKNNASIEELYDFDNEVAERIGEIHAYT EEVNKTLEKLITNSVKNNTDNIDKNKADIQALENNVEEGLLELSGHLIDQKADLT KD IKALESNV EEGLLDLSGRLLDQKADIAKNQADIAQNQTDIQLAAYNELQDQYAKQKQTEAIDALNKASSENT QNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNK ASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKVSAANTDRIAKNKADADASFET LTKNQNTLIEKDEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVK VDGFDSRV TALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKS AVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (616 amino acids)
Norwegian 36 (SEQ ID NO: 16)	MKTMKLLPLKIAVTSALIVGLGAASTANAQATETFTLPNLFNDNDYTETTDPLYHGMILGNTAITQ DTQYKFYAENGNEVPDSLFFNKILHDDQQLNGFKEGDTIIPLDENGKPVYKLEITENGVKRKVY SVTTKTATREDVEQSAYSRTGQGDIDDLYEANKENVNRLIEHGDKIFANEESVQYLNKEVQNNI ENIHELAAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKADLT KD IKALESNVEEGLLDLSGHL IDQKADLT KD IKALESNVEEGLLDLSGRLLDQKADIAKNQADIAQNQTDIQLAAYNELQDQY QKQTEAIDALNKASSENTQNIEDLAAYNELQDQYAKQKQTEAIDALNKASSENTQNIEDLAAYNE LQDQYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDQYAKQKQTEAIDALNKASSENTQNI AKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEK DKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGDFGRVTAL DTKVNALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAI GAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (676 amino acids)
BC5SV (SEQ ID NO: 17)	MKTMKLLPLKIAVTSALIVGLGAASTANAQNGTSTKLKLNKEYAQYLDNYAQYLDLDDIDDL DKEVGELSQNIAKNQANIKDLNKKLSRDIDSLREDVYDNQYEIVNNQADIEKNQDDIKELE NNVGKELLNLSGRLLDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLI DQKSDIAQNQTDIQLLATYNELQDQYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAY AKQKQTEAIDALNKASSENTQNIQLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLA AYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSE NTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQ

	DQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAID ANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKDAFDGRVTALDTKVNAFDGRITA LDSKVENGMAAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAA INTSGNKKGSSYNIGVNYEF (629 amino acids)
Norwegian 14 (SEQ ID NO: 18)	MKTMKLLPLKIAVTSAMIVGLGMASTANAQQQSRSPKTETFLPNIFFNEYADDLDTLYHNMI LGDTAITHDDQYKFYADDATEVPSLFFNKILHDQQLLYGFKEGDKIIFLDENGKPVYKLDK RLDNGVQKTVYSVTTKTATADDVNSAYSRGIQGDIDDLYEANKENVNRLIEHGDKIFANEE SVQYLNKEVQNNIENIHELAQQQDQHSSDIKTLKKNVEEGGLELSGHLIDQKTDIAQNQTD IQDLATYNELQDQYAKQTEAIDALNKASSENTQNIAKNSNRIKALENNIEEGGLELSGHL IDQKADLT KD IKALESNVEEGLLDLSGRLLDQKADIAQNQANIQDLAAYNELQDAYAKQQT EAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNEL QDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKAS AANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATA DAITKNGNAITKNAKSITDLGTVKDGFDGRVTALDTKVNALDTKVNAFDGRITALDSKVEN GMAAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNK KSSYNIGVNYEF (683 amino acids)
Norwegian 3 (SEQ ID NO: 19)	MKTMKLLPLKIAVTSAMIVGLGAASTANAQAQSNRSLDQVQALLRGIDETRIKKEIQQSQQ PELNKYLTFNQLANALNIEELNNNVQKNTQRLDSAATLYGDLSTVPKSIKENKESIKENK ESIKENKESIKENKESIKENKESIKENKESITTLTRKSFQNOVDIVRNNASTIEDLYAYGQE VAKSIGEIHAYTEEVNKTLENLITNSVENTNNTKNKADIQALENNVVEELENLSGRLLDQ KADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGGLELSGHLIDQKADLT KD IKTLESNV EEGLLDLSGRLLDQKADIAQNQANIQDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI DLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKA SSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELA QQQDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKT VIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKDGFDGRVTALDTKVNALDTK VNAFDGRITALDSKVENGMAAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVN PNLAFKAGAAINTSGNKKGSSYNIGVNYEF (700 amino acids)
Finnish 414 (SEQ ID NO: 20)	MKTMKLLPLKIAVTSALIVGLGAASTANAQATETFLPNLFDNDYIETTDPLYHGMITLGNTA ITQDTQYKFYAENGNEVPDSLFFNKILHDQQLNGFKEGDTIIFLDENGKPVYKLDEITENG VKKRVYSVTTKTATREDVEQSAYSRGIQGDIIDDLYEANKENVNRLIEHGDKIFANEESVQY LNKEVQNNIENIHELAQQQDQHSSDIKTLKKNVEEGGLELSGHLIDQKADLT KD IKTLENN VEEGGLELSGHLIDQKADLT KD IKALESNVEEGLLDLSGRLLDQKADIAKNQADIAQNQTD

	IQDLAAYNELQDQYAKQKTEAIDALNKASSENTQNIEDLAAYNELQDQYAKQKTEAIDALN KASSENTQNIEDLAAYNELQDQYAKQKTEAIDALNKASSENTQNIEDLAAYNELQDQYAKQ QTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRI AKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNG NAITKNAKSITDLGTVKVDGFDGRVTALDTKVNALDTKVNAFDGRITALDSKVENGMAAQAA LSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIG VNYEF (676 amino acids)
Japanese Z7476 (SEQ ID NO: 21)	MKTMKLLPLKIAVTSAMTIGLGAASTANAQLAEQFFPNIFSNHAPVKQHYHNVVVGDTISIV ENLQDSDDTQLKFYSNDEYSVPDSLLENKMLHEQQQLNGFKKGPDTITPLDENGKPVYKVDYK LDGQEPRRVYSVTTKIATQDDVDNSPYSRGIQGDIDDLYEANKENVNRLLIEHGDKRIFANEE SVQYLNKEVQNNIENIYELAQQQDQHSSDIKTLKKNVEEGLLELSGRLIDQKADIAQNQAN IQDLAAYNELQDQYAKQKTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALN KASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQ QTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYN ELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAK VSAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAA TADAITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMAAQ AALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYN IGVNYEF (678 amino acids)
Belgian Z7530 (SEQ ID NO: 22)	MKTMKLLPLKIAVTSAMIIGLGAASTANAQSRDRSLEDIQDSISKLVQDDINTLKQDQQKM NKYLILNQLANTLITDELNNNVIKNTNSTEALGDEIGWLENDIADLEEVEELTKNQNTLI EKDEEHDRLIAQNQADIQTLENNVVEELFNLSGRLIDQEADIAKNNASIEELYDFDNEVAE RIGEIHAYTEEVNKTLENLITNSVKNTDNIDKNKADIDNNINHIYELAQQQDQHSSDIKTL KNNVEEGLLELSGHLIDQKADLTDKIKALESNVEEGLLDLSGRLLDQKADLTDKIKALESN VEEGLLDLSGRLLDQKADIAQNQTDIQDLAAYNELQDQYAKQKTEAIDALNKASSENTQNI EDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSS DIKTLAKASAANTNRIATAELGIAENKKDAQIAKAQANANKTAIDENKASADTKFAATADA ITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMAAQAALS GLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVN YEF (613 amino acids)
German Z8063	MKTMKLLPLKIAVTSALIVGLGAASTANAQATNKDITLEDVLKSIEETDPYELRDYIEYPT AIERFLLLSQYGNLTLEEFNDNIELLDQDVEDLEESVTELAQNQNSLIEQGEAIREDLQG LADFVERQEDKILQNETSIRKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAKSI

(SEQ ID NO: 23)	GEIHAHNEAQNETLKDITNSVKNTDNITKNKADIQALESNVEKGLELSGHLIDQKADID NNINNIHELAAQQQDQHSSDIKTLKKNVEEGLELSGHLIDQKSDIAQNQANIQDLATYNEL QDQYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNI AKNQADIANNINNIYELAAQQQDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQN TLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDG FDSRVTAALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKS AVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (589 amino acids)
American Q12E (SEQ ID NO: 24)	MKTMKLLPLKIAVTSAMMVGLGMASTANAQQQKSPKTEIFLPLNLFNDNTELTDFLYHNMI LGNTALLTQENQYKFYADDGNGVDPDSLLFNKILHDQLLHGFKEGDTIIPLDENGKFPVYKLD SIVEQGGKTKTVYSVTTKTATADDVNSAYSARGIQGDIDDLYEANKENVNRLIEHGDKIFANE ESVQYLNREVQNNIENIHELAAQQQDQHSSDIKTLKKNVEKDLLDLSGRLIAQKEDIAQNQT DIQDLATYNELQDQYAKQKQTEAIDALNKASSENTQNIAKNSNHIKTLNENIIEGLELSGH LIDQKADLTRDIKALESNVEEGLLDLSGRLLIDQKADIAQNQANIQDLAAYNELQDAYAKQQ TEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNE LQDAYAKQKQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAAQQQDQHSSDIKTLAKA SAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAAT ADAITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVE NGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGN KKGSYNIGVNYEF (684 amino acids)
Greek MC317 (SEQ ID NO: 25)	MKTMKLLPLKIAVTSALIVGLGAASTANAQQQKTKTEVFLPLNLFYNDYIEETDLLYHNMI LGDFAALVDRQNYNSQLKFYSNDEESVFPDSLLFSKMLNNQQLNGFKAGDIIIPVDANGQV IYQKDTVRVEGGKTRTVLSVTTKIATQQDVDSAYSARGIQGKVNDLDEMNFLNHDITSLYDV TANQQDDIKGLKKGVKDLKKGVEGLNKELEKELDKVEGVLSRDIKSLNDDVAQNNESIEDLY DFSQEVADSIGEIHAKHKAQNETLQDLITNSVENTNNTKNKADIQALENNVVEELFNLSG RLIDQKADLTRDIKTLESNVEEGLLELSGHLIDQKADIAKNQADIAQNQANIQDLAAYNEL QDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNI AKNQADIANNINNIYELAAQQQDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQN TLIEKDKEHDKLITANKTAIDENKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDG FDGRVTALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKS AVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (650 amino acids)
American V1122	MKTMKLLPLKIAVTSALIVGLGAVSTTNAQAQSRSLDQIQTKLADLAGKIAAGKNGGGQNN QNNQNDINKYFLFSQYANILTMEELNNNVVKNSSSIETLETDFGWLENDVADLEDGVEELT KNQNTLIEKDEEHDRLIAQNQADIQTLENNVVEELFNLSDRLIDQKADIAKNQADIAQNNE

(SEQ ID NO: 26)	SIEELYDFDNEVAEKIGIEIHAYTEEVNKTLDLITNSVRNTDNIDKNKADIDNNINHIYEL AQQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKADLTCDIKTLENNVEEGLLDLSGRLIDQ KADIAKNQADIAQNQTDIQLDAAAYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNE LQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKA SAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDENKASADTKFAAT ADAITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMAAQA ALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNI GVNYEF (616 amino acids)
American P44 (SEQ ID NO: 27)	MKTMKLLPLKIAVTSALIVGLGTASTANAQVASEPANQKIQQKIKKVRKELRQDIKSLRNDI DSNTADIGSLNDDVADNQDDILDNQADIAKNQDDIEKNQADIKELDKEVGVLSREIGSLND DIADNYTDIIDNYTDIIDNQANIAKNQDDIEKNQADIKELDKEVGVLSREIGSLNDDVADN QDDIAKNQADIQTLENNVEEGLLELSGHLIDQKADIDNNINNIYELAQQQDQHSSDIKTLK KNVEEGLLELSGHLIDQKTDIAQNQANIQLDQATYNELODQYAEQTEAIDALNKASSENTQ NIAKNSNRIKALESNVEEGLLELSGHLIDQKADLTCDIKALESNVEEGLLELSGHLIDQKA DIAQNQANIQLDAAAYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQ TEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAAANTDRIA KNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKVSADTKFAATADAITKNGN AITKNAKSITDLGTVKVDADFSDRVTAALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPY SVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNI GVNYEF (668 amino acids)
American V1171 (SEQ ID NO: 28)	MKTMKLLPLKIAVTSAMIVGLGATSTVNAQVVEQFFFNIFFNENHDELD DAYHNMILGDTA IVSNSQDNSTQLKFYSNDEDSVPDSLLFSKLLHEQQLNGFKAGDTIIPLDKDGKPVYTKDT RTRKDGKVVETVYSVTFKRIATQDDVEQSAYSRSIGQDIDDLYDINREVNEYLKATHDYNERQT EAIDALNKASSANTDRIDTAERIDKNEYDIKALESNVEEGLLELSGHLIDQKADLTCDIK ALESNVEEGLLELSGHLIDQKADLTCDIKALESNVEEGLLDLSGRLIDQKADIAQNQANIQ DLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNK ASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQT EAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAAANTDRIAK NKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNA ITKNAKSITDLGTVKVDGFDGRVTALDTKVNAALDTKVNAFDGRITALDSKVENGMAAQAALS GLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNI GVNYEF (674 amino acids)
American	MKTMKLLPLKIAVTSAMIIGLGAASTANAQSRDRSLEDIQDSISKLVQDDIDTLKQDQQKM

TTA24 (SEQ ID NO: 29)	NKYLLLNQLANTLITDELNNNVIKNTNSIEALGDEIGWLENDIADLEEGVEELTKNQNTLI EKDEEHDRLIAQNQADIQTLENNVVEELFNLSGRLLDQEAADIKNNASIEELYDFDNEVAE RIGEIHAYTEEVNKTLENLITNSVKNTDNIDKNKADIDNNINHIYELAQQQDQHSSDIKTL KNNVEEGLLELSGHLIDQKADLTCDIKALESNVEEGLLDLSGRLLDQKADLTCDIKALESN VEEGLLDLSGRLLDQKADIAQNQTDIQLAAYNELQDQYAKQQTAEIDALNKASSENTQNI EDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIARNQADIANNINNIYELAQQQDQHSS DIKTLAKASAAANTNRIATAELGIAENKKDAQIAKAQANANKTAIDENKASADTKFAATADA ITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMQAQAALS GLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVN YEF (613 amino acids)
American O35E (SEQ ID NO: 30)	MKTMKLLPLKIAVTSAMIVGLGATSTVNAQVVEQFFFNIFENENHDELDDAYHNMILGDTA IVSNSQDNSTQLKFYSNDEDSVPDSLLFSKLLHEQQNLNGFKAGDTIIFLDKDGKPVYTKDT RTKDGKVVETVYSVTTKIATQDDVEQSAYSIRGIQGDIDDLIDINREVNEYLKATHDYNERQT EAIDALNKASSANTDRIDTAERIDKNEYDIKALESNVEEGLLELSGHLIDQKADLTCDIK ALESNVEEGLLELSGHLIDQKADLTCDIKALESNVEEGLLDLSGRLLDQKADIAKNQADIA QNQTDIQLAAYNELQDAYAKQQTEAIDALNKASSENTQNIARNQADIANNINNIYELAQQ QDQHSSDIKTLAKASAAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAI DANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNALDTKVN AFDGRITALDSKVENGMQAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPN LAFKAGAAINTSGNKKGSYNIGVNYEF (576 amino acids)
American SP12-6 (SEQ ID NO: 31)	MKTMKLLPLKIAVTSAMMVGLGMASTANAQQQKSPKTEIFLPNLFNDNTELTDPYHNMIL LGNTALLTQENQYKFYADGNGVDPDSLLFNKILHDQLLHGFKEGDTIIFLDENGKPVYKLD SIVEQGGTKTVYSVTTKTATADDVNSAYSIRGIQGDIDDLIEANKENVNRLIEHGDKIFANE ESVQYLNREVQNNIENIHELAAQQQDQHSSDIKTLKKNVEKDLLDLSGRLLAQKEDIAQNQT DIQDLATYNELQDQYAKQQTAEIDALNKASSENTQNIARNQADIANNINNIYELAQQQDQHSS LIDQKADLTCDIKALESNVEEGLLDLSGRLLDQKADIAQNQANIQLAAYNELQDAYAKQQ TEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNE LQDAYAKQQTEAIDALNKASSENTQNIARNQADIANNINNIYELAQQQDQHSSDIKTLAKA SAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAAT ADAITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNALDTKVNAFDGRITALDSKVE NGMAQAQAALSGLFQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGN KKGSYNIGVNYEF (684 amino acids)
American	MKTMKLLPLKIAVTSAMIIGLGAASTANAQATETFLPNLFNDNDYTETTDPLYHGMILCNTA

SP12-5 (SEQ ID NO: 32)	ITQDTQYKFYAENGNEVPDSLFFNKILHDQQLNGFKEGDTIIPLDENGKFPVYKLDEITENG VKKRVYSVPTKTATREDVEQSAYSRSIGQGDIDDLYEANKENVNRLIEHGDKIFANEESVQY LNKEVQNNIENIHELAAQQDQHQSSDIKTLKKNVEEGLLELSGRLLIAQKEDIAQNQTDIQLD ATYNELQDQYAKQKQTEAIDALNKASSENTQNIAKNSNHIKTLNENIEEGLLELSGHLIDQK ADLTRDIKALESNVEEGLLDLSGRLLDQKADIAKNQADIAQNQTDIQLDLAAYNELQDQYAK KQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAY NELQDAYAKQKQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAAQQDQHQSSDIKTLA KASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFA ATADAITKNGNAITRNAKSITDLGTVKVDGDFGRVTALDTKVNAFDGRITALDSK VENGMAAQAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTS GNKKGSYNIGVNYEF (686 amino acids)
Swedish BC5 (SEQ ID NO: 33)	MKTMKLLPLKIAVTSAMITGLGAASTANAQAKNDITLEDLPYLIKIDQNELEADIGDITALEK YLALSOYGNILALEELNKALEELEDVGVNQNNDIANLEDDVETLTKNQNALAEQGEAIKEDLQG LADPVEGQEGKILQNETSIKKNQNTQNLVNGFEIEKNKDAIAKNMESIEDLYDFGHEVAESIGE HAHNEAQNETLKGLITNSIENTNNITKNKADIALENNVVEELFNLSGRLLIDQKADIDNNINNI YELAAQQDQHQSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQANIQLLATYNELQDQYAKQK EADALNKASSENTQNIEDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNELQDA YAKQKQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAAQQDQHQSSDIKTLAKASAANTDR IAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNA ITRNAKSITDLGTVKVDGDFSRVTALDTKVNAFDGRITALDSKVENGMAAQAALSGLFQPYSVGK FNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (630 amino acids)
American 7169 (SEQ ID NO: 34)	MKTMKLLPLKIAVTSALIVGLGAASTANAQAQDRSLEQIQDKLANIVEKIEQAKSQNGQSQ KDINQYLLLSQYANVLTMEELNNNVKNSSSIETLDNDIAWLNDDITDLDKEVGVLSRDIG SLHDDVAQNQADIKTLKNNVVEELFNLSDRLLDQEAQNNESIEDLYDFGREVAESIGE IHAHNEAQNETLKDLITNSVKNQNTNITKNKADIALENDVGVKELNLSGRLLIDQKADIDNN INHIYELAAQQDQHQSSDIKTLKNNVEEGLLELSGHLIDQKADLTRDIKALESNVEEGLLDL SGRLLDQKADIAQNQANIQLDLAAYNELQDAYAKQKQTEAIDALNKASSENTQNIEDLAAYNE LQDAYAKQKQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAAQQDQHQSSDIKTLAKA SAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAAT ADAITKNGNAITRNAKSITDLGTVKVDGDFSRVTALDTKVNAFDGRITALDSKVENGMAAQA ALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNI GVNYEF (616 amino acids)
Finnish	MKTMKLLPLKIAVTSAMITGLGATSTVNAQVVEQFFENIFFNENHDELDDAYHNMILGDTA

FIN2344 (SEQ ID NO: 35)	IVSNSQDNSTQLKFYSNDEDSVPDSLLFSKLLHEQQLNGFKAGDTIIPLDKDGKPVYTKDT RTKDGKVEFVYSVTTKIATQDDVEQSAYSIRGIQGGIDDDLYDINREVNEYLKATHDYNERQT EAIIDALNKASSANTDRIDTAEERIDKNEYDIKALESNVGKDLLDLSGRLIAQKEDIDNNIN HIYELAQQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKADLTCKDIKTLESNVEEGLLDLSG RLIDQKADIAQNQANIQDLAAYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYNELQ DAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLARVSA ANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATAD AITKNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMMAQAAL SGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGV NYEF (614 amino acids)
American V1118 (SEQ ID NO: 36)	MKTMKLPPLKIAVTSAMIIGLGAASTANAQTTETFLPNLFDNDYTETTDPLYHGMILGDTA ITQDTQYKFAENGNEVPDSLFFNKILHDQLLNGFKAGDTIIPLDENGKPVYKLDERTENG VKRKVYSVTTKTATQADVEQSAYSIRGIQGGIDDDLYEANKENVNRLIEHGDKIFANEESVQY LNREVQNNIENIHELAAQQDQHSSDIKTLKNNVEKDLLDLSGRLIAQKEDIAQNQTDIQL ATYNELQDQYAQKQTEAIDALNKASSENTQNIAKNSNHIKTLENNIERCLLELSGHLIDQK ADLTCKDIKALESNVEEGLLDLSGRLIDQKADIAQNQANIQDLAAYNELQDAYAKQQTEAID ALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAY AKQQTEAIDALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANT DRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAIT KNGNAITKNAKSITDLGTVKVDGFDGRVTALDTKVNALDTKVNAFDGRITALDSKVENGMMA QAALSGLFQPYSVGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSY NIGVNYEF (679 amino acids)
American V1145 (SEQ ID NO: 37)	MKTMKLLPLKIAVTSALIVGLGAASTANAQETLEEVLESIKQINEQDLQDDIGYNSAJDRY LVLSQYGNLITAKELNENVEKNSNSIAKNSNSIADLEADVGYLAENQNTLIEQNETINQEL EGITHELESFIAYAHQQDQKNLVNEFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEITHA YTREVNKTLENLITNSVENTDNITENKADIQALESNVEKELLNLSGRLIDQKADIDNNINH IYELAQQQDQHSSDIKTLKNNVEEGLLELSGHLIDQKSDIAQNQTDIQLATYNELQDQYA QKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAA YNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSEN TQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAID ALNKASSENTQNIAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKAD ADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKN AKSITDLGTVKVDGFDGRVTALDTKVNAFDGRITALDSKVENGMMAQAALSGLFQPYSVGKF NATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYEF (724

	amino acids)
American V1156 (SEQ ID NO: 38)	MKTMKLLPLKIAVTSALIVGLGAASTANAQAQARDRSLEDIQALIGNIDVDKIRSQKQKNP EIFQYLLLNQLSNTLITDELNNNVIKNTNSIETLDNDIAWLNDLIDLDKEVGVLSRDIGS LHDDVAQNQADIKTLENNVVEELFNLSDRLLIDQEAETIAQNNESIEDLYDFGREVAESIGEI HAHNEAQNETLKDLITNSVKNTDNIDKNKADIQALENNVEEGLLELSGHLIDQKADLTKDI KALESNVEEGLIDLSGRLLDQKADIKNQADIAQNQTDIQDLAAYNELQDQYAKQTEAID ALNKASSENTQNIEDLAAYNELQDAYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAY AKQQTEAIDALNKASSENTQNIKNQADIANNINNIYELAQQQDQHSSDIKTLAKVSAANT DRIAKNKADADASFETLTKNQNTLIEKDKEHDKLITANKTAIDANKASADTKFAATADAIT KNGNAITKNAKSITDLGTVKVDGFSRVTAIDTKVNAFDGRITALDSKVENGMAAQAALSGL FQPYSGKFNATAALGGYGSKSAVAIGAGYRVNPNLAFKAGAAINTSGNKKGSYNIGVNYE F (611 amino acids)

[34] UspA2 may be UspA2 from *M. catarrhalis* strain ATCC(a US registered trademark) 25238™, American 2933, American 2912, American 2908, Finnish 307, Finnish 353, Finnish 358, Finnish 216, Dutch H2, Dutch F10, Norwegian 1, Norwegian 13, Norwegian 20, Norwegian 25, Norwegian 27, Norwegian 36, BC5SV, Norwegian 14, Norwegian 3, Finish 414, Japanese Z7476, Belgium Z7530, German Z8063, American O12E, Greek MC317, American V1122, American P44, American V1171, American TTA24, American O35E, American SP12-6, American SP12-5, Swedish BC5, American 7169, Finnish FIN2344, American V1118, American V1145 or American V1156. UspA2 may be UspA2 as set forth in any of SEQ ID NO: 1 – SEQ ID NO: 38. UspA2 may be UspA2 from another source which corresponds to the sequence of UspA2 in any one of SEQ ID NO: 1 – SEQ ID NO: 38. Corresponding UspA2 sequences may be determined by one skilled in the art using various algorithms. For example, the Gap program or the Needle program may be used to determine UspA2 sequences corresponding to any one of SEQ ID NO: 1 – SEQ ID NO: 38.

[35] UspA2 may be a sequence with at least 95% identity, over the entire length, to any of SEQ ID NO: 1 – SEQ ID NO: 38.

[36] Immunogenic fragments of UspA2 comprise immunogenic fragments of at least 450 contiguous amino acids of SEQ ID NO: 1, 490 contiguous amino acids of SEQ ID NO: 1 (for example, the UspA2 fragment of MC-004 or MC-005), 511 contiguous amino acids of

SEQ ID NO: 1 (for example, the UspA2 fragment of construct MC-001, MC-002, MC-003 or MC-004), 534 contiguous amino acids of SEQ ID NO: 1 (for example, the UspA2 fragment of MC-009 or MC-011) or 535 contiguous amino acids of SEQ ID NO: 1 (for example, the UspA2 fragment of MC-007, MC-008 or MC-010). The immunogenic fragments may elicit antibodies which can bind SEQ ID NO: 1.

[37] Immunogenic fragments of UspA2 may comprise immunogenic fragments of at least 450, 490, 511, 534 or 535 contiguous amino acids of any of SEQ ID NO: 1 – SEQ ID NO: 38. Immunogenic fragments of UspA2 may comprise immunogenic fragments of UspA2 from any of SEQ ID NO: 2 – SEQ ID NO: 38 which correspond to the UspA2 fragment of SEQ ID NO: 1 in any of the UspA2 constructs MC-001, MC-002, MC-003, MC-004, MC-005, MC-006, MC-007, MC-008, MC-009, MC-010 or MC-011. The immunogenic fragments may elicit antibodies which can bind the full length sequence from which the fragment is derived.

[38] Alignments between polypeptides pairs may be calculated by various programs. 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(a US registered trademark) package (Accelrys Inc.) may be used.

[39] The Gap and Needle programs are an implementation of the Needleman-Wunsch algorithm described in: Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453. These programs are using frequently the BLOSUM62 scoring matrix (Steven Henikoff and Jorja G. Henikoff (1992), "Amino acid substitution matrices from protein blocks", Proc. Natl. Acad. Sci. USA 89 (Biochemistry): 10915-10919) with gap open and extension penalties of, respectively, 8 and 2. Sometimes, the PAM250 scoring matrix (Dayhoff 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) is also used.

[40] Scoring matrices are describing by numbers the tendency of each amino acid to mutate in another, or to be conserved. These numbers are generally computed from statistics of mutations observed in faithful pairwise or multiple alignments, or even in fragments of multiple alignments. Generally, in these tables, if a high positive number is associated with a pair of identical amino acids, it is indicating that this residue has a low tendency for mutation. At the

opposite, a high positive number associated with a pair of different amino acids is indicating a high tendency of mutation between these two. And this is called a "conservative substitution".

[41] Looking at a pairwise alignment, aligned identical residues ("identities") between the two sequences can be observed. A percentage of identity can be computed by multiplying by 100 (1) the quotient between the number of identities and the length of the alignment (for example, in the Needle program output), or (2) the quotient between the number of identities and the length of the longest sequence, or (3) the quotient between the number of identities and the length of the shortest sequence, or (4) the quotient between the number of identities and the number of aligned residues (for example, in the Gap program output).

[42] The percentage of identities of Table 8 have been calculated according the definition (3) of the previous paragraph, using the pairwise alignments computed by the Gap software.

[43] 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.

[44] The invention further provides proteins of formula (I) containing conservative amino acid substitutions. For example, the proteins of formula (I) may contain a conservative substitution of any amino acid from UspA2 of *Moraxella catarrhalis* as described in any of the sequences set forth herein (for example, any UspA2 sequence set forth in SEQ ID NO. 1 – SEQ ID NO. 38).

[45] 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.

[46] A heterologous signal peptide may be cleaved from a 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.

[47] 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.

[48] 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.

[49] 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* (NTHi), 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)). Recent antibiotic pressure and vaccination with the pneumococcal conjugate vaccine have resulted in the emergence of β -lactamase-producing *Haemophilus influenzae* and *Moraxella catarrhalis* as the leading organisms causing acute otitis media in North America, followed by *Streptococcus pneumoniae* (Pediatr Clin N Am 60 (2013) 391-407).

[50] 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)).

[51] 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)).

[52] 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.

[53] 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)).

[54] 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)).

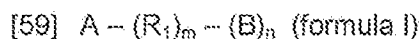
[55] The course of COPD is characterized by progressive worsening of airflow limitation and a decline in pulmonary function. COPD may be complicated by frequent and recurrent acute exacerbations (AE), which are associated with enormous health care expenditure and high morbidity. (Proceedings of the American 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)). *Haemophilus influenzae* (*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 Phillipines, 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 (acute exacerbation of COPD) 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)).

[56] Community-acquired pneumonia (CAP) has been described as the leading cause of death from infectious disease and the six-ranked cause of death overall in the United States. *Moraxella catarrhalis* is one of the pathogens associated with CAP in North America (Clin Chest Med 26 (2005) 37 -55) and is one of the pathogens associated with moderate to severe community acquired pneumonia in Japan (J Infect Chemother. 2014 Nov 20. pii: S1341-321X(14)00398-1. doi: 10.1016/j.jiac.2014.11.006. [Epub ahead of print]).

[57] There is a need for effective vaccines against *M. catarrhalis*.

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



[60] wherein:

A is UspA2 from *Moraxella catarrhalis* or an immunogenic fragment thereof;

R₁ is an amino acid;

m is 0 or 2;

B is histidine; and

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

[61] In one particular embodiment, R_1 and m are defined wherein $(R_1)_m$ is AS (alanine serine). In another embodiment, R_1 is non-native amino acids.

[62] In one embodiment, the proteins of formula (I) and proteins of the invention are defined wherein m is 0. In one embodiment, when m is 0, n is 2. In another embodiment of the invention, when m is 0, n is not 0.

[63] In one embodiment, m is 2.

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

[65] In one embodiment, n is selected from the group consisting of 0, 1, 2, and 6, or any subset thereof.

[66] In one embodiment n is 0. In another embodiment, when n is 0, m is 2.

[67] In one embodiment, n is 1. In one embodiment, n is 3. In one embodiment, n is 4. In one embodiment, n is 5.

[68] In one embodiment, the proteins of formula (I) further contain a methionine (M) at the amino terminus; a protein with the following formula: methionine- A - $(R_1)_m$ - $(B)_n$. These are included within proteins of the invention. In one particular embodiment, when m is 0 and n is 0, the proteins of formula (I) and proteins of the invention are non-native proteins.

[69] In one embodiment, the proteins of formula (I) and proteins of the invention are non-native proteins.

[70] In one embodiment, the proteins of formula (I) are defined wherein A is UspA2 from *M. catarrhalis*. In another embodiment, the proteins of formula (I) are defined wherein A is UspA2 as set forth in an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, 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 and SEQ ID NO: 38 or any subset of SEQ ID NO: 1 through SEQ ID NO: 38. In another embodiment, the proteins of formula (I) are defined wherein A is UspA2, wherein UspA2 is at least 63%, 66%, 70%, 72%, 74%, 75%, 77%, 80%, 84%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical, over the entire length, to SEQ ID NO: 1. In another embodiment, the proteins of formula (I) are defined wherein A is UspA2, wherein UspA2 is approximately 75% to 100% identical to the UspA2 amino acid sequence set forth in SEQ ID NO: 1. In another embodiment, A is UspA2 wherein UspA2 is approximately 90% to 100% identical to the UspA2 amino acid sequence set forth in SEQ ID NO: 1. In another embodiment, A is UspA2 wherein UspA2 is at least 95% identical to the UspA2 amino acid sequence set forth in SEQ ID NO: 1. In another embodiment, the proteins of formula (I) are defined wherein A is UspA2, wherein UspA2 is approximately 75% to 100% identical to the UspA2 amino acid sequence set forth in any one of SEQ ID NO: 1 – SEQ ID NO: 38. In another embodiment, A is UspA2 wherein UspA2 is approximately 90% to 100% identical to the UspA2 amino acid sequence set forth in any one of SEQ ID NO: 1 – SEQ ID NO: 38. In additional embodiment, A is UspA2 wherein UspA2 is at least 95% identical to UspA2 as set forth in any of SEQ ID NO: 1 – SEQ ID NO: 38. In a particular embodiment, A is UspA2 having the amino acid sequence set forth in SEQ ID NO: 1.

[71] In another embodiment, the proteins of formula (I) are defined wherein A is an immunogenic fragment of UspA2 from *M. catarrhalis*. In another embodiment, A is an immunogenic fragment of UspA2 wherein UspA2 has an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, 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 and SEQ ID NO: 38 or any subset of SEQ ID NO: 1 through SEQ ID NO: 38. In another embodiment, A is an immunogenic fragment of UspA2, wherein UspA2 is approximately 75% to 100% identical to the amino acid sequence set forth in SEQ ID NO: 1. In another embodiment, A is an immunogenic fragment of UspA2, wherein UspA2 is approximately 90% to 100% identical to SEQ ID NO: 1. In an another embodiment, A is an immunogenic fragment of UspA2, wherein

UspA2 is at least 95% identical to SEQ ID NO: 1. In another embodiment, A is an immunogenic fragment of UspA2, wherein UspA2 is approximately 75% to 100% identical to the amino acid sequence set forth in any one of SEQ ID NO: 1 – SEQ ID NO: 38. In another embodiment, A is an immunogenic fragment of UspA2, wherein UspA2 is approximately 90% to 100% identical to any one of SEQ ID NO: 1 – SEQ ID NO: 38. In an additional embodiment, A is an immunogenic fragment of UspA2, wherein UspA2 is at least 95% identical to any of SEQ ID NO: 1 – SEQ ID NO: 38. In a particular embodiment, A is an immunogenic fragment of UspA2 wherein UspA2 has the amino acid sequence set forth in SEQ ID NO: 1.

[72] In another embodiment, A is an immunogenic fragment of UspA2 from *M. catarrhalis* selected from the group consisting of amino acids 30-540 of SEQ ID NO: 1 (SEQ ID NO: 39), amino acids 31-540 of SEQ ID NO: 1 (SEQ ID NO: 40), amino acids 30-519 of SEQ ID NO: 1 (SEQ ID NO: 41), amino acids 30-564 of SEQ ID NO: 1 (SEQ ID NO: 42) and amino acids 31-564 of SEQ ID NO: 1 (SEQ ID NO: 43). More specifically, in one embodiment, A is SEQ ID NO: 43, amino acids 31-564 of SEQ ID NO: 1. In an additional embodiment, A is SEQ ID NO: 42, amino acids 30-564 of SEQ ID NO: 1. In another embodiment, A is an immunogenic fragment of UspA2 from *M. catarrhalis* selected from the group consisting of amino acids 30-540 of SEQ ID NO: 1 (SEQ ID NO: 39), amino acids 31-540 of SEQ ID NO: 1 (SEQ ID NO: 40) and amino acids 30-519 of SEQ ID NO: 1 (SEQ ID NO: 41). In another embodiment, A is an immunogenic fragment of UspA2 with at least 52% (American 2908), 55% (Norwegian 25), 57% (Japanese Z7476), 62% (Finnish FIN2344), 64% (American 2912), 69% (American P44), 73% (American 7169), 76% (Norwegian 27), 81% (American V1145), 88% (German Z8063) or 100% (Swedish BC5) identity to SEQ ID NO: 39. In another embodiment, A is an immunogenic fragment of UspA2 with at least 52 % (American 2908), 57% (Dutch F10), 62% (American 2933), 65% (Greek MC317), 67% (American V1122), 70% (American P44), 73% (American 7169), 76% (Norwegian 3), 81% (German Z8063), 100% (Swedish BC5) identical to SEQ ID NO: 43.

[73] In another embodiment, A is an immunogenic fragment of UspA2 from *M. catarrhalis* from SEQ ID NO: 2 through SEQ ID NO: 38 where the fragment comprises the amino acids that align with amino acids 30-540 of SEQ ID NO: 1 (SEQ ID NO: 39), amino acids 31-540 of SEQ ID NO: 1 (SEQ ID NO: 40), amino acids 30-519 of SEQ ID NO: 1 (SEQ ID NO: 41), amino acids 30-564 of SEQ ID NO: 1 (SEQ ID NO: 42) or amino acids 31-564 of SEQ ID NO:

1 (SEQ ID NO: 43). In one embodiment, the Gap program (from the GCG package), or the Needle program (from the EMBOSS package), implementing the Needleman-Wunsch algorithm, may be used to align the sequences.

[74] UspA2 - SEQ ID NO: 1

MKTMKLLPLKIAVTSAMIIGLGAASTANAQAKNDITLEDLPYLIKKIDQNELEADIGDIT
ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD
FGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
NIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDTK
VNAFDGRITALDSKVENGMAAQAALSGLFQPYSVCKFNATAALGGYGSKSAVAIGAGYRV
NPNLAFKAGAAINTSGNKKGSYNIGVNYEF

[75] Amino acids 30-540 of UspA2 from SEQ ID NO: 1, SEQ ID NO:39

QAKNDITLEDLPYLIKKIDQNELEADIGDIT
ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD
FGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
NIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDTK

[76] Amino acids 31-540 of UspA2 from SEQ ID NO: 1, SEQ ID NO: 40

AKNDITLEDLPYLIKKIDQNELEADIGDIT
ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD

FGHEVAESIGEIH AHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
 RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
 NIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
 LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
 IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
 ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTKVDGFDsrvTALDTK

[77] Amino acids 30-519 of UspA2 from SEQ ID NO: 1, SEQ ID NO: 41

QAKNDITLEDLPYLIKKIDQNELEADIGDIT
 ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
 AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD
 FGHEVAESIGEIH AHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
 RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
 NIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
 LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
 IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
 ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKS

[78] Amino acids 30-564 of UspA2 from SEQ ID NO: 1, SEQ ID NO: 42

QAKNDITLEDLPYLIKKIDQNELEADIGDIT
 ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
 AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD
 FGHEVAESIGEIH AHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
 RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
 NIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
 LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
 IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
 ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTKVDGFDsrvTALDTK
 VNAFDGRITALDSKVENGMAAQAA

[79] Amino acids 31-564 of UspA2 from SEQ ID NO: 1, SEQ ID NO: 43

AKNDITLEDLPYLIKKIDQNELEADIGDIT
 ALEKYLALSQYGNILALEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE
 AIKEDLQGLADDFVEGQEGKILQNETSIKKNTQNLVNGFEIEKNKDAIAKNNESIEDLYD
 FGHEVAESIGEIHAAHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
 RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
 NIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
 LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
 IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
 ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSEVTALDTK
 VNAFDGRITALDSKVENGMAAQAA

[80] In another embodiment, A is an immunogenic fragment of UspA2 from *M. catarrhalis* that differs from SEQ ID NO: 1 in one or more of the following amino acids: AA (amino acid) 30 to 298, AA 299 to 302, AA 303 to 333, AA 334 to 339, AA 349, AA 352 to 354, AA 368 to 403, AA 441, AA 451 to 471, AA 472, AA 474 to 483, AA 487, AA 490, AA 493, AA 529, AA 532 or AA 543. In another embodiment, A is an immunogenic fragment of UspA2 from *M. catarrhalis* that differs from SEQ ID NO: 1 in that it contains at least one amino acid insertion in comparison to SEQ ID NO: 1.

[81] In another embodiment, A is an immunogenic fragment of UspA2 that contains a laminin binding domain and a fibronectin binding domain.

[82] In an additional embodiment, A is an immunogenic fragment of UspA2 that contains a laminin binding domain, a fibronectin binding domain and a C3 binding domain.

[83] In a further embodiment, A is an immunogenic fragment of UspA2 that contains a laminin binding domain, a fibronectin binding domain, a C3 binding domain and an amphipathic helix.

[84] The laminin binding domain, fibronectin binding domain, C3 binding domain or amphipathic helix may be as defined for SEQ ID NO: 1 or may be the corresponding sequence in any one of SEQ ID NO: 2 through SEQ ID NO: 38.

[85] Proteins of formula (I) and proteins of the invention are useful as immunogens in subjects such as mammals, particularly humans. In particular, the proteins of formula (I) and proteins of the invention are useful in inducing an immune response against *M. catarrhalis* in subjects, particularly humans. The proteins of formula (I) and proteins of the invention are useful in the treatment or prevention of *M. catarrhalis* infection or disease. More specifically, the proteins of formula (I) and proteins of the invention are useful in the treatment or prevention of otitis media and/or COPD and/or AECOPD and/or pneumonia.

[86] The present invention relates to immunogenic compositions comprising UspA2 from *M. catarrhalis* or an immunogenic fragment thereof. The present invention also relates to vaccines comprising such immunogenic compositions and therapeutic uses of the same. Immunogenic compositions and vaccines of the present invention are useful in the treatment or prevention of *M. catarrhalis* infection or disease. More specifically, immunogenic compositions and vaccines described herein are useful in the treatment or prevention of otitis media and/or COPD and/or AECOPD and/or pneumonia.

[87] In one embodiment, the immunogenic composition comprises UspA2 from *M. catarrhalis*. UspA2 may be any one of SEQ ID NO: 1 through SEQ ID NO: 38 or a UspA2 sequence at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to any one of SEQ ID NO: 1 through SEQ ID NO: 38. UspA2 may also be a UspA2 sequence at least 63% (American 2908), 66% (Japanese Z7476), 70% (Dutch F10), 72% (Finnish 358), 74% (American P44), 77% (Finnish 307), 80% (Norwegian 3), 84% (American V1145), 90% (German Z8063) or 100% (Swedish BC5) identical to that of SEQ ID NO. 1.

[88] In another embodiment, the immunogenic composition comprises an immunogenic fragment of UspA2. The immunogenic fragment of UspA2 may be SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42 or SEQ ID NO.43, or a sequence having at least 90%, 95%, 96%, 97%, 98%, 99% sequence identity to any one of SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42 or SEQ ID NO.43. The immunogenic fragment of UspA2 may be a UspA2 sequence at least 52% (American 2908), 55% (Norwegian 25), 57% (Japanese Z7476), 62% (Finnish FIN2344), 64% (American 2912), 69% (American P44), 73% (American 7169), 76% (Norwegian 27), 81% (American V1145), 88% (German Z8063) or 100% (Swedish BC5) identical to SEQ ID NO. 39. The immunogenic fragment of UspA2 may also be a UspA2 sequence at least 52 % (American 2908), 57% (Dutch F10), 62% (American 2933), 65% (Greek MC317), 67% (American V1122), 70% (American P44), 73% (American

7169), 76% (Norwegian 3), 81% (German Z8063), 100% (Swedish BC5) identical to SEQ ID NO. 43. Amino acid differences have been described in UspA2 from various *Moraxella catarrhalis* species.

[89] UspA2 contains a laminin binding domain (for example, amino acids 30-177 of SEQ ID NO: 1, SEQ ID NO: 44). In one embodiment, the fragment of UspA2 comprises the laminin binding region of SEQ ID NO: 1. In an additional embodiment, the fragment of UspA2 comprises the laminin binding region of any one of SEQ ID NO: 2 – SEQ ID NO: 38.

[90] Amino acids 30-177 of SEQ ID NO: 1, SEQ ID NO: 44:
QAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDVG
WNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNL
VNGFEIEKNKDAIAKNNESIED.

[91] UspA2 contains a fibronectin binding domain (for example, amino acids 165-318 of SEQ ID NO: 1, SEQ ID NO: 45). In one embodiment, the fragment of UspA2 comprises the fibronectin binding region of SEQ ID NO: 1. In an additional embodiment, the fragment of UspA2 comprises the fibronectin binding region of any one of SEQ ID NO: 2 – SEQ ID NO: 38. The fibronectin binding domain of SEQ ID NO: 45 also has C3 binding properties.

[92] Amino acids 165-318 of SEQ ID NO: 1, SEQ ID NO: 45:

KDAIAKNNESIEDLYD
FGHEVAESIGEIIHAHNEAQNETLKGLITNSIENTNNITKNKADIQALENNVVEELFNLSG
RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
NIQDLATYNELQDQYAQK.

[93] UspA2 contains a complement component 3 (C3) binding domain (for example, amino acids 30-539 of SEQ ID NO: 1, SEQ ID NO: 46, or amino acids 165-318 of SEQ ID NO: 1, SEQ ID NO: 45). In one embodiment, the fragment of UspA2 comprises the C3 binding region of SEQ ID NO: 1. In an additional embodiment, the fragment of UspA2 comprises a C3 binding domain of any one of SEQ ID NO: 2 – SEQ ID NO: 38.

[94] Amino acids 30-539 of SEQ ID NO: 1, SEQ ID NO: 46:

QAKNDITLEDLPYLIKKIDQNELEADIGDIT
ALEKYLALSQYGNITLAEELNKALEELDEDVGWNQNDIANLEDDVETLTKNQNALAEQGE

AIKEDLQGLADFVEGQEGKILQNETSIKKNTQRNLVNGFEIEKNKDAIAKNNESIEDLYD
 FGHEVAESIGETIAHNEAQNETLKGILITNSIENTNNITKNKADIQALENNVVEELFNLSG
 RLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGLLELSGHLIDQKTDIAQNQA
 NIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDA
 LNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIAKNQADIANNINN
 IYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKDKEHDKL
 ITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSDRVTAIDT

[95] UspA2 contains an amphipathic helix (for example, amino acids 519-564 of SEQ ID NO: 1 or or amino acids 520-559 of SEQ ID NO:1). In one embodiment, the fragment of UspA2 comprises amino acids 519-564 of SEQ ID NO: 1. In another embodiment, the fragment of UspA2 comprises amino acids 520-559 of SEQ ID NO:1. In an additional embodiment, the fragment of UspA2 comprises an amphipathic helix of any one of SEQ ID NO: 2 – SEQ ID NO:38.

[96] In one embodiment, the immunogenic composition comprises a protein of formula (I) wherein A is an immunogenic fragment of UspA2 that comprises a laminin binding domain and a fibronectin binding domain.

[97] In an additional embodiment, the immunogenic composition comprises a protein of formula (I) wherein A is an immunogenic fragment of UspA2 that comprises a laminin binding domain, a fibronectin binding domain and a C3 binding domain.

[98] In a further embodiment, the immunogenic composition comprises a protein of formula (I) wherein A is an immunogenic fragment of UspA2 that comprises a laminin binding domain, a fibronectin binding domain, a C3 binding domain and an amphipathic helix.

[99] In another embodiment, the immunogenic composition comprises a protein as defined by formula (I). The immunogenic composition may contain, for example, a protein of formula (I) with an additional methionine at the amino terminus.

[100] In one embodiment, the present immunogenic compositions may be administered with other antigens. For example, the present immunogenic composition may be administered with antigens from *H. influenzae*. For example, the protein of formula (I) may be administered with Protein D (PD) from *H. influenzae*. Protein D may be as described in WO91/18926. The present immunogenic composition may be administered with Protein E (PE) and Pilin A (Pila)

from *H. Influenzae*. Protein E and Pilin A may be as described in WO2012/139225; the contents of which are incorporated herein by reference. Protein E and Pilin A may be presented as a fusion protein.

[101] 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, COPD, AECOPD or pneumonia.

[102] 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 *M. catarrhalis* in a human, for instance, may be expected to lie in the range of from about 0.001 mg - 0.120 mg. More specifically, a typical dose for the treatment of a condition caused wholly or in part by *M. catarrhalis* in a human may lie in the range of from about 0.003 mg to about 0.03 mg of protein. The present invention provides an immunogenic composition comprising a protein of formula (I) or a protein of the invention for use in the treatment or prevention of a condition or disease caused wholly or in part by *M. catarrhalis*. 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 0.005 mg to about 0.05 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). The present invention also provides an immunogenic composition comprising a protein of formula (I) or a protein of the invention for use in the treatment or prevention of a condition or disease caused wholly or in part by *Moraxella catarrhalis* in combination with at least one antigen from *Haemophilus influenzae*.

[103] 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.

[104] 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.

[105] 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 protein may be adsorbed onto aluminium phosphate. In another embodiment, the protein may be adsorbed onto aluminium hydroxide. In a third embodiment, alum may be used as an adjuvant.

[106] 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 (Thoenen et al. Vaccine (1998) 16:708-14; EP 689454-B1).

[107] 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.

[108] AS02 is an Adjuvant System containing MPL and QS21 in an oil/water emulsion. AS02V is an Adjuvant System containing MPL and QS21 in an oil/water emulsion (50 µg MPL and 50 µg QS21).

[109] AS03 is an Adjuvant System containing α-Tocopherol and squalene in an oil/water (o/w) emulsion. AS03_A is an Adjuvant System containing α-Tocopherol and squalene in an o/w emulsion (11.86 mg tocopherol). AS03_B is an Adjuvant System containing α-Tocopherol and squalene in an o/w emulsion (5.93 mg tocopherol). AS03_C 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.

[110] AS04 is an Adjuvant System containing MPL (50 µg MPL) adsorbed on an aluminum salt (500 µg Al³⁺). In one embodiment, the immunogenic composition or vaccine comprises AS04.

[111] 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.

[112] 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.

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

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

[115] Possible excipients include arginine, pluronic acid and/or polysorbate. In a preferred embodiment, polysorbate 80 (for example, TWEEN(a US registered trademark) 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.

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

[117] 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.

[118] 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.

[119] 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 a protein of formula (I) or a protein of the invention.

[120] The present invention provides a method for the treatment or prevention of exacerbations in chronic obstructive pulmonary disease. The exacerbation of COPD may be an acute exacerbation. The method comprises administering to a subject in need thereof a therapeutically effective amount of the protein of formula (I) or a protein of the invention.

[121] 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 protein of formula (I) or a protein of the invention.

[122] The present invention provides a pharmaceutical composition comprising a protein of formula (I) or a protein of the invention for use in the treatment or prevention of a condition or

disease caused wholly or in part by *Moraxella catarrhalis*. Pharmaceutical compositions may further comprise a pharmaceutically acceptable adjuvant.

[123] The present invention provides use of (a) proteins of formula (I) and proteins of the invention, (b) an immunogenic composition comprising a protein of formula (I) or protein of the invention or (c) a vaccine comprising (c1) a protein of formula (I) or protein of the invention or (c2) immunogenic composition comprising a protein of formula (I) or protein of the invention for the manufacture of a medicament for treating or preventing *M. catarrhalis* infection or disease.

[124] The present invention provides use of (a) proteins of formula (I) and proteins of the invention, (b) an immunogenic composition comprising a protein of formula (I) or protein of the invention or (c) a vaccine comprising (c1) a protein of formula (I) or protein of the invention or (c2) immunogenic composition comprising a protein of formula (I) or protein of the invention for the manufacture of a medicament for treating or preventing otitis media.

[125] The present invention provides use of (a) proteins of formula (I) and proteins of the invention, (b) an immunogenic composition comprising a protein of formula (I) or protein of the invention or (c) a vaccine comprising (c1) a protein of formula (I) or protein of the invention or (c2) immunogenic composition comprising a protein of formula (I) or protein of the invention for the manufacture of a medicament for treating or preventing acute exacerbations of chronic obstructive pulmonary disease (AECOPD).

[126] The present invention provides a use of (a) proteins of formula (I) and proteins of the invention, (b) an immunogenic composition comprising a protein of formula (I) or protein of the invention or (c) a vaccine comprising (c1) a protein of formula (I) or protein of the invention or (c2) immunogenic composition comprising a protein of formula (I) or protein of the invention for the manufacture of a medicament for treating or preventing pneumonia.

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

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

6xhis = six histidines;

xg = centrifugal force (number gravities)

AS = alanine serine

BSA = bovine serum albumin;
°C = degrees Celsius;
CaCl₂ = calcium chloride;
CD = circular dichroism;
CHCl₃ = chloroform;
CH₃CN = acetonitrile;
CO₂ = carbon dioxide;
Da = dalton;
DNA = deoxyribonucleic acid;
DO = dissolved oxygen;
DSC = differential scanning calorimetry;
EDTA = ethylenediaminetetraacetic acid;
h = hour;
H₂O = water;
H₂O₂ = hydrogen peroxide;
HCDI = high cell density induction;
HCl = hydrogen chloride;
His = his = histidine;
IMAC = immobilized metal affinity chromatography;
IPTG = isopropyl β-D-1-thiogalactopyranoside;
kVolts = kilovolts
L = liter;
LB = Luria-Bertani;
LCDI = low cell density induction;
MeOH = methanol;
ml = milliliter;
NaCl = sodium chloride;
RPM = rpm = revolutions per minute;
min = minute;
mM = millimolar;
μg = microgram;
μL = microliter;
MW = molecular weight;
m/z = mass/charge;

NaCl = sodium chloride;

NaPO₄ = sodium phosphate;

ng = nanogram;

NH₄OH = ammonium hydroxide;

nm = nanometer;

O.D. = optical density;

PBS = phosphate buffered saline;

PCR = polymerase chain reaction;

psi = pounds per square inch;

PVDF = polyvinylidene difluoride;

SDS-PAGE = sodium dodecyl sulphate polyacrylamide gel electrophoresis;

TFA = trifluoroacetic acid

T_m = melting point;

T_{m1} = first melting point;

T_{m2} = second melting point;

w/v = weight/volume.

[129] EXAMPLES

Example 1: Protein Constructs

Protein constructs were produced with different fragments of UspA2 with and without additional amino acids. The following table describes protein constructs made.

Table 2: Protein constructs containing UspA2 protein.

Construct ID	Description	N-terminal	C-Terminal
MC-001	UspA2 + 1/2 helix + 6His	UspA2 fragment (A.A.: 30-540 of SEQ ID NO: 1, SEQ ID NO: 47)	ASHHHHHH
	A.A.	30	540 541 548
MC-002	UspA2 + 1/2 helix	UspA2 fragment (A.A.: 30-540 of SEQ ID NO: 1, SEQ ID NO: 47)	
	A.A.	30	540
MC-003	UspA2 + 1/2 helix +1His	UspA2 fragment (A.A.: 30-540 of SEQ ID NO: 1, SEQ ID NO: 47)	H
	A.A.	30	540 541
MC-004	UspA2 + 1/2 helix + 2His	UspA2 fragment (A.A.: 30-540 of SEQ ID NO: 1, SEQ ID NO: 47)	HH
	A.A.	30	540 541 542
MC-005	UspA2 Δ helix + 6His	UspA2 fragment (A.A.: 30-519 of SEQ ID NO: 1, SEQ ID NO: 48)	ASHHHHHH
	A.A.	30	519 520 527
MC-006	UspA2 Δ helix	UspA2 fragment (A.A.: 30-519 of SEQ ID NO: 1, SEQ ID NO: 48)	
	A.A.	30	519
MC-007	UspA2 + helix + 6His	UspA2 fragment (A.A.: 30-564 of SEQ ID NO: 1, SEQ ID NO: 49)	ASHHHHHH
	A.A.	30	564 565 572
MC-008	UspA2 + helix + 2His	UspA2 fragment (A.A.: 30-564 of SEQ ID NO: 1, SEQ ID NO: 49)	HH
	A.A.	30	564 565 566
MC-009	UspA2 + helix + 2His Δ Q	UspA2 fragment (A.A.: 31-564 of SEQ ID NO: 1, SEQ ID NO: 50)	HH
	A.A.	31	564 565 566
MC-010	UspA2 + helix	UspA2 fragment (A.A.: 30-564 of SEQ ID NO: 1, SEQ ID NO: 49)	
	A.A.	30	564
MC-011	UspA2 + 1/2helix + 6His Δ Q	UspA2 fragment (A.A.: 31-540 of SEQ ID NO: 1, SEQ ID NO: 51)	ASHHHHHH
	A.A.	31	540 541 548

A.A. = amino acid

The DNA and amino acid sequences for each protein constructs listed in Table 2 are set forth below.

PROTEIN CONSTRUCT SEQUENCES:

MC-001 (DNA) - SEQ ID NO: 52

ATGCAGGCCAAAAATGATATTACCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGCCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTTGAAATTGAAAAAACAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAACGTCTGGTCACCTGATCGATCAGAAAACCTGATATTGCCCAGAATCAGGCAAAT
ATTGAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAAC
AACGAACTGCAGGATGCCATATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACC
CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAGTTGATGGTTTTTGATAGC
CGTGTGACCGCACTGGATACCAAAGCAAGCCATCATCATCACCACCACTAA.

MC-001 (protein) – (M)(UspA2 amino acids 30 - 540)(ASHHHHHH) SEQ ID NO: 53

MQAKNDITLEDLPYLIIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNTQRN
LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGITNSIENTNNI
TKNKADIQALENNVVEELENLSGRLLDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELODQYAQKQTEAIDALNKASSENTQNIEDLAAY

NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDS
RVTALDTKASHHHHHH

MC-002 (DNA) - SEQ ID NO: 54

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAAATCAG
AATGCACCTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCCAAAATTCTGCAGAACGAAACAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCCTGAAAAAACGTTGAAGAAGGT
CTGCTGGAACCTGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAAT
ATTGAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAACACCCAGAATATCGAAGATCTGECAGCATAAC
AACGAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCGAATGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATAACCAATAA

MC-002 (Protein) – (M)(UspA2 amino acids 30-540) SEQ ID NO: 55

MQAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNQQRN
LVNGFETIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHANNEAQNETLKLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELQDQYQKQTEAIDALNKASSENTQNIEDLAAY
NELQD

AYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNIKNQ
ADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTLIEKD
KEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDSESRVTAL
DTK

SEQ ID NO: 56

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTGAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAGCAGATATTCAGGCGCTGGAAAATAATGTTGTGGAAGAACTGTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAAGTGTCTGGTCACCTGATCGATCAGAAAACCTGATATTGCCCAGAATCAGGCAAAAT
ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAAC
AACGAACTGCAGGATGCCCTATGCAAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACC
CTGATTGAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATACCAAAGTTAATGCATTTGATGGTCGTATTACCGCTCTGGATAGT
AAAGTTGAAAATGGAATGGCAGCACAAAGCAGCACACTAA

SEQ ID NO: 57

MQAKNDITLEDLPYLIKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNQORN
LVNGFEEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
TKNKADIQALENNVVEELENLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELQDQYQKQTEAIDALNKASSENTQNIEDLAAY

NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGDFS
RVTALDTKVNAFDGRITALDSKVENGMAAQAAH

MC-003 (DNA) – SEQ ID NO: 87

ATGCAGGCCAAAAATGATATTACCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAACCCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAACCTGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAAT
ATTGAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAACACCCAGAATATCGAAGATCTGGCAGCATAAC
AACGAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCAGAACACC
CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATACCAAACACTAA

MC-003 (Protein) – (M)(UspA2 amino acids 30-540)(H) – SEQ ID NO: 88

MQAKNDITLEDLPYLIKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLOGGLADFEVGEQEGKILQNETSIKKNQTN
LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNAIQDLATYNELQDQYQKQTEAIDALNKASSENTQNIEDLAAY

NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDS
RVTALDTKH

MC-004 (DNA) - SEQ ID NO: 58

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCCAAAAATTTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATCAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAGCAGATATTCAGGCGCTGGAAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAACGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAAT
ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAAACACCCAGAATATCGAAGATCTGGCAGCATA
AACGAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCAGAACACC
CTGATTGAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATACCAAACATCATTA

MC-004 (Protein) – (M)(UspA2 amino acids 30-540)(HH) SEQ ID NO: 59

MQAKNDITLEDLPYLIKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNTQRN
LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHANEAQNETLKGLITNSIENTNNI
TKNRADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNOANIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAY
NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN

IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDG
RVTALDTKHH

MC-005 (DNA) - SEQ ID NO: 60

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATFACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAATAATGTTGTGCAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAACCTGTCTGGTCACCTGATCGATCAGAAAACCTGATATTGCCCAGAATCAGGCAAAT
ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAC
AACGAACTGCAGGATGCCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATPGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACC
CTGATTGAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAAT
GCCATCACCAAAAATGCCAAAAGCGCAAGCCATCATCATCACCACTAA

MC-005 (Protein) – (M)(UspA2 amino acids 30-519)(ASHHHHHH) SEQ ID NO: 61

MQAKNDITLEDLPYLIKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNTQRN
LVNGFEEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHANNEAQNETLKLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELQDQYAKQTEAIDALNKASSENTQNIEDLAAY
NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSASHHHHHH

MC-006 (DNA) - SEQ ID NO: 62

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG
TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCTGACCAAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAGCAGATATTCAGGCGCTGGAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCTGAAAAAAAACGTTGAAGAAGGT
CTGCTGGAACGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAAT
ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAAACACCCAGAATATCGAAGATCTGGCAGCATA
AACGAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCAGAACACC
CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAAATGGCAAT
GCCATCACCAAAAAATGCCAAAAGCTAA

MC-006 (Protein) – (M)(UspA2 amino acids 30-519) SEQ ID NO: 63

MQAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNTQRN
LVNGFEEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAY
NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKS

MC-007 (DNA) - SEQ ID NO: 64

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG

TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCTGACCAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCTGAAAAAACGTTGAAGAAGGT
CTGCTGGAAGTGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAAT
ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAACACCCAGAATATCGAAGATCTGGCAGCATAC
AACGAACTGCAGGATGCCATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACATATCAATAATATCTATGAACTGGCCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCAGAACACC
CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAAT
GCCATCACCAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATACCAAAGTTAATGCATTTGATGGTCTGATTACCGCTCTGGATAGT
AAAGTTGAAAATGGTATGGCAGCACAGGCAGCAGCAAGCCATCATCATCACCACCACCTAA

MC-007 (Protein) – (M)(UspA2 amino acids 30-564)(ASHHHHHH) SEQ ID NO: 65

MQAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFVEGQEGKILQNETSIKKNTRQ
LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKT DIAQNQANIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAY
NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQ
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFD
RVTAIDTKVNAFDGRITALDSKVENGMAAQAAASHHHHHH

MC-008 (DNA) - SEQ ID NO: 66

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAATCGATCAG
AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG

TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCTGACCAAAAATCAG
AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
GAAGCACAGAATGAAACCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
ACCAAAAAACAAAGCAGATATTCAGGCGCTGGAAAATAATGTTGTGGAAGAACTGTTTAATCTG
AGCGGTCTGCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCTGAAAAAACGTTGAAGAAGGT
CTGCTGGAACCTGTCTGGTCACCTGATCGATCAGAAAACCTGATATTGCCCAGAATCAGGCAAT
ATTGAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
ATTGATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAAC
AACGAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
ATCGCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCAGCAG
CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAAATACCGATCGT
ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACC
CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAAT
GCCATCACCAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAGTTGATGGTTTTGATAGC
CGTGTGACCGCACTGGATACCAAAGTTAATGCATTTGATGGTTCGTATTACCGCTCTGGATAGT
AAAGTTGAAAATGGTATGGCAGCACAGGCAGCACACCACTAA

MC-008 (Protein) – (M)(UspA2 30-564)(HH) SEQ ID NO: 67

MQAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
GWNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNTQRN
LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
LLELSGHLIDQKTDIAQNQANIQDLATYNELODQYAQKQTEAIDALNKASSENTQNIEDLAAY
NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELODAYAKQQTEAIDALNKASSENTQNI
IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNT
LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTRVDFGDS
RVTALDTKVNAFDGRITALDSKVENGMAAQAHH

MC-009 (DNA) - SEQ ID NO: 68

ATGGCGAAAAATGATATTACCCTGGAAGATCTGCCGTATCTGATCAAAAAATCGATCAGAAC
GAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAGTAT

GGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTGGGT
 TGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCCTGACCAAAAATCAGAAT
 GCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGTGAA
 GGTGAGGAAGGCCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAATCTG
 GTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATTGAA
 GATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAACGAA
 GCACAGAATGAAACCCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATTACC
 AAAAAACAAAGCAGATATTCAGGCGCTGGAATAATGTTGTGGAAGAACTGTTAATCTGAGC
 GGTGCTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAACAACATTTATGAACTGGCA
 CAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCCTGAAAAAACGTTGAAGAAGGTCTG
 CTGGAACCTGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAATATT
 CAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCAATT
 GATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATACAAC
 GAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCAAGC
 TCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCATATAATGAATTACAGGATGCGTATGCC
 AAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAATATC
 GCCAAAAATCAGGCCGATATTGCCAACAATATCAATAATATCTATGAACTGGCCAGCAGCAG
 GATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGTATT
 GCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACCCCTG
 ATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCAAAAT
 AAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAATGCC
 ATCACCAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGCCGT
 GTGACCGCACTGGATACCAAAGTTAATGCATTTGATGGTTCGTATTACCGCTCTGGATAGTAAA
 GTTGAAAATGGTATGGCAGCACAGGCAGCACACCACTAA

MC-009 (Protein) – (M)(UspA2 31-564)(HH) SEQ ID NO: 69

MAKNDITLEDLPYLIKKTIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDVG
 WNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNTQARNL
 VNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHANNEAQNETLKGLITNSIENTNNIT
 KNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGL
 LELSGHLIDQKTDIAQNQANIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYN
 ELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI
 AKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTL
 IEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDNR
 VTALDTKVNAFDGRITALDSKVENGMAAQAHH

MC-010 (DNA) - SEQ ID NO: 70

ATGCAGGCCAAAAATGATATTACCCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAG
 AACGAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAG

TATGGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTG
 GGTTGGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCTGACCAAAAATCAG
 AATGCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGT
 GAAGGTCAGGAAGGCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAAT
 CTGGTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATT
 GAAGATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAAC
 GAAGCACAGAATGAAACCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATT
 ACCAAAAACAAAGCAGATATTCAGGCGCTGGAAAATAATGTTGTGGAAGAACTGTTTAATCTG
 AGCGGTCGTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTG
 GCACAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCTGAAAAAACCGTTGAAGAAGGT
 CTGCTGGAACCTGTCTGGTCACCTGATCGATCAGAAAACCTGATATTGCCCAGAATCAGGCAAAT
 ATTCAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCA
 ATTGATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAC
 AACGAACTGCAGGATGCCATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCA
 AGCTCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTAT
 GCCAAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAAT
 ATCGCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCAGCAG
 CAGGATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAAATACCGATCGT
 ATTGCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACAGAACACC
 CTGATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCA
 AATAAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAAT
 GCCATACCAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTFTTTGATAGC
 CGTGTGACCGCACTGGATACCAAAGTTAATGCATTTGATGGTCGTATTACCGCTCTGGATAGT
 AAAGTTGAAAATGGTATGGCAGCACAGGCAGCATAA

MC-010 (Protein) – (M)(UspA2 amino acids 30-564) SEQ ID NO: 71

MQAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLALSQYGNILALEELNKALEELDEDV
 GWNQNDIANLEDDVETILTQKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNTQRN
 LVNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNI
 TKNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEG
 LLELSGHLIDQKTDIAQNQANIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAY
 NELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQN
 IAKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTQNT
 LIEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFDS
 RVTALDTKVNAFDGRITALDSKVENGMAAQAA

MC-011 (DNA) - SEQ ID NO: 72

ATGGCGAAAAATGATATTACCCTGGAAGATCTGCCGTATCTGATCAAAAAAATCGATCAGAAC
 GAACTGGAAGCCGATATTGGTGATATTACCGCACTGGAAAAATATCTGGCACTGAGCCAGTAT

GGAAATATTCTGGCCCTGGAAGAACTGAATAAAGCTCTGGAAGAGCTGGATGAAGATGTGGGT
 TGAATCAGAATGATATCGCCAATCTGGAAGATGATGTTGAAACCCTGACCAAAAATCAGAAT
 GCACTGGCAGAACAGGGTGAAGCAATTAAAGAAGATCTGCAGGGTCTGGCAGATTTTGTGAA
 GGTGAGGAAGGCCAAAATTCTGCAGAACGAAACCAGCATCAAAAAAACACCCAGCGTAATCTG
 GTGAATGGCTTTGAAATTGAAAAAACAAAGATGCCATTGCCAAAAACAACGAAAGCATTGAA
 GATCTGTATGATTTTGGTCATGAAGTTGCCGAAAGCATTGGTGAAATTCATGCACATAACGAA
 GCACAGAATGAAACCCTGAAAGGTCTGATTACCAACAGCATCGAAAATACCAATAACATTACC
 AAAAACAAAGCAGATATTCAGGCGCTGGAATAATGTTGTGGAAGAACTGTTTAATCTGAGC
 GGTGCTCTGATTGATCAGAAAGCCGATATCGATAATAACATTAAACAACATTTATGAACTGGCA
 CAGCAGCAGGATCAGCATAGCAGCGATATCAAAACCCTGAAAAAACGTTGAAGAAGGTCTG
 CTGGAACCTGTCTGGTCACCTGATCGATCAGAAACTGATATTGCCCAGAATCAGGCAAATATT
 CAGGATCTGGCCACCTATAATGAACTGCAGGATCAGTATGCACAGAAACAGACCGAAGCAATT
 GATGCCCTGAATAAAGCGAGCAGCGAAAACACCCAGAATATCGAAGATCTGGCAGCATAACA
 GAACTGCAGGATGCCTATGCAAAACAGCAGACTGAAGCCATCGACGCACTGAACAAGGCAAGC
 TCTGAAAACACGCAGAACATTGAAGATCTGGCTGCCTATAATGAATTACAGGATGCGTATGCC
 AAACAGCAGACCGAAGCGATTGATGCGCTGAACAAAGCCTCTTCTGAAAATACACAGAATATC
 GCCAAAAATCAGGCCGATATTGCCAACAAATATCAATAATATCTATGAACTGGCCAGCAGCAG
 GATCAGCACTCTTCTGATATCAAAACACTGGCAAAAGCAAGCGCAGCAAATACCGATCGTATT
 GCGAAAAACAAAGCCGATGCAGATGCAAGCTTTGAAACACTGACGAAAAACCAGAACACCCTG
 ATTGAAAAAGATAAAGAACATGATAAACTGATCACCGCCAATAAAACCGCAATTGATGCAAAT
 AAAGCCAGCGCAGATACCAAATTTGCAGCAACCGCAGATGCAATTACCAAAAATGGCAATGCC
 ATCACCAAAAATGCCAAAAGCATTACCGATCTGGGCACCAAAGTTGATGGTTTTGATAGCCGT
 GTGACCGCACTGGATACCAAAGCAAGCCATCATCATCACCACCCTAA

MC-011 (Protein) -- (M)(UspA2 amino acids 31-540)(ASHHHHHH) SEQ ID NO: 73

MAKNDITLEDLPYLIKKIDQNELEADIGDITALEKYLLALSQYGNILALEELNKALEELDEDVG
 WNQNDIANLEDDVETLTKNQNALAEQGEAIKEDLQGLADFEVGEQEGKILQNETSIKKNTQRNL
 VNGFEIEKNKDAIAKNNESIEDLYDFGHEVAESIGEIHAHNEAQNETLKGLITNSIENTNNIT
 KNKADIQALENNVVEELFNLSGRLIDQKADIDNNINNIYELAQQQDQHSSDIKTLKKNVEEGL
 LELSGHLIDQKTDIAQNQANIQDLATYNELQDQYAQKQTEAIDALNKASSENTQNIEDLAAYN
 ELQDAYAKQQTEAIDALNKASSENTQNIEDLAAYNELQDAYAKQQTEAIDALNKASSENTQNI
 AKNQADIANNINNIYELAQQQDQHSSDIKTLAKASAANTDRIAKNKADADASFETLTKNQNTL
 IEKDKEHDKLITANKTAIDANKASADTKFAATADAITKNGNAITKNAKSITDLGTVKVDGFSR
 VTALDTKASHHHHHH

Vector Construction and Transformation

DNA Sequence for UspA2 from strain ATCC 25238 – SEQ ID NO: 74.

ATGAAAACCATGAACTTCTCCCTCTAAAAATCGCTGTAACCAAGTGCCATGATTATTGGCTTGGGTG
CGGCATCTACTGCGAATGCGCAGGCTAAAAATGATATAACTTTAGAGGATTTACCATATTTAATAAA
AAAGATTGACCAAAATGAATTGGAAGCAGATATCGGAGATATTACTGCTCTTGAAAAGTATCTAGCA
CTTAGCCAGTATGGCAATATTTTAGCTCTAGAAGAGCTCAACAAGGCTCTAGAAGAGCTCGACGAG
GATGTTGGATGGAATCAGAATGATATTGCAAACTTGGAAGATGATGTTGAAACGCTCACCAAAAAT
CAAAATGCTTTGGCTGAACAAGGTGAGGCAATTAAAGAAGATCTTCAAGGGCTTGACGATTTTGTA
GAAGGGCAAGAGGGTAAAATTCTACAAAATGAACTTCAATTAATAAAAAATACTCAGAGAAACCTTG
TCAATGGGTTTTGAGATTGAGAAAAATAAGATGCTATTGCTAAAAACAATGAGTCTATCGAAGATCT
TTATGATTTTGGTCATGAGGTTGCAGAAAGTATAGGCGAGATACATGCTCATAATGAAGCGCAAAA
TGAAACTCTTAAAGGCTTGATAACAAACAGTATTGAGAATACTAATAATATTACCAAAAACAAAGCT
GACATCCAAGCACTTGAAAACAATGTCGTAGAAGAAGTATTCAATCTAAGCGGTGCGCTAATTGAT
CAAAAAGCAGATATTGATAATAACATCAACAATATCTATGAGCTGGCACAACAGCAAGATCAGCATA
GCTCTGATATCAAAACACTTAAAAAAAATGTCGAAGAAGGTTTGTGAGCTAAGCGGTACCTAAT
TGATCAAAAACAGATATTGCTCAAAACCAAGCTAACATCCAAGATCTGGCCACTTACAACGAGCTA
CAAGACCAGTATGCTCAAAAGCAAACCGAAGCGATTGACGCTCTAAATAAAGCAAGCTCTGAGAAT
ACACAAAACATCGAAGATCTGGCCGCTTACAACGAGCTACAAGATGCCTATGCCAAACAGCAAACC
GAAGCAATTGACGCTCTAAATAAAGCAAGCTCTGAGAATACACAAAACATCGAAGATCTGGCCGCT
TACAACGAGCTACAAGATGCCTATGCCAAACAGCAAACCGAAGCCATTGACGCTCTAAATAAAGCA
AGCTCTGAGAATACACAAAACATTGCTAAAAACCAAGCGGATATTGCTAATAACATCAACAATATCT
ATGAGCTGGCACAACAGCAAGATCAGCATAGCTCTGATATCAAAACCTTGCCAAAAGCAAGTGCTG
CCAATACTGATCGTATTGCTAAAAACAAAGCCGATGCTGATGCAAGTTTTGAAACGCTCACCAAAAA
TCAAAATACTTTGATTGAAAAAGATAAAGAGCATGACAAATTAATTACTGCAAAACAAACTGCGATT
GATGCCAATAAAGCATCTGCGGATACCAAGTTTGCAGCGACAGCAGACGCCATTACCAAAAATGG
AAATGCTATCACTAAAAACGCAAAATCTATCACTGATTTGGGCACTAAAGTGGATGGTTTTGACAGT
CGTGTAAGTGCATTAGACACCAAAGTCAATGCCTTTGATGGTCGTATCACAGCTTTAGACAGTAAA
GTTGAAAACGGTATGGCTGCCCAAGCTGCCCTAAGTGGTCTATTCCAGCCTTATAGCGTTGGTAAG
TTAATGCGACCGCTGCACCTGGTGGCTATGGCTCAAAATCTGCGGTTGCTATCGGTGCTGGCTAT
CGTGTGAATCCAAATCTGGCGTTTAAAGCTGGTGCAGCGATTAAATACCAGTGGTAATAAAAAAGGC
TCTTATAACATCGGTGTGAATTACGAGTTCTAA

Protein Sequence for UspA2 from strain ATCC 25238. - SEQ ID NO. 1 as described above.

Vector Construction

To generate the construct MC-001, DNA fragment coding for an UspA2 gene fragment (amino acids 30 to 540 from strain ATCC 25238) including the *NdeI* / *XhoI* restriction sites to facilitate the cloning (the starting methionine is encoded by *NdeI* site) and the DNA sequence corresponding to the AS (alanine serine) amino acids linker and 6xhis amino acids was codon-optimized (non-native) and synthesized by GENEART(a US registered trademark). Codon-optimized means that the nucleotide sequence was changed from the native sequence without changing the amino acid sequence in order to better fit with the codon usage in *Escherichia coli* for optimal expression. The UspA2 fragment was cloned according to standard methods into the pET-26b expression vector using the *NdeI* / *XhoI* restriction sites.

To generate MC-002, MC-003, and MC-004 constructs, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 30-540 from strain ATCC 25238) using the MC-001 construct as a template, the primer UspA2Nde opt (which contains the methionine start codon), and the primer UspA2opt delta His, A2opt 1His delta AS, and A2opt 2His delta AS, respectively. The UspA2 fragment was cloned according to standard methods into the pET-26b expression vector using the *NdeI* / *XhoI* restriction sites.

To generate the construct MC-005, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 30-519 from strain ATCC 25238) using the MC-001 vector as a template with the primers UspA2Nde opt (which contains the methionine start codon) and R delta hairpin A2opt His. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *XhoI* restriction site was incorporated into the 3' primer. In addition, DNA sequence corresponding to the AS amino acids linker and 6xhis amino acids was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN(a US registered trademark)).

To generate the construct MC-006, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 30-519 from strain ATCC 25238) using the MC-005 construct as a template with the primers UspA2Nde opt (which contains the methionine start codon) and delta His delta hélice. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *XhoI* restriction site was incorporated into the 3' primer.

The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN(a US registered trademark)).

To generate the construct MC-007, DNA fragment coding for an UspA2 gene fragment (amino acids 30 to 564 from strain ATCC 25238) including the *NdeI* / *XhoI* restriction sites to facilitate the cloning (starting methionine is encoded by *NdeI* site) and the DNA sequence corresponding to the AS amino acids linker and 6xhis amino acids was codon-optimized and synthesized by GENEART(a US registered trademark) (plasmid: 1026399). The UspA2 fragment was cloned according to standard methods into the pET-26b expression vector using the *NdeI* / *XhoI* restriction sites.

To generate the construct MC-008, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 30-564 from strain ATCC 25238) using the MC-007 construct as a template with the primers UspA2Nde opt (which contains the methionine start codon) and 2His hélice deltaAS. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *XhoI* restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN(a US registered trademark)).

To generate the construct MC-009, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 31-564 from strain ATCC 25238) using the 1026399 plasmid as the template and the primers N-term cyto Abis (which contains the methionine start codon) and 2His hélice deltaAS. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer including the glutamine deletion and *XhoI* restriction site was incorporated into the 3' primer including two histidine residues. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN (a US registered trademark)). DNA sequencing of the final construct was performed to confirm the correct sequence.

To generate the construct MC-010, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 30-564 from strain ATCC 25238) using the MC-007 construct as a template with the primers UspA2 Nde opt (which contains the methionine start codon) and cyto hélice dHis dAS. DNA sequence corresponding to *NdeI* restriction site was incorporated into the 5' primer and *XhoI* restriction site was incorporated into the 3' primer.

The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN (a US registered trademark)).

To generate the construct MC-011, a polymerase chain reaction was performed to amplify the UspA2 gene fragment (amino acids 31-540 from strain ATCC 25238) using the MC-001 construct as a template with the primers N-term cyto Abis (which contains the methionine start codon) and N-term reverse. DNA sequence corresponding to NdeI restriction site was incorporated into the 5' primer and XhoI restriction site was incorporated into the 3' primer. The generated PCR product was then inserted into the pET-26b(+) cloning vector (NOVAGEN(a US registered trademark)).

A detailed list of PCR primer sequences used for amplifications is illustrated in Table 3. The polymerase chain reaction was performed using Expand High Fidelity PCR System kit (Roche) according to manufacturer's recommendations. Ligation was performed using Rapid DNA Ligation Kit (Roche) according to manufacturer's recommendations.

Table 3: PCR primer sequences used for UspA2 amplifications

Primer ID	DNA Sequence 5' – 3'
UspA2 Nde opt	GAATTCTTAATTAACATATGCAGGCCAAAAATGATATTACCCTG (SEQ ID NO:75)
UspA2opt delta His	GGCGCGCCTCGAGTTATTATTTGGTATCCAGTGCGGTCACACG (SEQ ID NO:76)
UspA2opt 1His delta AS	GGCGCGCCTCGAGTTAGTGTTTGGTATCCAGTGCGGTCACACG (SEQ ID NO:77)
UspA2opt 2His delta AS	GGCGCGCCTCGAGTTAGTGTTTGGTATCCAGTGCGGTCACACG (SEQ ID NO:78)
R delta hairpin A2opt His	GGCGCGCCTCGAGTTAGTGTTGTTGATGATGATGGCTTGCGCTTTTGGCATTTTG GTGATGGCAT (SEQ ID NO:79)
Delta His delta hélice	CCGCTCGAGCTAGCTTTTGGCATTTTTGGTATGGC (SEQ ID NO:80)
N term cytoAbis	GGAATTCCATATGGCGAAAAATGATATTACCCTGGAAGATCTG (SEQ ID NO:81)
2His hélice delta	GGCGCGCCTCGAGTTAGTGTTGCTGCTGCTGCCATACCATT (SEQ ID

AS	NO:82)
Cyto hélice dHis dAS	GGCGCGCCTCGAGTTATGCTGCCTGTGCTGCCATACCATT (SEQ ID NO:83)
N term reverse	CAGTTCATTATAGGTGGCCAGATCCTG (SEQ ID NO:84)

Transformation

Escherichia coli (*E. coli*) BLR (DE3), modified BLR (DE3) or B834(DE3) cells were transformed with plasmid DNA according to standard methods with CaCl₂-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) competent cells were gently thawed on ice. Approximately 4 µl of plasmid (10-100 ng) were mixed using 50-100 µl competent cells. Thereafter, this formulation was incubated on ice for 5 min. To perform the transformation reaction, the formulation was heat pulsed at 42°C for 30 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 150 µl of transformed cell culture was plated and incubated overnight at 37°C.

BLR (DE3): BLR is a *recA*⁻ derivative of BL21 (F⁻ *ompT* *hsdSB*(rB⁻ mB⁻) *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(a US registered trademark). (F⁻ *ompT* *hsdSB* (rB⁻ mB⁻) *gal* *dcm* Δ(*srl-recA*)306::Tn10 (TetR) (DE3).

B834 (DE3) is the parental strain for BL21. These hosts are methionine auxotrophs and allow high specific activity labeling of target proteins with 35S-methionine and selenomethionine for crystallography. The detailed genotype of *E. coli* B834 (DE3) has been published by NOVAGEN(a US registered trademark): F⁻ *ompT* *hsdSB*(rB⁻ mB⁻) *gal* *dcm* *met* (DE3).

Modified BLR (DE3): In order to prevent (phospho)gluconoylation, *Pgl* gene was inserted in the biotin locus located in the BLR (DE3) genome. In addition, to prevent the Ile-Val substitutions, the C219Y mutation in the threonine deaminase gene was corrected.

Genotype: (F- ompT hsdSB (rB- mB-) gal dcm Δ (srl-recA)306::Tn10 (TetR); Δ (bioA-bioD)::Pgl; TD + (C219Y) (DE3).

Example 2: Protein expression using shake flask

Escherichia coli strains transformed with recombinant plasmid were used to inoculate 100 ml of LB broth (Becton, Dickinson and Company) $\pm$ 1% (weight/volume, w/v) glucose (Laboratoire MAT, catalogue number: GR-0101) and 50 μ g/ml kanamycin (Sigma). This preculture was generally grown overnight at 37°C. Twelve ml of the preculture is used to inoculate 500 ml LB broth+50 μ g/ml kanamycine. Cultures were incubated at 37°C with agitation of 150 RPM to reach an O.D._{600nm} of ~0.6.

At an O.D._{600nm} ~0.6, the BLR (DE3) cultures were induced for the expression of the recombinant protein by addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23°C with agitation of 150 RPM. After the induction period, the cultures were centrifuged at 6370g for 20 minutes and the pellets from 350 ml culture were frozen at -20°C separately.

Example 3: Protein purification using phosphate buffer (MC-001 construct and MC-011 construct)

Each bacterial pellet obtained after induction in shake flask was resuspended in 30 ml 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl and Roche COMPLETE(a US registered trademark) Protease Inhibitor Cocktail (1 tablet/50 ml ml buffer). Cell lysis is performed by 3 X French Press extractions (20 000 psi) and clarification is performed by 30 minutes centrifugation at 23700g. Supernatant is harvested and filtrated on 0.22 μ m.

6xHis tagged-proteins were purified on immobilized metal affinity chromatography (IMAC) using XK16 column and 20 ml NINTA resin (Qiagen) previously equilibrated with 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl or PBS buffer pH 8.0 containing 500 mM arginine. The soluble components were loaded on at up to 4 ml/min (producing a "flow through fraction"). After loading on the column, the column was washed with 60ml of 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl at a rate of 4 ml/min

producing a "wash fraction #1. A second wash using same buffer + 10 mM imidazole was performed, producing a "wash fraction #2. Elution was performed using same buffer containing 200 or/and 500 mM imidazole.

Samples from elution fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples containing the protein were dialyzed against 5 liters of 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl. Protein concentration was determined using Lowry method.

Example 4: Protein purification using arginine containing buffer (MC-001, MC-005 and MC-007)

Each bacterial pellet obtained after induction in shake flask (Example 3) or fermenter (Example 5) was resuspended in 30 ml PBS buffer + 500 mM arginine pH8.0 and Roche COMPLETE(a US registered trademark) Protease Inhibitor Cocktail (1 tablet/50 ml buffer). Alternatively, fermentation cell paste ($\approx$ 7g) was resuspended in 90 ml PBS buffer containing 500 mM arginine pH8.0 and Roche COMPLETE(a US registered trademark) Protease Inhibitor Cocktail (1 tablet/50 ml buffer).

Cell lysis was performed by 2 or 3 X French Press extractions (20 000 psi) and clarification was performed by 30 minutes centrifugation at 23 700g 4°C. Supernatant was harvested and filtrated on 0.22 μ m. 6xHis tagged-proteins were purified on immobilized metal affinity chromatography (IMAC) using XK16 column and 80 ml NINTA resin (Qiagen) previously equilibrated with PBS buffer + 500 mM arginine pH 8.0 . The soluble components were loaded on at up to 4 ml/min (producing a "flow through fraction"). After loading on the column, the column was washed with the same buffer, then with 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl at a rate of 4-6 ml/min producing a "wash fraction #1." A second wash using same buffer + 10 mM imidazole was performed, producing a "wash fraction #2." Elution was performed using same buffer + 200 mM imidazole or 500 mM imidazole. In further elution vials, 5 mM EDTA final concentration was added.

Samples from elution fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples containing the proteins were dialyzed against 5 liters of 20 mM potassium phosphate buffer (pH 8.0) containing 10 mM NaCl and 5 mM EDTA. Protein concentration was determined using Lowry method.

This protocol may be used with other 6xHis tagged-proteins.

Example 5: Fermentation

The following fermentation procedure may be used:

Working seeds are frozen aliquots of flask-grown *Escherichia coli* BLR(DE3) or BLR(DE3)-derived strains transformed with a pET26b derivative containing a sequence coding for a specific antigen candidate recombinant protein construct.

A working seed (WS) is removed from frozen storage, thawed and used to inoculate an Erlenmeyer flask containing pre-culture media. Manipulation of the seed and flask culture are performed aseptically under a Laminar Air Flow (LAF) Hood or Biological Safety Cabinet (BSC). The pre-culture flask is incubated typically between 30°C - 37°C under 200 RPM agitation speed for the time needed to reach an Optical Density at 650 nm (OD_{650nm}) between 1.0 and 3.0, typically between 4 – 6 hrs.

A 20L fermentor is prepared by Clean-In-Place followed by an automated steam sterilisation sequence. Starting medium is transferred aseptically into the fermentor. A bottle filled with NH_4OH 25 % is aseptically connected to the fermentor for automatic pH control. The initial pH of the starting medium is adjusted to target pH by addition of NH_4OH solution. Irradiated antifoam is added using a syringe through a septum in the head-plate. A bottle filled with Feed medium is aseptically connected to the fermentor. Feed addition is controlled by either a pO_2 -cascade (control dissolved oxygen) to or a pre-programmed feed-curve. Agitation is controlled either by a pO_2 -cascade or a pre-programmed agitation-curve.

Initial fermentor parameters are typically as follows:

- Temperature: 28°C - 32°C
- Pressure: 0.5 barg (7 psi)
- Air flow rate: 2 VVM (Vessel Volumes per Minute)
- pH: Regulated at 6.8 by addition of NH_4OH 25%.

An aliquot of this pre-culture (typically between 5ml – 50ml) is used to inoculate the starting fermentor media by syringe addition through a septum on the fermentor head-plate. The phases of Fermentation Culture are:

- Batch Phase: Biomass is accumulated using carbon source in starting media.

- Fed-batch Phase: Feed media is introduced either according to pO₂-cascade control or pre-programmed feed curve. Biomass accumulation continues on carbon source in feed media.
- Induction Phase: Expression of the recombinant protein antigen is induced by addition of IPTG solution to the culture in the fermentor.

At harvest, the culture is collected typically in 1L centrifugation bottles and centrifuged to separate the solid pellet (cell-paste) fraction from the liquid supernatant fraction. The supernatant is discarded, and the wet cell weight (solid pellet) is recorded and the cell-paste bags stored at -20°C.

The following procedure may also be used:

Escherichia coli standard pre-culture

Each standard pre-culture were prepared using a frozen seed culture of *Escherichia coli* strains. These strains are BLR(DE3) strains transformed with a pET26b derivative containing a sequence coding for the specific construct to be evaluated.

The seed culture was thawed to room temperature and 400 µl were used to inoculate a 2 litre Erlenmeyer flask containing 400 ml of preculture medium (adapted from Zabriskie et al. (J. Ind. Microbiol. 2:87-95 (1987))).

The inoculated flask was then incubated at 37 °C ($\pm$ 1°C) and 200 rpm. The pre-culture was stopped after 6h of incubation. At this step the optical density at 650nm (OD_{650nm}) is about 2. The pre-culture was used to inoculate medium in a fermenter as soon as the culture was stopped.

20L scale fedbatch fermentation

Method

A 20 litre fermenter (Biolafitte) was used. Nine litres of batch phase medium were aseptically transferred into the fermenter. The pH of the medium was readjusted to 6.8 with base addition. 1 ml of undiluted irradiated antifoam (SAG 471) was also added to the fermenter. The temperature (28°C), head pressure (0.5 bar), aeration rate (20 litres sparged air per minute) and initial agitation speed (300 rpm) were then set prior to inoculation. The level of

dissolved oxygen in these conditions was 100%. The head pressure and aeration rate were maintained at a constant level during the fermentation.

Inoculation was achieved by the addition of an equivalent 10 ml OD650nm = 2 of pre-culture (prepared as described above, in Example 2) following the next formula:

$$\text{Preculture Volume (ml)} = \frac{20}{\text{Preculture Final OD650nm}}$$

During batch phase (0-15h), the temperature was maintained at 28°C. The level of dissolved oxygen was set at 20%. The level of dissolved oxygen (DO) was regulated by increasing stirring when the DO fell below 20%. Glucose exhaustion resulted in an increase in DO and a concomitant decrease in stirring.

When glucose is exhausted, the feeding rate is started based on a pH signal that increases above 7.0. From this point forward, the feeding rate was controlled by oxygen demand, increasing the flow rate when dissolved oxygen tends to drop below the 20% set point. At this step the agitation speed is maintained at 900 rpm.

During the fed-batch phase (before induction), the pH was maintained at 6.8 by addition of base, the temperature was regulated at 30°C.

Two strategies were applied to produce the protein:

The "High Cell Density Induction" (HCDI) is applied when culture is induced with 1 mM IPTG (isopropyl-beta-D-thiogalactopyranoside) at an optical density of 80 ± 5 , typically reached after 40h of culture. The temperature was maintained at 28°C and feeding rate still controlled by oxygen demand with a constant agitation speed at 900 rpm.

The "Low Cell Density Induction" (LCDI) process means an induction at an optical density of 40 ± 5 usually reached after 24h of culture. The temperature was decreased to 30°C and the constant feeding rate of 0.5 ml/min is applied. Then 1 mM IPTG is added to the culture. At this step, the DO level was maintained at 20% by controlling the stirring rate.

At the end of the induction phase (72h), cell paste was collected by centrifugation (6,500xg, 4°C for 1h), and stored at -20°C.

Figures 1 and 2 illustrate a typical fermentation profile with the HCDI and the LCDI processes and the parameters monitored during 20L-scale fed-batch fermentation.

Table 4 sets forth the constructs evaluated in fermenter and UspA2 yield obtained for each one.

Table 4

Construct ID	Name	Helix	His tag	Process used	UspA2 Yield (g/l)
MC-008	UspA2 + Helix + 2 His	Full	2 res	HCDI	2.21
MC-007	UspA2 + Helix + 6 His	Full	6 res	LCDI	2.60
MC-010	UspA2 + Helix	Full	No	HCDI	0.22
MC-005	UspA2 Δ Helix + 6 His	No	6 res	LCDI	1.92
MC-006	UspA2 Δ Helix	No	No	LCDI	1.14
MC-004	UspA2+ $\frac{1}{2}$ Helix + 2 His	1/2	2 res	LCDI	0.92
MC-001	UspA2+ $\frac{1}{2}$ Helix + 6 His	1/2	6 res	HCDI	3.68
MC-002	UspA2+ $\frac{1}{2}$ Helix	1/2	No	HCDI	0.49

His=histidine

Figure 3 depicts in graphical form the UspA2 yield in Table 4 from the constructs evaluated in fermenter.

In this figure, UspA2 yield is affected by histidine residues present in the construct . ($p < 0.05$, one way, three levels, Type II ANOVA). A positive correlation between the number of histidine residues and UspA2 fermentation yield was observed, with a yield increase higher than 400% between 0 and 6 residues in fed-batch fermentations.

It was also observed that one histidine residue added to half-helix pattern (MC-003 construct) produced a UspA2 yield of about 1 g/l of protein.

Example 6: Protein characterization

Analytical ultracentrifugation

Analytical ultracentrifugation is used to determine the homogeneity and size distribution in solution of the different species within a protein sample by measuring the rate at which molecules move in response to a centrifugal force. This is based on the calculation of the coefficients of sedimentation of the different species that are obtained by sedimentation velocity experiment, which depend on their molecular shape and mass.

The following protein samples were spun in a Beckman-Coulter ProteomeLab XL-1 analytical ultracentrifuge at 28 000RPM after the AN-60Ti rotor had been equilibrated to 15°C.

- a. MC-005 lot BMP53 , 675µg/ml in 20mM NaPO₄, 10mM NaCl, pH8,0
- b. MC-001 lot BMP13 , 545µg/ml in 20mM NaPO₄, 10mM NaCl, pH8,0
- c. MC-001 lot BMP14 , 545µg/ml in 20mM NaPO₄, 10mM NaCl, pH8,0
- d. MC-001 lot BMP54 , 445µg/ml in 20mM NaPO₄, 10mM NaCl, pH8,0
- e. MC-007 lot BMP70 , 510µg/ml in 20mM NaPO₄, 10mM NaCl, pH8,0

For data collection, from 133 to 325 scans were recorded at 280nm every 5 minutes.

Data analysis was performed using the program SEDFIT (available through the National Institutes for Health) for determination of the C(S) distribution. The C(S) distribution is a representation of the relative intensity of the different components in a mixture of macromolecules separated by their coefficient of sedimentation, which is a function of molecule size and conformation. Determination of the partial specific volume of the proteins at 15°C was performed with the SEDNTERP software from their amino acid sequence. SEDNTERP (SEDNTERP is distributed and supported through the Biomolecular Interaction Technologies Center at the University of New Hampshire) was also used to determine the viscosity and the density of the buffer at 15°C.

Determination of the relative abundance of all species has been performed by considering the total area under the curve of the overall distribution as 100% of the sample and by calculating the percentage of this total area represented by the contribution of every species. C(S) distribution plot (concentration vs sedimentation coefficient) has been used for that calculation, considering that it's a better representation of the raw data than the C(M) distribution (concentration vs molecular weight).

Analytical ultracentrifugation of the different purified constructs allowed observation that UspA2 Δ helix, UspA2 ½ helix and UspA2 full helix with C-terminal his tag are present mainly as trimers in solution when 500mM L-arginine is added during cell lysis prior to purification (Figures 4, 5, 7 and 8).

A heterogeneous size distribution has been observed for UspA2 ½ helix when no L-arginine was added during cell lysis. Two major populations are observed. It was not possible to confirm the molecular weight of the species detected by AUC (analytical ultracentrifugation) with this protein preparation, since the frictional ratio which is essential for molecular weight estimation needs to be calculated from an homogeneous sample. However, based on sedimentation coefficients, none of the observed species seem to correspond to the trimer observed in other samples.

Figure 4 illustrates the molecular weight distribution of purified MC-005 determined by sedimentation velocity analytical ultracentrifugation. The majority of protein is found as a trimer, with a small proportion of a higher molecular weight oligomer that may correspond to dimer of trimer.

Figure 5 illustrates the molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. Majority of protein is found as a trimer.

Figure 6 illustrates the molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. The sample presents multiple species and is highly polydisperse. The sedimentation coefficient of the major species detected doesn't correspond to the one of the trimers normally detected in the other lots.

Figure 7 illustrates the molecular weight distribution of purified MC-001 determined by sedimentation velocity analytical ultracentrifugation. The majority of protein is found as a trimer.

Figure 8 illustrates the molecular weight distribution of purified MC-007 determined by sedimentation velocity analytical ultracentrifugation. The majority of protein is found as a trimer.

Circular dichroism/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 different temperatures by a Peltier thermostated RTE-111 cell block. A nitrogen flow of 10L/min is maintained during the measurements.

Concentration of the following protein constructs was adjusted to 400µg/ml in 20mM NaPO₄, 10 mM NaCl, pH8,0 buffer.

- a. MC-005 lot BMP53 , in 20mM NaPO₄, 10mM NaCl, pH8,0
- b. MC-001 lot BMP13 , in 20mM NaPO₄, 10mM NaCl, pH8,0
- c. MC-001 lot BMP14 , in 20mM NaPO₄, 10mM NaCl, pH8,0
- d. MC-001 lot BMP54 , in 20mM NaPO₄, 10mM NaCl, pH8,0
- e. MC-007 lot BMP70 , in 20mM NaPO₄, 10mM NaCl, pH8,0

Calculations of secondary structures have been done using the following algorithms:

Selcon 3 (Sreerama and Woody, Anal. Biochem. (1993), 209, 32 ; Sreerama and Woody, Biochemistry, 33, 10022-25 (1994) ; Sreerama et al. Protein Science, 8, 370-380 (1999) ; Johnson W.C.Jr., Proteins:Struc.Func.Genet.35,307-312 (1999))
CDSSTR (Johnson W.C. Proteins:Struc.Func.Genet. 35, 307-312 (1999) modified by Sreerama. N.(Anal.Biochem.,287,252 (2000)).

Displayed results are an average of the percentage calculated with both algorithms and are subjected to a 5% error margin.

Results of the secondary structure calculations for fermentor expressed proteins are displayed in Table 5, considering a 5% error margin.

Table 5: Secondary structure calculations at 22°C

Protein	Helix	Beta	random
MC-005 BMP53	40.8	26.4	34.1
MC-007 BMP70	58.2	18.2	24.7
MC-001 BMP54	53.7	14.2	34.4

Calculations are compatible with the shape and visual analysis of the spectra, where helical content increases with the intensity of minima at 208 and 220nm. Proteins are composed of a high proportion of helical structures, with presence of beta structures.

Superposition of the spectra on Figure 9 shows no significant difference in the shape between constructs. Spectra of MC-005 helix shows a lower intensity which could account for a lower alpha structure that is coherent with the absence of C-terminal helix.

Figure 9 illustrates the far-UV circular dichroism spectra of UspA2 constructs MC-001, MC-005 and MC007 giving an indication of protein secondary structures. Superposition of spectra clearly shows that constructs containing half and full C-terminal helix have no detectable difference in their secondary structures, while the construct without helix generates a spectra with a difference in intensity that could account for a different secondary structure content.

Thermal unfolding

Measurement of far-UV CD spectra at different temperatures during thermal unfolding suggested that MC-005 is less thermally stable than MC-007. The spectra observed at 33°C for MC-005 is similar to the typical spectra of an unfolded protein. For the MC-007 construct, even if partial loss of secondary structures is observed at 33°C, the complete unfolding seems to occur between 35°C and 37°C. This may be an indication of higher thermal stability of the full helix containing construct MC-007.

Figure 10 illustrates secondary structures monitoring by circular dichroism during thermal unfolding of MC-005 (UspA2 Δ helix+6His). Visual analysis of the spectra clearly shows that the protein loses most of its secondary structures at 33°C.

Figure 11 illustrates secondary structures monitoring by circular dichroism during thermal unfolding of MC-007 (UspA2 + helix + 6His). Visual analysis of the spectra shows that loss of secondary structure is slower compared to the construct without helix. Structural changes are detectable upon heating to 33°C, but complete unfolding seems to occur between 35°C and 37°C.

Differential Scanning Calorimetry (DSC) thermal unfolding

Thermal transitions of different UspA2 constructs were compared in order to evaluate the effect of C-terminal helix modifications on thermal stability of the proteins.

Analysis was done on VP-DSC from MicroCal (part of GE Healthcare). The buffer 20mM NaPO₄, 10mM NaCl, 5mM EDTA, pH8 was used as reference and subtracted from the scans. Proteins were equilibrated at initial temperature for 15 minutes before temperature ramping. DSC scans were then conducted from 10°C to 60°C at a heating rate 90°C/hr.

Two transitions were detected in MC-001 and MC-007 constructs and only one in MC-005. Values of the transitions (or T_m) of the different constructs can be found on Table 6.

While the lower T_m of all three proteins is around 32°C, the main difference is the value of the second T_m. The construct containing a full helix (MC-007) has a higher T_m at 37.5°C compared to 34.5°C for the half helix (MC-001).

It has been demonstrated that for MC-001 and MC-007, the first T_m around 32°C is reversible, while the higher T_m is irreversible. For MC-005, the only T_m detected is irreversible.

This may be an indication of a higher thermal stability of full helix containing construct MC-007.

Table 6: Melting points of UspA2 constructs measured by DSC

Constructs	[mg/mL]	T _{m1} (°C)	T _{m2} (°C)
------------	---------	----------------------	----------------------

MC-005 lotBMP53	0.400	31.74	----
MC-001 lotBMP54	0.400	32.02	34.51
MC-007 lotBMP70	0.400	32.19	37.50

Mass spectrometry

UspA2 protein samples were prepared by protein precipitation with CHCl_3 / MeOH / H_2O system. The protein pellet was centrifuged at the bottom of the eppendorf tube before being gently dried under nitrogen. The dried pellet was then dissolved in 2 μl of pure formic acid before being diluted with 3 μl of ultrapure water and 5 μl of sinapinic acid. Sinapinic acid used as matrix for MALDI-TOF (Matrix-Assisted Laser Desorption/Ionisation followed by Time-Of-Flight spectrometry analyser) analysis is prepared in 50% CH_3CN / 50% H_2O supplemented by TFA 0.1% final concentration.

1 μl of the sample + matrix mixture was spotted onto a Bruker 384 ground stainless steel MALDI target and let to dry for crystallization at room temperature and atmospheric pressure (dried droplet method).

UspA2 mass spectrometry analysis was performed on a Bruker Ultraflex 2 MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) in positive ionization and linear mode. Protein samples, co-crystallized in sinapinic acid matrix, were irradiated by a smartbeam laser. Mass measurement of intact UspA2 protein were done over 10.000-100.000 Da mass range with an acceleration voltage of 25 kVolts. Laser attenuation was fine-tuned in order to get the best protein signal as possible and to avoid any fragmentation as well as background over-ionization phenomena. Calibration of the mass spectrometer was performed in close external method with homologous matrix and using the commercial Bruker Protein Calibration mixture 2, by accurate measures on the following calibrators: $[\text{M}+2\text{H}]^{2+}$ (mass measured by MS detector following addition of two H^+ ions to the protein during ionisation) species of

protein A at m/z 22307 Da, $[M+H]^+$ species of trypsinogen at m/z 23982 Da, $[M+H]^+$ species of protein A at m/z 44613 Da and $[M+H]^+$ species of bovine albumin at m/z 66431 Da^o. Each presented spectrum results from the sum of 500 individual shots.

The following samples were analyzed:

MC-001 construct with MQAK amino acids (SEQ ID NO: 85) in N-terminal produced in shake flask, lot opt-01, MC-011 construct with MAK amino acids in N-terminal produced in shake flask, lot BMP37.

In Table 7 and Figure 12, MC-001 protein with MQAK amino acids (SEQ ID NO: 85) at the N-terminal extremity has been shown to be at least partially demethionylated, as shown by the measured molecular mass of 57427Da, compared to the expected mass of 57565Da. The other peak of 57620Da may represent the complete non-demethionylated protein, N-acetylated protein, or another modified protein population.

Figure 12 illustrates the MALDI spectrum of MC-001 lot opt-01. The mass observed at 57427Da may be coherent with the demethionylated protein, while the peak at 57620Da could correspond to the complete protein.

As shown in Table 7 and figure 13, MC-011 protein with MAK amino acids at the N-terminal extremity gave a major population in MALDI-MS that may correspond to the demethionylated protein, with a mass of 57265Da, compared to 57437Da for the expected mass based on complete amino acid sequence. The two other peaks at +186Da and +366Da are not close to any expected post-translational modifications, so they couldn't be identified by this experiment.

Table 7: Molecular mass of two UspA2 constructs as measured by MALDI-MS. Both constructs have a main measured mass lower than the one expected from amino acid sequence. The mass of the major population obtained with both constructs may be coherent with a demethionylated protein.

Protein	Theoretical mass (Da)	Measured mass (Da)	Comment
---------	-----------------------	--------------------	---------

MC-001 lot opt-01	57565.8	-57427.9 -57620.3	-Coherent with demethionylation (57434.6) - Coherent with protein containing N-terminal methionine
MC-011 lot BMP37	57437.6	57265.2	Coherent with demethionylation (57306.4)

N-terminal sequencing by Edman's degradation

In order to evaluate if the optimisation of the N-terminal region (optimisation of the amino acid sequence next to the N-terminal methionine) leads to demethionylation of the protein, N-terminal sequencing has been done on the MC-011 construct carrying MAK amino acids on his N-terminal extremity.

The proteins were separated by SDS PAGE on a Novex 4%-20% polyacrylamide gel from Invitrogen, before transfer onto Problot PVDF (polyvinylidene difluoride) (Bio-Rad) membrane. The membrane was stained with amidoblack. The band of interest was then cut and analysis was carried out according to the manufacturer's protocol using an Applied Biosystems Procise sequencer system. Twelve cycles of Edman's degradation were performed.

The N-terminal amino acid sequence obtained is AKNDITLEDLP (SEQ ID NO:86), which corresponds to the N-terminal extremity of the protein starting at the amino acid number two after the initial methionine. This indicates that the mature protein is mainly demethionylated.

Example 7: UspA2 construct MC-001: Bactericidal activity.

Bactericidal Assay

Moraxella catarrhalis was cultivated overnight on Petri dish at 37°C + 5% CO₂. Bacteria were transferred in 12 ml HBSS-BSA (Hank's Buffered Salt Solution with Bovine Serum Album) 0.1% buffer in order to get an OD₆₂₀ of 0.650. Serum samples were heated for 45 min at 56°C to inactivate the endogenous complement. Serial two-fold dilutions of sera in SBA buffer (HBSS-BSA 0.1%) were added on a 96-well round bottom microtitre plate (25 µl/well). Subsequently, 50 µl of SBA buffer were added in each well. Then 25 µl of *Moraxella catarrhalis* strains at 4 10⁶ cfu/ml were added to the wells containing sera and incubated for 15 min at room temperature. Finally 25 µl of freshly thawed baby rabbit complement diluted 1/8 in HBSS-BSA 0.1% were

added to reach a final volume of 125 μ l. Plates were incubated for 1 h at 37 °C with orbital shaking (210 rpm). The reaction was stopped by laying the microplate on ice for at least 5 min.

After homogenization, various dilutions of the suspension (a mixture of bacteria, serum, complement and buffer, at a volume of 125 μ l as discussed in the previous paragraph) were added onto chocolate agar plates and incubated for 24 hours at 37°C with 5% CO₂ and *Moraxella catarrhalis* colonies were counted.

Eight wells without serum sample were used as bacterial controls to determine the number of *Moraxella catarrhalis* colonies per well. The mean number of CFU (colony forming unit) of the control wells was determined and used for the calculation of the killing activity for each serum sample. The bactericidal titers were expressed at the reciprocal dilution of serum inducing 50% of killing.

Anti-UspA2 antisera generated in mice, guinea pigs and rabbits against MC-001 were tested in the bactericidal assay described here above against 20 different *Moraxella catarrhalis* strains isolated from various tissues (blood, sputum, nose, middle ear fluids) in various countries (US, Finland, Netherlands, Norway, Sweden).

As shown below anti-UspA2 antibodies were able to induce a cross-bactericidal killing of *Moraxella catarrhalis*, whatever the percentage of homology of the UspA2 expressed by the tested strain. Moreover bactericidal activity was also shown against strains which only express UspA1 or the chimeric protein UspA2H. As expected, no or only weak bactericidal antibody titres were measured against UspA1 and UspA2 double knock-out mutants.

Table 8: Cross-bactericidal activity of anti-UspA2 MC-001 antibodies generated in mouse, guinea pig and rabbit. 1+2 KO is a double knock out, UspA1 & UspA2. 1KO is a UspA1 knockout only. MEF (AOM) = Middle Ear Fluid (Acute Otitis Media). "/" in the Isolate source column = not aware of the isolate source.

Strains		Isolate source	UspA gene present	Identity % * versus the vaccine sequence ATCC25238	Anti-UspA2 antiserum bactericidal activity		
					Mouse	Guinea pig	Rabbit
ATCC	25238	/	UspA1/UspA2	45.2/100	+++	++	+++
	43617	Bronchitis	UspA1	40.1	+	+	+
	2926	/	UspA1	36.4	-	+/	-
	2933	/	UspA1/UspA2	44.6/62.4	-	++	++
	2912	/	UspA1/UspA2	39.3/64.6	++	++	+++
Finnish	2908	/	UspA1/UspA2	43.3/52.8	+++	+++	+++
	307	MEF (AOM)	UspA1/UspA2	47.6/70.1	-	++	+++
	353	MEF (AOM)	UspA1/UspA2	45/61.8	+	++	++
	358	MEF (AOM)	UspA1/UspA2	47/61.5	+++	+++	+++
	216	MEF (AOM)	UspA1/UspA2	46.6/66.9	+++	+++	+++
Dutch	N9	nose	UspA1/UspA2H	41.1/70.1	++	+++	+++
	H2	sputum	UspA1/UspA2	47/61.6	+++	+++	+++
	F10	sputum	UspA1/UspA2	42.9/61.1	++	+++	+++
Norwegian	1	Tracheotomy (Pneumonia)	UspA1/UspA2	44.6/60.9	+++	+++	+++
	13	Tracheotomy (Pneumonia)	UspA1/UspA2	47.8/55	++	+++	+++
	20	Tracheotomy (Pneumonia)	UspA1/UspA2	33.7/50.6	+++	++	+++
	25	Tracheotomy (Pneumonia)	UspA1/UspA2	47.8/55	++	+++	+++
	27	Tracheotomy (Pneumonia)	UspA1/UspA2	46/76.1	++	+++	+++
Swedish	36	Tracheotomy (Pneumonia)	UspA1/UspA2	53.6/61.6	+	++	++
	BBH18 WT	sputum	UspA1/UspA2H	42.9/57.3	+++	+++	+++
	BBH18 (1+2KO)		-	-	+	-	-
	BBH18 (1KO)		UspA2H	57.3	+++	+++	+++
	RH4 WT	blood	UspA1/UspA2H	37.4/59.9	not done	++	++
	RH4 (1+2KO)		-	-	-	-	+

* determined using the software Gapt/ClustalX versus the ATCC25238 UspA2 fragment AA30-540.

+++ >50000
++ >20000
+ >500
- <200

Table 9: UspA Expression in the *M. catarrhalis* strains in Table 8.

Strains		UspA1 expression	UspA2 expression	UspA2H expression
ATCC	25238	Yes	Yes	No
	43617	Yes	stop codon	No
American	2926	Yes	No	stop codon
	2933	Yes	Yes	No
	2912	Yes	Yes	No
	2908	Yes	Yes	No
Finnish	307	Yes	Yes	No
	353	Yes	Yes	No
	358	Yes	Yes	No
	216	Yes	Yes	No
Dutch	N9	Yes	No	Yes
	H2	Yes	Yes	No
	F10	Yes	Yes	No
Norwegian	1	Yes	Yes	No
	13	Yes	Yes	No
	20	Yes	Yes	No
	25	Yes	Yes	No
	27	Yes	Yes	No
Swedish	36	Yes	Yes	No
	BBH18 WT	Yes	No	Yes
	BBH18 (1+2KO)	No	No	No
	BBH18 (1KO)	No	No	Yes
	RH4 WT	Yes	No	Yes
	RH4 (1+2KO)	No	No	No

Example 8: Protection in a mouse model of lung colonization (MC-001)

Five weeks-old female Balb/c mice (n=8/5 groups) were immunized by the intramuscular route at days 0, 14 and 28 with 50 µl of vaccine containing 10 µg of UspA2 construct MC-001 formulated within AS02V. Mice were intranasally challenged at day 42 with $5 \cdot 10^5$ CFU of various *Moraxella catarrhalis* strains. Bacteria were counted in lungs collected 0, 3 and 6 hours post-challenge. Differences between groups were analysed using the Dunnet test.

As summarised in Table 10, UspA2 construct MC-001 induced a significant protection against both homologous and heterologous strains, including the strain 43617 which does express UspA1 but not UspA2 and the BBH18 strain which expresses the chimeric protein UspA2H (constituted of the N-terminal sequence from UspA1 and the C-terminal sequence from UspA2).

Table 10: Protective efficacy of UspA2. MC-001 construct.

Strain	UspA expressed	Identity % * versus the vaccinal sequence ATCC25238	Log ₁₀ cfu / ml		p value
			Control group	Vaccine group	
25238	UspA1 & UspA2	45.2/100	5.2	3.1	0.01
43617	UspA1	40.1	4.9	3.6	0.01
F10	UspA1 & UspA2	42.9/61.1	4.3	3.9	0.25
F10	UspA1 & UspA2	42.9/61.1	4.4	3.6	0.01
BBH18	UspA1 & UspA2H	42.9/57.3	4.3	3.5	0.01
20	UspA1 & UspA2	33.7/60.6	4.4	3.9	0.02

* determined using the Gap/ClustalX software versus the ATCC25238 UspA2 fragment AA 30-540

p values in bold are significant (p < 0.05)

Example 9: UspA2 construct MC-007: Antibody bactericidal activity.**Bactericidal Assay**

Moraxella catarrhalis 25238 was cultivated overnight on Petri dish at 37°C + 5% CO₂. Bacteria were transferred in 12 ml HBSS-BSA 0.1% buffer in order to get an OD₆₂₀ of 0.650. Serum samples were heated for 45 min at 56°C to inactivate the endogenous complement. Serial two-fold dilutions of sera in SBA buffer (HBSS-BSA 0.1%) were added on a 96-well round bottom microtitre plate (25 µl/well). Subsequently, 50 µl of SBA buffer were added in each well. Then 25 µl of *Moraxella catarrhalis* 25238 strain at $4 \cdot 10^{14}$ cfu/ml were added to the wells containing sera

and incubated for 15 min at room temperature. Finally 25 μ l of freshly thawed baby rabbit complement diluted 1/8 in HBSS-BSA 0.1% were added to reach a final volume of 125 μ l. Plates were incubated for 1 h at 37 °C with orbital shaking (210 rpm). The reaction was stopped by laying the microplate on ice for at least 5 min. After homogenization, various dilutions of the suspension were added onto chocolate agar plates and incubated for 24 hours at 37°C with 5% CO₂ and *Moraxella* colonies were counted. Eight wells without serum sample were used as bacterial controls to determine the number of *Moraxella catarrhalis* colonies per well. The mean number of CFU of the control wells was determined and used for the calculation of the killing activity for each serum sample. The bactericidal titers were expressed at the reciprocal dilution of serum inducing 50% of killing.

Anti-UspA2 antisera generated in mice with UspA2 construct MC-001 or MC-007 were tested in a bactericidal assay using the protocol described above against the 25238 *Moraxella catarrhalis* homologous strain.

As shown in Table 11, the MC-007 UspA2 construct elicited a high bactericidal response, similar to that induced by the MC-001.

Table 11: Bactericidal activity of anti-UspA2 MC-001 and MC-007 antibodies. Normal mouse sera = sera from mice immunized with AS02V only, not with UspA2.

Samples	Bactericidal titers
Normal mouse sera (AS02V)	-
Mouse Anti-Killed whole cells 25238	++
Mouse anti-UspA2 against UspA2 MC-001	+++
Mouse anti-UspA2 against UspA2 MC-007	+++

Example 10: UspA2 construct MC-007: Protective efficacy in a lung challenge model**Protection in a mouse model of lung colonization**

Five weeks-old female Balb/c mice (8 mice per group, 5 groups max per time point) were immunized by the intramuscular route at days 0, 14 and 28 with 50 µl of vaccine containing 10 µg of UspA2 construct MC-001 formulated with AS02V or MC-007 formulated within AS02V. Mice were intranasally challenged at day 42 with $5 \cdot 10^5$ CFU of *Moraxella catarrhalis* strain ATCC(a US registered trademark) 25238™. Mice were immunized with 10 µg of killed whole cells from *Moraxella catarrhalis* strain ATCC(a US registered trademark) 25238™ (as positive control) (M.cat. WC 25238 in Figure 14) or with AS02V alone (as negative control). Bacteria were counted in lungs collected 0, 3 and 6 hours post-challenge. Differences between groups were analysed using the Dunnet test.

As shown in Figure 14, both UspA2 constructs were similarly protective against ATCC(a US registered trademark) strain 25238™.

Example 11: Immunogenicity of UspA2 MC-009 protein formulations in mice

Groups of 25 female Balb/c mice were immunized by the intramuscular (IM) route at days 0, 14 and 28 with 50 µl of the following formulations:

- MC-009 (1 µg) AlPO₄ (1000 µg/ml)
- MC-009 (1 µg) AS04C (AlPO₄/MPL 100/100 per ml)
- MC-009 (1 µg) AS01E (QS21/MPL 50/50 per ml)

Anti-IgG levels were determined in individual sera collected at days 28 (PII) and 42 (PIII) using the following protocol:

ELISA to measure anti-UspA2 antibodies.

Plates were coated overnight at 4°C with 100 µl per well of UspA2 construct MC-009 at 4 µg/ml in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 20 (polysorbate 20) 0.05%. After washing, serial two fold dilutions of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxidase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA Sigma P8787) and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH (pH) 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter).

The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

As shown in Figure 15, UspA2 induced high antibody levels with each adjuvant formulation.

Bactericidal Assay

The bactericidal assay was performed against *M. catarrhalis* strain (ATCC(a US registered trademark) 25238™) expressing a homologous full length UspA2 using the following protocol: *Moraxella catarrhalis* strain ATCC(a US registered trademark) 25238™ was cultivated overnight on Petri dish at 37°C + 5% CO₂. Bacteria were transferred in 10 ml of BHI (broth heart infusion) in order to get an OD₆₂₀ of 0.650. Serum samples were heated for 45 min at 56°C to inactivate the endogenous complement. Serial two-fold dilutions of sera in SBA buffer (HBSS-BSA 0.1%) were added on a 96-well round bottom microtitre plate (25 µl/well). Subsequently, 50 µl of SBA buffer were added in each well. Then 25 µl of *Moraxella catarrhalis* strain 25238™ at 4 10⁸ cfu/ml were added to the wells containing sera and incubated for 15 min at room temperature. Finally 25 µl of freshly thawed baby rabbit complement diluted 1/8 in HBSS-BSA 0.1% were added to reach a final volume of 125 µl. Plates were incubated for 1 h at 37 °C with orbital shaking (210 rpm). The reaction was stopped by laying the microplate on ice for at least 5 min. A 20 µl aliquot of each well of the

plate was then transferred into the corresponding well of a 96-well flat bottom microplate and 50 μ l of Mueller Hinton Broth–0.9% agar was added to each well. 50 μ l of PBS 0.9% agar was added as a second layer. After 3 hours at 37°C with 5% CO₂ plates were incubated overnight at 25°C, Moraxella colonies were counted using an automated image analysis system (KS 400, Zeiss, Oberkochen, Germany). Eight wells without serum sample were used as bacterial controls to determine the number of Moraxella per well. The mean number of CFU of the control wells was determined and used for the calculation of the killing activity for each serum sample. The bactericidal titers were expressed at the reciprocal dilution of serum inducing 50% of killing.

Figure 16 illustrates the bactericidal titers induced by UspA2 against a homologous strain. In this experiment, UspA2 induced high levels of bactericidal antibodies for each adjuvant formulation. Sera were tested at PIII; five pools of five sera samples were tested.

Example 12: Immunogenicity of UspA2 in combination with PD and PE-PilA NTHi antigens.

Immunization Protocol

Groups of 25 female Balb/c mice were immunized by the intramuscular (IM) route at days 0, 14 and 28 with 50 μ l of the following formulations:

- UspA2 construct MC-009 (1 μ g) AIPO4
- UspA2 construct MC-009 (1 μ g) AS04C
- UspA2 construct MC-009 (1 μ g) AS01E
- UspA2-PD-PEPilA (UspA2 construct MC-009, PEPilA construct LVL-735) AIPO4 (1 μ g of each of UspA2, PD and PEPilA; 1000 mg/ml AIPO4)
- UspA2-PD-PEPilA (UspA2 construct MC-009, PEPilA construct LVL-735) AS04C AIPO4 (1 μ g of each of UspA2, PD and PEPilA; 100/100 per ml AIPO4/MPL)
- UspA2-PD-PEPilA (UspA2 construct MC-009, PEPilA construct LVL-735) AS01E (1 μ g of each of UspA2, PD and PEPilA; 50/50 per ml QS21/MPL)

ELISA to measure anti-UspA2 antibodies

Anti-UspA2 IgG levels were determined in individual sera collected at days 28 and 42 using the following protocol.

Plates were coated overnight at 4°C with 100 µl per well of UspA2 construct MC-009 at 4 µg/ml in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 20 (polysorbate 20) 0.05%. After washing, serial two fold dilutions of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA Sigma P8787) and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH (pH) 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

ELISA to measure anti-PE antibodies

Plates were coated overnight at 4°C with 100 µl per well of 2 µg/ml of UspA2 in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

ELISA to measure anti-PilA antibodies

Plates were coated overnight at 4°C with 100 µl per well of 4 µg/ml of PilA in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in

PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 μ l per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 μ l of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 μ l/well) for 15 min in darkness. The reaction was stopped by addition of 50 μ l of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

ELISA to measure anti-PD antibodies

Plates were coated overnight at 4°C with 100 μ l per well of 8 μ g/ml of PD in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 μ l per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 μ l of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 μ l/well) for 15 min in darkness. The reaction was stopped by addition of 50 μ l of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

Bactericidal assay

Bactericidal titres were measured in pooled sera (5 pools/ group) collected at day 42 using the following protocol:

Moraxella catarrhalis was cultivated overnight on Petri dish at 37°C + 5% CO₂. Bacteria were transferred in 10 ml BHI (broth heart infusion) medium in order to get an OD₆₂₀ of 0.650. Serum samples were heated for 45 min at 56 °C to inactivate the endogenous complement. Serial two fold dilutions of sera in SBA buffer (HBSS-BSA 0.1%) were added on a 96-well round bottom microtitre plate (25 μ l/well). Subsequently, 50 μ l of SBA buffer were added in each well. Then 25 μ l of *Moraxella catarrhalis* strain 25238 at 4 10³ cfu/ml were added to the

wells containing sera and incubated for 15 min at room temperature. Finally 25 µl of freshly thawed baby rabbit complement diluted 1/8 in HBSS-BSA 0.1% were added to reach a final volume of 125 µl. Plates were incubated for 1 h at 37 °C with orbital shaking (210 rpm). The reaction was stopped by laying the microplate on ice for at least 5 min. A 20 µl aliquot of each well of the plate was then transferred into the corresponding well of a 96-well flat bottom microplate and 50 µl of Mueller Hinton Broth–0.9% agar was added to each well. 50 µl of PBS 0.9% agar was added as a second layer. After 3 hours at 37°C with 5% CO₂ plates were incubated overnight at 25°C. *Moraxella* colonies were counted using an automated image analysis system (KS 400, Zeiss, Oberkochen, Germany). Eight wells without serum sample were used as bacterial controls to determine the number of *Moraxella* per well. The mean number of CFU of the control wells was determined and used for the calculation of the killing activity for each serum sample. The bactericidal titers were expressed at the reciprocal dilution of serum inducing 50% of killing.

The bactericidal assay was performed against *Moraxella catarrhalis* strain 25238™, expressing a homologous UspA2.

A negative impact of the presence of PD and PE-PilA antigens on UspA2 IgG levels was observed in AS04C (post III) and AS01E (post II) formulations (Figure 17). However the impact remained limited (≤ 2 fold antibody decrease) and was not confirmed in the bactericidal assay (Figure 18). The IgG responses induced against PD, PE and PilA in mice by PE-PEPIL-A-UspA2 vaccine are shown in figure 19, figure 20 and figure 21, respectively.

Therefore, UspA2 was immunogenic when combined with PD and PE-PilA.

Example 13: UspA2 construct MC-009: Immunogenicity of PD and PE-PilA NTHi antigens in combination with UspA2 in mice

Immunization protocol

Groups of 25 female Balb/c mice were immunized by the intramuscular (IM) route at days 0, 14 and 28 with 50 µl of the following formulations:

- PD-PEPIL-A (1 µg of PD and 1 µg of PEPIL-A construct LVL-735) AS01E
- UspA2-PD-PEPIL-A (1 µg of UspA2 construct MC-009, PD and PEPIL-A construct LVL-735) AS01E

The ELISA IgG levels to PD, PE and PIIA were determined in individual sera collected at days 28 (PII) and 42 (PIII).

ELISA to measure anti-PE antibodies

Plates were coated overnight at 4°C with 100 µl per well of 2 µg/ml of UspA2 in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

ELISA to measure anti-PIIA antibodies

Plates were coated overnight at 4°C with 100 µl per well of 4 µg/ml of PIIA in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX (a US registered trademark) Pro software.

ELISA to measure anti-PD antibodies

Plates were coated overnight at 4°C with 100 µl per well of 8 µg/ml of PD in carbonate buffer pH 9.6. The plates were washed three times with NaCl 0.09% TWEEN(a US registered

trademark) 0.05%. After washing, serial two fold dilution of sera were added to microwells in PBS TWEEN(a US registered trademark) 20 0.05%. The plates were placed at room temperature for 30 minutes with shaking. After washing, anti-mouse IgG antibodies (Jackson 115-035-003) conjugated to peroxylase (100 µl per well) were added, and the plates were placed at room temperature for 30 minutes with shaking. Plates were washed as above and the solution of revelation (4 mg of OPDA and 5 µl of H₂O₂ in 10 ml of citrate 0.1M PH 4.5) was added to each well (100 µl/well) for 15 min in darkness. The reaction was stopped by addition of 50 µl of HCl 1N and the absorbance was read at 490 nm (620 nm for the reference filter). The titers were calculated by the 4-parameters method using the SOFTMAX(a US registered trademark) Pro software.

No major impact of the addition of UspA2 on PD and PEPilA immunogenicity in AS01E was observed as shown in figures 22, 23 and 24.

Example 14: Safety of a tetravalent vaccine formulation containing UspA2 in a mouse *Moraxella catarrhalis* lung inflammation model.

To mitigate the risk of inducing undesirable inflammatory responses in the lungs of COPD patients upon immunization with a candidate vaccine aiming at preventing the exacerbations due to Non-typeable *Haemophilus influenzae* (NTHi) and *Moraxella catarrhalis* (*M. cat.*), various animal models were developed and used to assess the safety of this vaccine. The formulation tested contained three NTHi antigens (PD, PE and PilA, with the two last ones combined as a PEPilA fusion protein), one *M. cat.* antigen (UspA2) and the Adjuvant System 01E (AS01E).

Two models were particularly dedicated to evaluate the safety of the UspA2 component of the vaccine.

Model 1:

- Objective

This model aimed at assessing the possible induction of undesirable immune responses in inflamed lungs upon vaccination.

• Study design

C57Bl/6 mice were sensitized by three intranasal administrations of 25 µg of heat-inactivated *M. cat.* strain ATCC(a US registered trademark) 25238™ whole cells (expressing an UspA2 which is 100% homologous to the vaccine UspA2) at days 0, 7 and 14. This treatment induced in the lungs a perivascular and peribronchiolar inflammation (with formation of lymphoid aggregates), alveolitis, pneumonitis, fibrosis and a strong *M.cat.* whole cell-specific IL-17⁺ CD4⁺ T cell response, which altogether mimicked the inflammatory process observed in the lungs of COPD patients (except emphysema).

The mice were then vaccinated at day 42 by the intramuscular route with 1/10th of human dose of the following formulations:

- PD 10 µg/ PEPiIA (LVL735 construct, described in WO2012/139225) 10 µg/ UspA2 (MC009 construct) 10 µg/ AS01_E
- PD 10 µg/ PEPiIA (LVL735 construct) 10 µg/ UspA2 (MC009 construct) 3.3 µg/ AS01_E
- AS01_E (negative control)
- PBS (negative control)

To assess the impact of these formulations on the sensitization-induced lung inflammation:

- Mice were daily monitored from day 43 to day 49 to look at mortality and any clinical signs indicating the induction of adverse events (prostration, piloerection, hunched position).
- A histological analysis of the lungs was performed at days 2, 7 and 14 post-vaccination (with 5 mice per group and time-point) to look at a possible aggravation of the inflammation.
- The induction of potentially undesirable T cell responses was evaluated on pools of lungs collected at days 7 and 14 post-vaccination (with 4 pools/group/time-point and the lungs of 3 mice per pool). The lung T cells were re-stimulated overnight either with UspA2 peptides, heat-inactivated *M. cat.* whole cells (VVC) or medium (as a negative control) and then analyzed by flow cytometry for the expression of CD5, CD4, CD8, IL-17, IL-13, TNFα and IFNγ.

• Results

- No mortality or adverse event was reported.
- Lung histology (Figures 25 to 29):
 - o The alterations observed in the lungs were similar in severity in all groups and characterized by slight to moderate perivascular/bronchiolar mononuclear cell infiltrates.
 - o No alveolitis and/or pneumonitis related to vaccination were observed.
- T cell response:
 - o Strong CD4⁺ T cell responses (mainly IL-17 and TNF α producing cells) were measured in the lungs upon re-stimulation with WC, but regardless of the formulation administered (vaccines or adjuvant alone or PBS) (Figures 30 to 33). Low or no lung CD8⁺ T cell responses were observed (data not shown).
 - o No detectable T cell response was re-stimulated by UspA2 peptides, whatever the group, indicating that no UspA2-specific response was primed or boosted post-vaccination (data not shown).

Model 2:

• Objective

This model aimed at assessing the possible induction of undesirable immune responses in inflamed lungs upon vaccination and *M. cat.* challenge.

• Study design

C57Bl/6 mice were successively:

- Sensitized by three intranasal administrations of 25 μ g of heat-inactivated *M. cat.* strain 25238 WC (expressing an UspA2 which is 100% homologous to the vaccine UspA2) at days 0, 7 and 14 (as in Model 1).
- Vaccinated at day 42 by the intramuscular route with 1/10th of human dose of the following formulations (as in Model 1):
 - PD (10 μ g/ PEPiA (LVL735 construct) 10 μ g/ UspA2 (MC009 construct) 10 μ g/ AS01_E

- PD 10 µg/ PEPiIA (LVL735 construct) 10 µg/ UspA2 (MC009 construct) 3.3 µg/ AS01_E
- AS01_E (negative control)
- PBS (negative control)
- Challenged by one intranasal administration of 25 µg of heat-inactivated *M. cat.* strain F10 WC (expressing an UspA2 which shares 53% homology with the vaccine UspA2) or by one intranasal administration of PBS as a control, both at day 56. The challenge strain was different from the sensitization strain to mimic the situation observed in COPD patients who experience new exacerbations due to newly acquired *M. cat.* strains.

To assess the impact of vaccination and challenge on the sensitization-induced lung inflammation:

- Mice were daily monitored from day 43 to day 63 to look at mortality and any clinical signs indicating the induction of adverse events (prostration, piloerection, hunched position).
- The induction of potentially undesirable T cell responses was evaluated on pools of lungs collected at days 7 and 14 post-challenge (with 4 pools/group/time-point and the lungs of 3 mice per pool). The lung T cells were re-stimulated overnight either with UspA2 peptides, heat-inactivated *M. cat.* F10 WC or medium (as a negative control) and then analyzed by flow cytometry for the expression of CD5, CD4, CD8, IL-17, IL-13, TNF α and IFN γ .

• Results

- No mortality or adverse event was reported.
- T cell response:
 - Strong post-challenge CD4⁺ T cell responses (mainly IL-17 and TNF α producing cells) were measured in the lungs upon re-stimulation with F10 WC, regardless of the formulation administered (vaccines or adjuvant alone or PBS) (Figures 34 to 37). Not surprisingly, these responses were higher in mice challenged with inactivated bacteria than in mice challenged with PBS. Whatever the challenge, low or no lung CD8⁺ T cell responses were observed (data not shown).
 - No detectable T cell response was re-stimulated by UspA2 peptides, whatever the group, indicating that no UspA2-specific response was primed or boosted post-challenge (data not shown).

Conclusion

The PD/ PEPiIA/ UspA2/ AS01_E formulations tested and more specifically the UspA2 component of these vaccines were shown safe in a mouse *M. cat.* lung inflammation model.

CLAIMS

1. A protein consisting of formula I:

$A - (R_1)_m - (B)_n$ (formula I)

wherein:

A is an immunogenic fragment of UspA2 from *Moraxella catarrhalis* having at least 90% identity to SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42 or SEQ ID NO: 43;

R_1 is an amino acid;

m is 0 or 2;

B is histidine; and

n is 1, 2, 3, 4, 5 or 6.

2. The protein according to claim 1 wherein $(R_1)_m$ is AS (alanine serine).
3. A protein according to any of claims 1-2 further comprising a methionine at the amino terminal.
4. The protein according to any of claims 1-3 wherein A is an immunogenic fragment of UspA2 selected from the group consisting of amino acids 30-540 of SEQ ID NO: 1 (SEQ ID NO: 39), amino acids 31-540 of SEQ ID NO: 1 (SEQ ID NO: 40), amino acids 30-519 of SEQ ID NO: 1 (SEQ ID NO: 41), amino acids 30-564 of SEQ ID NO: 1 (SEQ ID NO: 42) and amino acids 31-564 of SEQ ID NO: 1 (SEQ ID NO: 43).
5. The protein according to any of claims 1-3 wherein A is an immunogenic fragment of UspA2 of SEQ ID NO: 43 or a sequence having at least 90%, 95%, 96%, 97%, 98%, 99% sequence identity to SEQ ID NO: 43.
6. The protein according to any of claims 1-5 wherein the protein is selected from the group consisting of SEQ ID NO: 53, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, SEQ ID NO: 73 and SEQ ID NO: 88.
7. The protein according to any of claims 1-6 wherein the protein is SEQ ID NO: 69 (MC-009).
8. An immunogenic composition comprising a protein of formula (I) as defined in any of claims 1-7.
9. The immunogenic composition of claim 8 further comprising at least one antigen from *Haemophilus influenzae*, optionally Protein D.

10. The immunogenic composition of claim 8-9 further comprising *Haemophilus influenzae* antigens Protein E, PilA or Protein E and PilA as a fusion protein.
11. A vaccine comprising the protein of any of claims 1-7 or immunogenic composition of any of claims 8-10 and optionally further comprising an adjuvant, optionally AS01E.
12. The vaccine of claim 11 wherein the immunogenic composition contains a protein of SEQ ID NO: 69, Protein D and a PE-PilA fusion protein, optionally LVL-735.
13. A process for preparing the immunogenic composition of claims 8-10 or the vaccine of claims 11 and 12 comprising combining a protein of formula (I) with a pharmaceutically acceptable excipient.