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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 July 2011 (07.07.2011)

PCT

(10) International Publication Number
WO 2011/080595 A2

(51) International Patent Classification:
A61K 47/48 (2006.01) *A61K 39/02* (2006.01)

(21) International Application Number:
PCT/IB2010/003484

(22) International Filing Date:
30 December 2010 (30.12.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/291,262 30 December 2009 (30.12.2009) US

(71) Applicant *(for all designated States except US)*: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35, CH-4056 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **COSTANTINO, Paolo** [RU/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **SERINO, Laura** [IT/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **ROMANO, Maria, Rosaria** [IT/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **TONTINI, Marta** [IT/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **BERTI, Francesco** [IT/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **MICOLI, Francesca** [IT/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentina 1, I-53100 Siena (IT). **KABANOVA, Anna** [RU/IT]; c/o Novartis Vaccines And Diagnostics Srl, Via Fiorentino 1, I-53100 Siena (IT).

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

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

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/080595 A2

(54) **Title**: POLYSACCHARIDE IMMUNOGENS CONJUGATED TO E. COLI CARRIER PROTEINS

(57) **Abstract**: *E. coli* proteins have been identified that are useful as carrier proteins to improve a response to a polysaccharide immunogen conjugated to such protein. In particular, AcfD precursor protein (orf3526 polypeptide), Flu antigen 43 protein (orf1364 polypeptide), and Sell repeat-containing protein (upec-5211 polypeptide) have been shown to be effective. Additionally, these *E. coli* proteins can enhance the immune response to glucans, particularly fungal glucans.

POLYSACCHARIDE IMMUNOGENS CONJUGATED TO E. COLI CARRIER PROTEINS

TECHNICAL FIELD

This invention relates to immunization using polysaccharide immunogens conjugated to *E. coli* carrier polypeptides. Of particular interest is use of such compositions as vaccines against bacterial and fungal infections and diseases.

BACKGROUND ART

Carrier proteins are used to improve the immune response to polysaccharide immunogens. Such carrier proteins can be particularly advantageous in the induction of an immune response in the very young and are therefore found in a number of pediatric vaccines. The recommended pediatric immunization schedule includes a significant number of vaccines including hepatitis B vaccine at birth; starting at six weeks, all of diphtheria/tetanus/pertussis (DTaP), rotavirus, *H. influenzae* type b (Hib) conjugate, inactivated poliovirus and pneumococcal conjugates; starting at six months, inactivated influenza vaccines; starting at 12 months, measles/mumps/rubella (MMR), varicella, and hepatitis A; and after two years, meningococcal conjugate. Among this list, the following are polysaccharide conjugates: Hib conjugate (e.g., HbOC – a diphtheria CRM_{197} conjugate); pneumococcal conjugates (e.g., Prevnar – a diphtheria CRM_{197} conjugate and Synflorix – a protein carrier derived from non-typeable Haemophilus influenzae strains); and meningococcal conjugate (e.g., Menactra – a diphtheria CRM_{197} conjugate).

Adding new vaccines to the current pediatric immunization schedule can encounter two potential problems that must be addressed. First, the issue of carrier-induced epitopic suppression (or "carrier suppression", as it is generally known) must be addressed, particularly suppression arising from carrier priming. "Carrier suppression" is the phenomenon whereby pre-immunization of an animal with a carrier protein prevents it from later eliciting an immune response against a new antigenic epitope that is presented on that carrier (Herzenberg *et al.* (1980) *Nature* 285: 664-667).

As reported in Schutze *et al.* (1985) *J Immunol* 135:2319-2322, where several vaccine antigens contain the same protein component (being used as an immunogen and/or as a carrier protein in a conjugate) then there is the potential for interference between those antigens. Schutze *et al.* observed that the immune response against an antigen that was conjugated to a tetanus toxoid (Tt) carrier was suppressed by pre-existing immunity against Tt.

Dagan *et al.* observed that a combination of DTP vaccines with a Hib conjugate vaccine was adversely affected where the carrier for the Hib conjugate was the same as the tetanus antigen

from the DTP vaccine ((1998) *Infect Immun* 66:2093-2098). Dagan *et al.* concluded that this "carrier suppression" phenomenon, arising from interference by a common protein carrier, should be taken into account when introducing vaccines that include multiple conjugates.

In contrast to Schutze *et al.* and Dagan *et al.*, Barington *et al.* reported that priming with tetanus toxoid had no negative impact on the immune response against a subsequently-administered Hib-Tt conjugate, but suppression was seen in patients with maternally acquired anti-Tt antibodies ((1994) *Infect Immun* 62:9-14). Di John *et al.*, however, observed an "epitopic suppression" effect for a Tt-based peptide conjugate in patients having existing anti-Tt antibodies resulting from tetanus vaccination ((1989) *Lancet* 2(8677):1415-8).

Granoff *et al.* suggested that a conjugate having CRM_{197} (a detoxified mutant of diphtheria toxin) as the carrier may be ineffective in children that had not previously received diphtheria toxin as part of a vaccine (*e.g.*, as part of a DTP or DT vaccine) ((1993) *Vaccine Suppl* 1: S46-51). This work was further developed in Granoff *et al.* (1994) *JAMA* 272:1116-1121, where a carrier priming effect by D-T immunization was seen to persist for subsequent immunization with Hib conjugates.

In Barington *et al.* (1993) *Infect Immun* 61:432-438, the authors found that pre-immunization with a diphtheria or tetanus toxoid carrier protein reduced the increase in anti-Hib antibody levels after a subsequent immunization with the Hib capsular saccharide conjugated to those carriers, with IgG1 and IgG2 being equally affected. Responses to the carrier portions of the conjugates were also suppressed. Furthermore, a more general non-epitope-specific suppression was seen, as pre-immunization with one conjugate was seen to affect immune responses against both the carrier and saccharide portions of a second conjugate that was administered four weeks later.

Thus, given the confusion over the impact of "carrier suppression," having additional carrier proteins available for conjugation will be beneficial to reduce such adverse interactions.

Second, given the already crowded immunization schedule, addition of new vaccines to the immunization schedule will become increasingly difficult due to possible adverse interactions, but also due simply to the number of separate injections required. Thus, being able to combine vaccines into a single injection such as the DTaP or MMR vaccines is advantageous. Having additional carrier proteins that can enhance an immune response to a polysaccharide immunogen as well as induce an immune response to itself will be beneficial as it can allow combination of vaccines against different pathogens into a single injectable composition.

It is an object of the invention to provide further and/or better carrier polypeptides for conjugation to polysaccharide immunogens. It is also an object of the invention to provide carrier polypeptides for conjugation to polysaccharide immunogens where the carrier

polypeptides can be used in immunization against pathogenic *E. coli* strains, and more particularly against intestinal pathotypes (*e.g.* EAEC, EIEC, EPEC and ETEC strains) as well as ExPEC pathotypes. It is also an object of the invention to provide conjugates with such further and/or better carrier polypeptides where the polysaccharide immunogen is a glucan polysaccharide, which in some embodiments can be used in immunization against fungal pathogens.

SUMMARY

Accordingly, one aspect of the invention provides a glucan polysaccharide conjugate comprising a glucan polysaccharide conjugated to a carrier polypeptide selected from the group consisting of an *E. coli* AcfD precursor protein (orf3526 polypeptide), an *E. coli* Flu antigen 43 protein (orf1364 polypeptide), and an *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide). When the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NO 50; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NO: 50; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 50; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NO: 50; and/or (e) when aligned with SEQ ID NO: 50 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. When the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NO 44; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NO: 44; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 44; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NO: 44; and/or (e) when aligned with SEQ ID NO: 44 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. When the carrier polypeptide is the *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NO 48; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NO: 48; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 48; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NO: 48; and/or (e) when aligned with SEQ ID NO: 48 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. In certain

embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide contains β-1,3-linkages and/or β-1,6-linkages. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide is a single molecular species. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide is conjugated to the carrier protein directly or is conjugated to the carrier protein via a linker. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide has a molecular weight of less than 100 kDa (e.g. less than 80, 70, 60, 50, 40, 30, 25, 20, or 15 kDa). In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide has 60 or fewer (*e.g.*, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, or 4) glucose monosaccharide units. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide is a β-1,3 glucan polysaccharide with some β-1,6 branching. In certain embodiments where the polysaccharide is a β-1,3 glucan polysaccharide with some β-1,6 branching, the glucan polysaccharide is a laminarin. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide comprises both β-1,3-linked glucose residues and β-1,6-linked glucose residues, with a ratio of β 1,3 linked glucose residues to β-1,6-linked residues of at least 8:1 and/or there are one or more sequences of at least five adjacent non-terminal residues linked to other residues only by β-1,3 linkages. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide comprises both β-1,3-linked glucose residues and β-1,6-linked glucose residues, with a ratio of β-1,3 linked glucose residues to β-1,6-linked residues of at least 8:1. In certain embodiments which may be combined with any of the preceding embodiments excluding those which have β-1,6-linked residues, the glucan polysaccharide has exclusively β-1,3 linkages. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide is a curdlan. In certain embodiments which may be combined with any of the preceding embodiments, the glucan polysaccharide conjugate further comprises an adjuvant.

Another aspect of the invention provides vaccine components comprising the glucan polysaccharide conjugate according to the preceding aspect in any of its embodiments.

Still another aspect of the invention provides vaccines comprising the vaccine component of the preceding aspect. In certain embodiments, the vaccine further comprises an adjuvant. In certain embodiments that may be combined with the preceding embodiment, the vaccine further comprises an additional vaccine component selected from: a *Neisseria meningitidis* antigen, a *Streptococcus pneumoniae* antigen, a *Streptococcus pyogenes* antigen, a *Moraxella catarrhalis* antigen, a *Bordetella pertussis* antigen, a *Staphylococcus aureus* antigen, a *Staphylococcus epidermis* antigen, a *Clostridium tetani* antigen, a *Cornynebacterium diphtheriae* antigen, a *Haemophilus influenzae* type B (Hib) antigen, a *Pseudomonas aeruginosa* antigen, a *Legionella pneumophila* antigen, a *Streptococcus agalactiae* antigen, a *Neiserria gonorrhoeae* antigen, a *Chlamydia trachomatis*

antigen, a *Treponema pallidum* antigen, a *Haemophilus ducreyi* antigen, an *Enterococcus faecalis* antigen, an *Enterococcus faecium* antigen, a *Helicobacter pylori* antigen, a *Staphylococcus saprophyticus* antigen, a *Yersinia enterocolitica* antigen, an additional *E. coli* antigen, a *Bacillus anthracis* antigen, a *Yersinia pestis* antigen, a *Mycobacterium tuberculosis* antigen, a Rickettsia antigen, a *Listeria monocytogenes* antigen, a *Chlamydia pneumoniae* antigen, a *Vibrio cholerae* antigen, a *Salmonella typhi* antigen, a *Borrelia burgdorferi* antigen, a *Porphyromonas gingivalis* antigen, a Shigella antigen and a Klebsiella antigen. In certain preferred embodiments that may be combined with the preceding embodiment, the vaccine further comprises an additional vaccine component selected from a bacteria associated with nosocomial infections, which can include: a *Staphylococcus aureus* antigen, a *Candida albicans* antigen, a *Clostridium difficile* antigen, or a *Pseudomonas aeruginosa* antigen.

Yet another aspect of the invention provides methods of inducing an enhanced immune response in a mammalian subject to a glucan polysaccharide comprising administering to the mammalian subject of the glucan polysaccharide conjugate according to the preceding aspect in any of its embodiments, the vaccine component according to the preceding aspect, or the vaccine according to the preceding aspect in any of its embodiments.

Another aspect of the invention provides uses of the glucan polysaccharide conjugate according to the preceding aspect in any of its embodiments, the vaccine component according to the preceding aspect, or the vaccine according to the preceding aspect in any of its embodiments to induce an enhanced immune response in a mammalian subject to the polysaccharide.

Still another aspect of the invention provides polysaccharide conjugates comprising a polysaccharide conjugated to a carrier polypeptide selected from the group consisting of an *E. coli* AcfD precursor protein (orf3526 polypeptide), an *E. coli* Flu antigen 43 protein (orf1364 polypeptide), and an *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide). In embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide), the carrier polypeptide may comprise a mutation reducing the toxicity and/or a deletion improving purification as compared to the *E. coli* AcfD precursor protein (orf3526 polypeptide) of SEQ ID NO: 39. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) with a mutation reducing the toxicity, the mutation may be selected from a deletion of all or a portion of the zincin metalloprotease domain and a point mutation in zincin metalloprotease domain which reduces the protease activity. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) with a mutation reducing the toxicity, the point mutation is a mutation of a zinc binding residue or a mutation of a catalytic residue, which in some cases may be amino acid number 1305 based upon alignment with SEQ ID NO: 39. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) with a mutation reducing the toxicity, the carrier polypeptide does not comprise at least the last 100 C-terminal amino

acids of the *E. coli* AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 300 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 400 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 500 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 600 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 700 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 750 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the last 758 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein or does not comprise at least the first 100 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 200 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 300 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 400 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 500 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 600 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 700 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 750 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the first 760 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide), the carrier polypeptide does not comprise at least the last 100 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 125 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 150 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 175 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 210 C-terminal amino acids of the E. coli AcfD (orf3526) protein, or at least the last 217 C-terminal amino acids of the E. coli AcfD (orf3526) protein and optionally do not comprise at least the first 10 N-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the first 20 N-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the first 30 N-terminal amino acids of the E. coli AcfD (orf3526) protein, or at least the first 33 N-terminal amino acids of the E. coli AcfD (orf3526) protein. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NOs 26-40; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NO: 26-40; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NOs 26-40; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NOs 26-40; and/or (e) when aligned with any of SEQ ID NOs: 26-40 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide), the carrier polypeptide further contains a deletion

relative to the E. coli AcfD (orf3526) protein which increases solubility of the carrier polypeptide as compared to the E. coli AcfD (orf3526) protein. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) with a deletion which increases the solubility, the deletion is removal of substantially all of the N-terminal amino acids up to the gly-ser region, removal of all or a part of the N-terminal proline-rich repeat, or both. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) with a deletion which increases the solubility, the deletion is removal of at least the first 10 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 20 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 30 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 33 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 40 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 50 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 60 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 70 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 80 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, at least the first 90 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, or at least the first 94 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein. In embodiments where the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), the carrier polypeptide may be a fragment of the *E. coli* Flu antigen 43 protein (orf1364 polypeptide) wherein the fragment contains a deletion relative to the full length *E. coli* Flu antigen 43 protein (orf1364 polypeptide) which deletion increases solubility of the fragment as compared to the full length protein. In certain embodiments where the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide) has a deletion which increases the solubility, the deletion comprises the carboxyl-terminal β-barrel domain or the carrier polypeptide corresponds to the amino acid sequence of SEQ ID NO:44. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide) has a deletion which increases the solubility, the fragment comprises less than 950 amino acids, less than 900 amino acids, less than 850 amino acids, less than 800 amino acids, less than 750 amino acids, less than 700 amino acids, less than 650 amino acids, less than 600 amino acids, less than 550 amino acids, less than 500 amino acids, less than 450 amino acids, less than 440 amino acids, or less than 430 amino acids of the flu antigen 43 (orf1364) protein. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NOs 1-22; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NOs 1-22; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NOs 1-22; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at

least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NOs 1-22; and/or (e) when aligned with any of SEQ ID NOs 1-22 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. In another embodiment which may be combined with the preceding embodiments where the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), the fragment does not comprise at least the first 10 N-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 20 N-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 30 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 40 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 50 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), or at least the first 52 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein, and/or the fragment does not comprise at least the last 50 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 100 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 150 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 200 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 250 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 300 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 325 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), or at least the last 328 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide). In embodiments where the carrier polypeptide is the *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide), the carrier polypeptide may: (a) have the amino acid sequence of SEQ ID NOs 23-25; (b) have an amino acid sequence having from 1 to 10 single amino acid alterations compared to SEQ ID NOs 23-25; (c) have at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NOs 23-25; (d) comprise a fragment of at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 125, or at least 150 consecutive amino acids from SEQ ID NOs 23-25; and/or when aligned with any of SEQ ID NOs: 23-25 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75. In certain embodiments which may be combined with any of the preceding embodiments, the polysaccharide may be: (a) a glucan, (b) a capsular saccharide from at least one of serogroups A, C, W135 and Y of *Neisseria meningitidis*, (c) a saccharide antigen from *Streptococcus pneumoniae*, (d) a capsular polysaccharide from *Staphylococcus aureus*, (e) a *Haemophilus influenzae* B polysaccharide, (f) a saccharide antigen from *Streptococcus aglactiae*, (g) a lipopolysaccharide from *Vibrio cholerae*, or (h) a capsular polysaccharide from *Salmonella typhi*. In certain embodiments which may be combined with any of the preceding embodiments, the polysaccharide has a molecular weight of less than 100 kDa (e.g.

less than 80, 70, 60, 50, 40, 30, 25, 20, or 15 kDa). In certain embodiments which may be combined with any of the preceding embodiments, the polysaccharide has 60 or fewer (*e.g.,* 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, or 4) monosaccharide units. In certain embodiments which may be combined with any of the preceding embodiments, the polysaccharide may be directly or indirectly conjugated to the carrier protein. In certain embodiments which may be combined with any of the preceding embodiments, the polysaccharide conjugate further comprises an adjuvant.

Another aspect of the invention provides vaccine components comprising the polysaccharide conjugate according to the preceding aspect in any of its embodiments.

Still another aspect of the invention provides vaccines comprising the vaccine component of the preceding aspect. In certain embodiments, the vaccine further comprises an adjuvant. In certain embodiments that may be combined with the preceding embodiment, the vaccine further comprises an additional vaccine component selected from: a *Neisseria meningitidis* antigen, a *Streptococcus pneumoniae* antigen, a *Streptococcus pyogenes* antigen, a *Moraxella catarrhalis* antigen, a *Bordetella pertussis* antigen, a *Staphylococcus aureus* antigen, a *Staphylococcus epidermis* antigen, a *Clostridium tetani* antigen, a *Cornynebacterium diphtheriae* antigen, a *Haemophilus influenzae* type B (Hib) antigen, a *Pseudomonas aeruginosa* antigen, a *Legionella pneumophila* antigen, a *Streptococcus agalactiae* antigen, a *Neiserria gonorrhoeae* antigen, a *Chlamydia trachomatis* antigen, a *Treponema pallidum* antigen, a *Haemophilus ducreyi* antigen, an *Enterococcus faecalis* antigen, an *Enterococcus faecium* antigen, a *Helicobacter pylori* antigen, a *Staphylococcus saprophyticus* antigen, a *Yersinia enterocolitica* antigen, an additional *E. coli* antigen, a *Bacillus anthracis* antigen, a *Yersinia pestis* antigen, a *Mycobacterium tuberculosis* antigen, a Rickettsia antigen, a *Listeria monocytogenes* antigen, a *Chlamydia pneumoniae* antigen, a *Vibrio cholerae* antigen, a *Salmonella typhi* antigen, a *Borrelia burgdorferi* antigen, a *Porphyromonas gingivalis* antigen, Shigella antigen and a Klebsiella antigen. In certain preferred embodiments that may be combined with the preceding embodiment, the vaccine further comprises an additional vaccine component selected from a bacteria associated with nosocomial infections, which can include: a *Staphylococcus aureus* antigen, a *Candida albicans* antigen, a *Clostridium difficile* antigen, or a *Pseudomonas aeruginosa* antigen.

Yet another aspect of the invention provides methods of inducing an enhanced immune response in a mammalian subject to a polysaccharide comprising administering to the mammalian subject of the polysaccharide conjugate according to the preceding aspect in any of its embodiments, the vaccine component according to the preceding aspect, or the vaccine according to the preceding aspect in any of its embodiments.

Another aspect of the invention provides uses of the polysaccharide conjugate according to the preceding aspect in any of its embodiments, the vaccine component according to the preceding

aspect, or the vaccine according to the preceding aspect in any of its embodiments to induce an enhanced immune response in a mammalian subject to the polysaccharide.

DETAILED DESCRIPTION

Pure polysaccharides are often poor immunogens and therefore need to be conjugated to a carrier polypeptide. Even where a polysaccharide has sufficient immunogenicity, conjugation to a carrier protein can enhance the immunogenicity so that less polysaccharide need be delivered. Furthermore, for protective efficacy in the very young, conjugation to a carrier polypeptides is often required. The use of conjugation to carrier proteins in order to enhance the immunogenicity of carbohydrate antigens is well known (see, *e.g.*, Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36, Buttery & Moxon (2000) *J R Coll Physicians Lond* 34: 163-8, Ahmad & Chapnick (1999) *Infect Dis Clin North Am* 13: 113-33, vii, Goldblatt (1998) *J. Med. Microbiol.* 47:563-567, EP-B-0477508, US Pat. No. 5,306,492, WO98/42721, Dick *et al.* in *Conjugate Vaccines* (eds. Cruse *et al.*) Karger, Basel, 1989, Vol. 10, 48-1 14, Hermanson *Bioconjugate Techniques*, Academic Press, San Diego CA (1996), *etc.*); and is used in particular for pediatric vaccines (Ramsay *et al.* (2001) *Lancet* 357(9251):195-6). The inventors have surprisingly found that three *E. coli* antigens can act as carrier polypeptides: accessory colonization factor D (AcfD) precursor protein (orf3526 polypeptide), Flu antigen 43 protein (orf1364 polypeptide), and Sell repeat-containing protein (upec-5211 polypeptide).

An aspect of the invention therefore provides a polysaccharide conjugate comprising a polysaccharide conjugated to a carrier protein and optionally, an adjuvant (as defined below).

The carrier polypeptide may be covalently conjugated to the polysaccharide directly or via a linker. Any suitable conjugation reaction can be used, with any suitable linker where necessary.

Attachment of the polysaccharide to the carrier polypeptide is preferably via a $-NH_2$ group, *e.g.*, through the side chain(s) of a lysine residue(s) or arginine residue(s) in the carrier polypeptide. Where the polysaccharide has a free aldehyde group, this group can react with an amine in the carrier polypeptide to form a conjugate by reductive amination. Attachment to the carrier may also be via a -SH group, *e.g.*, through the side chain(s) of a cysteine residue(s) in the carrier polypeptide. Alternatively the polysaccharide may be attached to the carrier protein via a linker molecule.

The polysaccharide will typically be activated or functionalized prior to conjugation. Activation may involve, for example, cyanylating reagents such as CDAP (l-cyano-4-dimethylamino pyridinium tetrafluoroborate). Other suitable techniques use carbodiimides, hydrazides, active esters, norborane, p-nitrobenzoic acid, N-hydroxysuccinimide, S-NHS, EDC, TSTU (see, *e.g.*, the introduction to WO98/42721).

Direct linkages to the carrier polypeptide may comprise oxidation of the polysaccharide followed by reductive amination with the carrier polypeptide, as described in, for example, U.S. Pat No. 4,761,283 and U.S. Pat No. 4,356,170.

Linkages via a linker group may be made using any known procedure, for example, the procedures described in U.S. Pat No. 4,882,317 and U.S. Pat No. 4,695,624. Typically, the linker is attached via an anomeric carbon of the poylsaccharide. A preferred type of linkage is an adipic acid linker, which may be formed by coupling a free -NH2 group (*e.g.*, introduced to a polysaccharide by amination) with adipic acid (using, for example, diimide activation), and then coupling a protein to the resulting saccharide-adipic acid intermediate (see, *e.g.*, EP-B-0477508, *Mol. Immunol*, (1985) 22, 907-919, and EP-A-0208375). A similar preferred type of linkage is a glutaric acid linker, which may be formed by coupling a free -NH group with glutaric acid in the same way. Adipid and glutaric acid linkers may also be formed by direct coupling to the polysaccharide, *i.e.*, without prior introduction of a free group, *e.g.*, a free -NH group, to the polysaccharide, followed by coupling a protein to the resulting saccharide-adipic/glutaric acid intermediate. Another preferred type of linkage is a carbonyl linker, which may be formed by reaction of a free hydroxyl group of a modified polysaccharide with CDI (Bethell G.S. *et al.* (1979) *J Biol Chem* 254, 2572-4 and Hearn M.T.W. (1981) *J. Chromatogr* 218, 509-18); followed by reaction with a protein to form a carbamate linkage. Other linkers include β-propionamido (WO00/10599), nitrophenyl-ethylamine (Gever *et al.* (1979) *Med Microbiol Immunol* 165, 171-288), haloacyl halides (U.S. Pat. No. 4,057,685), glycosidic linkages (U.S. Pat. Nos. 4,673,574; 4,761,283; and 4,808,700), 6-aminocaproic acid (U.S. Pat. No. 4,459,286), N-succinimidyl-3-(2-pyridyldithio)-propionate (SPDP) (U.S. Pat. No. 5,204,098), adipic acid dihydrazide (ADH) (U.S. Pat. No. 4,965,338), C4 to C12 moieties (U.S. Pat. No. 4,663, 160), *etc.* Carbodiimide condensation can also be used (WO2007/000343).

A bifunctional linker may be used to provide a first group for coupling to an amine group in the polysaccharide (*e.g.*, introduced to the polysaccharide by amination) and a second group for coupling to the carrier (typically for coupling to an amine in the carrier). Alternatively, the first group is capable of direct coupling to the polysaccharide, *i.e.*, without prior introduction of a group, *e.g.*, an amine group, to the polysaccharide.

In some embodiments, the first group in the bifunctional linker is thus able to react with an amine group (-NH2) on the polysaccharide. This reaction will typically involve an electrophilic substitution of the amine's hydrogen. In other embodiments, the first group in the bifunctional linker is able to react directly with the polysaccharide. In both sets of embodiments, the second group in the bifunctional linker is typically able to react with an amine group on the carrier polypeptide. This reaction will again typically involve an electrophilic substitution of the amine.

Where the reactions with both the polysaccharide and the carrier protein involve amines then it is preferred to use a bifunctional linker. For example, a homobifunctional linker of the formula X-L-X, may be used where: the two X groups are the same as each other and can react with the amines; and where L is a linking moiety in the linker. Similarly, a heterobifunctional linker of the formula X-L-X may be used, where: the two X groups are different and can react with the amines; and where L is a linking moiety in the linker. A preferred X group is N-oxysuccinimide. L preferably has formula

L'-L^2-L', where L' is carbonyl. Preferred L^2 groups are straight chain alkyls with 1 to 10 carbon atoms (*e.g.*, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}) *e.g.* -$(CH_2)_4$- or -$(CH_2)_3$-.

Other X groups for use in the bifunctional linkers described in the preceding paragraph are those which form esters when combined with HO-L-OH, such as norborane, p-nitrobenzoic acid, and sulfo-N- hydroxysuccinimide.

Further bifunctional linkers for use with the invention include acryloyl halides (*e.g.*, chloride) and haloacylhalides.

The linker will generally be added in molar excess to polysaccharide during coupling to the polysaccharide.

Conjugates may have excess carrier (w/w) or excess polysaccharide (w/w), *e.g.*, in the ratio range of 1:5 to 5:1. Conjugates with excess carrier protein are typical, *e.g.*, in the range 0.2:1 to 0.9:1, or equal weights. The conjugate may include small amounts of free (*i.e.*, unconjugated) carrier. When a given carrier protein is present in both free and conjugated form in a composition of the invention, the unconjugated form is preferably no more than 5% of the total amount of the carrier protein in the composition as a whole, and more preferably present at less than 2% (by weight).

The composition may also comprise free carrier protein as immunogen (WO96/40242).

After conjugation, free and conjugated polysaccharides can be separated. There are many suitable methods, *e.g.*, hydrophobic chromatography, tangential ultrafiltration, diafiltration, *etc.* (see also Lei *et al.* (2000) *Dev Biol (Basel)* 103:259-264 and WO00/38711). Tangential flow ultrafiltration is preferred.

The polysaccharide moiety in the conjugate is preferably a low molecular weight polysaccharide, as defined below (see section on glucans). Oligosaccharides will typically be sized prior to conjugation.

The protein-polysaccharide conjugate is preferably soluble in water and/or in a physiological buffer.

For some polysaccharides, the immunogenicity may be improved if there is a spacer between the polysaccharide and the carrier protein. In this context, a "spacer" is a moiety that is longer than a single covalent bond. This spacer may be a linker, as described above. Alternatively, it may be a moiety covalently bonded between the polysaccharide and a linker. Typically, the moiety will be covalently bonded to the polysaccharide prior to coupling to the linker or carrier. For example, the spacer may be moiety Y, wherein Y comprises a straight chain alkyl with 1 to 10 carbon atoms (*e.g.*, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}), typically 1 to 6 carbon atoms (*e.g.*, C_1, C_2, C_3, C_4, C_5, C_6). The inventors have found that a straight chain alkyl with 6 carbon atoms (*i.e.*, -$(CH_2)_6$) is particularly suitable, and may provide greater immunogenicity than shorter chains (*e.g.*, -$(CH_2)_2$). Typically, Y is attached to the anomeric carbon of the polysaccharide, usually via an –O– linkage. However, Y may be linked to other parts of the polysaccharide and/or via other linkages. The other end of Y is bonded to the linker by any suitable linkage. Typically, Y terminates with an amine group to facilitate

linkage to a bifunctional linker as described above. In these embodiments, Y is therefore bonded to the linker by an -NH- linkage. Accordingly, a conjugate with the following structure is specifically envisaged for use in the present invention: wherein n+2 is in the range of 2-60, *e.g.*, between 10-50 or between 2-40. Preferably, n+2 is in the range of 25-30 or 11-19, *e.g.*, 13-17. The inventors have found that n+2 = 15 is suitable. Y is as described above.

In one aspect, the invention provides a method for making a polysaccharide conjugated to a carrier protein, wherein the step of conjugation is carried out in a phosphate buffer with >10mM phosphate; and to a conjugate obtained by this method. The inventors have found that sodium phosphate is a suitable form of phosphate for the buffer. The pH of the buffer may be adjusted to between 7.0-7.5, particularly 7.2. The step of conjugation is typically carried out in a phosphate buffer with between 20-200 mM phosphate, *e.g.*, 50-150 mM. In particular, the inventors have found that a phosphate buffer with 90-110 mM, *e.g.*, about 100 mM, phosphate is suitable. The step of conjugation is usually carried out at room temperature. Similarly, the step of conjugation is usually carried out at room pressure. Typically, the polysaccharide is attached to a linker as described above prior to the step of conjugation. In particular, the polysaccharide may be attached to a bifunctional linker as described above. The free end of the linker may comprise a group to facilitate conjugation to the carrier protein. For example, the inventors have found that the free end of the linker may comprise an ester group, *e.g.*, an N-hydroxysuccinimide ester group.

The polysaccharide conjugates disclosed herein can further include a pharmaceutically acceptable carrier.

The polysaccharide conjugates disclosed herein can further include an adjuvant. The adjuvant can comprise one or more of the adjuvants described below.

The polysaccharide conjugates may also be used in methods for raising an immune response in a mammal (or avian), comprising administering to the mammal (or avian) a composition of the invention.

Polysaccharide Immunogens

Any polysaccharide capable of inducing an immune response in a mammal or avian (either alone or conjugated to a carrier protein) may be used in the polysaccharide conjugates as disclosed herein (*i.e.*, a polysaccharide immunogen). Preferably, the polysaccharide is capably of inducing an immune response against a pathogen of interest. The polysaccharide may be branched or linear.

Low molecular weight polysaccharides may be used, particularly those with a molecular weight of less than 100 kDa (*e.g.*, less than 80, 70, 60, 50, 40, 30, 25, 20, or 15 kDa). It is also possible to use oligosaccharides containing, for example, 60 or fewer (*e.g.*, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, or 4) monosaccharide

units. Within this range, oligosaccharides with between 10 and 50 or between 20 and 40 monosaccharide units are preferred.

Exemplary polysaccharide immunogens are detailed below.

Glucan polysaccharides: Glucans are glucose-containing polysaccharides found in, among other pathogens, fungal cell walls. The β-glucans include one or more α-linkages between glucose subunits, whereas β-glucans include one or more β-linkages between glucose subunits. The glucan used in accordance with the invention includes β linkages, and may contain only β linkages (*i.e.*, no α linkages).

The glucan may comprise one or more β-l,3-linkages and/or one or more β-l,6-linkages. It may also comprise one or more β-l,2-linkages and/or β-l,4-linkages, but normally its only β linkages will be β- 1,3-linkages and /or β-l,6-linkages.

The glucan may be branched or linear.

Full-length native β-glucans are insoluble and have a molecular weight in the megadalton range. It is preferred to use soluble glucans in immunogenic compositions of the invention. Solubilization may be achieved by fragmenting long insoluble glucans. This may be achieved by hydrolysis or, more conveniently, by digestion with a glucanase (*e.g.*, with a β-l,3-glucanase or a β-l,6-glucanase). As an alternative, short glucans can be prepared synthetically by joining monosaccharide building blocks.

Low molecular weight glucans are preferred, particularly those with a molecular weight of less than 100 kDa (*e.g.*, less than 80, 70, 60, 50, 40, 30, 25, 20, or 15 kDa). It is also possible to use oligosaccharides containing, for example, 60 or fewer (*e.g.*, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, or 4) glucose monosaccharide units. Within this range, oligosaccharides with between 10 and 50 or between 20 and 40 monosaccharide units are preferred.

The glucan may be a fungal glucan. A "fungal glucan" will generally be obtained from a fungus but, where a particular glucan structure is found in both fungi and non-fungi (*e.g.*, in bacteria, lower plants or algae) then the non-fungal organism may be used as an alternative source. Thus the glucan may be derived from the cell wall of a Candida, such as *C. albicans*, or from *Coccidioides immitis*, *Trichophyton verrucosum*, *Blastomyces dermatidis*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, *Saccharomyces cerevisiae*, *Paracoccidioides brasiliensis*, or *Pythiumn insidiosum*. Exemplary sources of fungal β-glucans may be found in WO2009/077854.

In some embodiments, the glucan is a β-1,3 glucan with some β-1,6 branching, as seen in, for example, laminarins. Laminarins are found in brown algae and seaweeds. The β(1-3):β(1-6) ratios of laminarins vary between different sources, for example, the ratio is as low as 3:2 in *Eisenia bicyclis* laminarin, but as high as 7:1 in *Laminaria digititata* laminarin (Pang *et al.* (2005) *Biosci Biotechnol Biochem* 69:553-8). Thus the glucan used with the invention may have a β(1-3): β(1-6) ratio of between 1.5:1 and 7.5:1 (*e.g.*, about 2:1, 3:1, 4:1, 5:1, 6:1 or 7:1). Optionally, the glucan may have a terminal mannitol subunit, e.g., a 1,1-α-linked mannitol residue (Read *et al.* (1996) *Carbohydr Res.* 281:187-201). The glucan may also comprise mannose subunits.

In other embodiments, the glucan has exclusively or mainly β-1,3 linkages, as seen in curdlan. The inventors have found that these glucans may be more immunogenic than glucans comprising other linkages, particularly glucans comprising β-1,3 linkages and a greater proportion of β-1,6 linkages. Thus the glucan may be made solely of β-1,3-linked glucose residues (*e.g.*, linear β-D-glucopyranoses with exclusively 1,3 linkages). Optionally, though, the glucan may include monosaccharide residues that are not β-1,3-linked glucose residues, *e.g.*, it may include β-1,6-linked glucose residues. The ratio of β-1,3-linked glucose residues to these other residues should be at least 8:1 (e.g. >9:1, >10:1, > 11:1, >12:1, >13:1, >14:1, >15:1, >16:1, >17:1, >18:1, >19:1, >20:1, >25:1, >30:1, >35:1, >40:1, >45:1, >50:1, >75:1, >100:1, *etc.*) and/or there are one or more (e.g. >1, >2, >3, >4, >5, >6, >7, >8, >9, >10, > 11, >12, etc.) sequences of at least five (e.g. >5, >6, >7, >8, >9, >10, >11, >12, >13, >14, >15, >16, >17, >18, >19, >20, >30, >40, >50, >60, etc.) adjacent non-terminal residues linked to other residues only by β-1,3 linkages. By "non-terminal" it is meant that the residue is not present at a free end of the glucan. In some embodiments, the adjacent non-terminal residues may not include any residues conjugated to a carrier molecule, linker or other spacer as described below. The inventors have found that the presence of five adjacent non-terminal residues linked to other residues only by β-1,3 linkages may provide a protective antibody response, *e.g.*, against *C. albicans*.

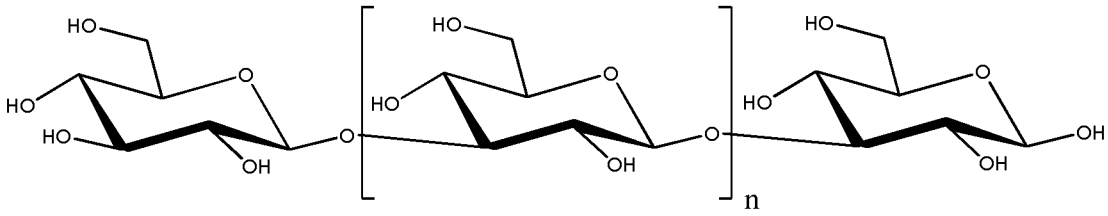
In further embodiments, a composition may include two different glucans, *e.g.*, a first glucan having a β(1-3): β(1-6) ratio of between 1.5:1 and 7.5:1, and a second glucan having exclusively or mainly β-1,3 linkages. For instance a composition may include both a laminarin glucan and a curdlan glucan.

Exemplary methods of preparing β-glucans may be found in WO2009/077854.

Laminarin and curdlan are typically found in nature as high molecular weight polymers e.g. with a molecular weight of at least 100kDa. They are often insoluble in aqueous media. In their natural forms, therefore, they are not well suited to immunization. Thus the invention may use a shorter glucan, *e.g.*, those containing 60 or fewer glucose monosaccharide units (e.g. 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4). A

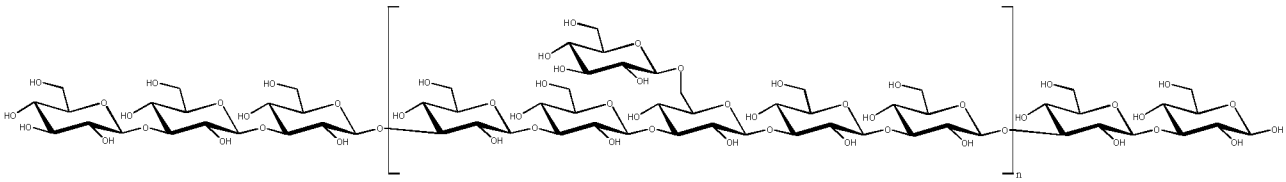
glucan having a number of glucose residues in the range of 2-60 may be used, for example, between 10-50 or between 20-40 glucose units. A glucan with 25-30 glucose residues is particularly useful. Suitable glucans may be formed e.g. by acid hydrolysis of a natural glucan, or by enzymatic digestion, *e.g.*, with a glucanase, such as a β-l,3-glucanase. A glucan with 11-19, *e.g.*, 13-19 and particularly 15 or 17, glucose monosaccharide units is also useful. In particular, glucans with the following structures (A) or (B) are specifically envisaged for use in the present invention:

(A)



wherein n+2 is in the range of 2-60, *e.g.*, between 10-50 or between 2-40. Preferably, n+2 is in the range of 25-30 or 11-19, *e.g.*, 13-17. The inventors have found that n+2 = 15 is suitable.

(B)



wherein n is in the range of 0-9, *e.g.*, between 1-7 or between 2-6. Preferably, n is in the range of 3-4 or 1-3. The inventors have found that n = 2 is suitable.

The glucan (as defined above) is preferably a single molecular species. In this embodiment, all of the glucan molecules are identical in terms of sequence. Accordingly, all of the glucan molecules are identical in terms of their structural properties, including molecular weight, *etc.* Typically, this form of glucan is obtained by chemical synthesis, *e.g.*, using the methods described above. For example, Jamois *et al.* (2005) *Glycobiology* 15(4):393-407, describes the synthesis of a single β-1,3 linked species. Alternatively, in other embodiments, the glucan may be obtained from a natural glucan, *e.g.*, a glucan from *L. digitata*, Agrobacterium or Euglena as described above, with the glucan being purified until the required single molecular species is obtained. Natural glucans that have been purified in this way are commercially available. A glucan that is a single molecular species may be identified by measuring the polydispersity (Mw/Mn) of the glucan sample. This parameter can conveniently be measured by SEC-MALLS, for example as

described in Bardotti *et al.* (2008) *Vaccine* 26:2284-96. Suitable glucans for use in this embodiment of the invention have a polydispersity of about 1, *e.g.*, 1.01 or less.

The solubility of natural glucans, such as curdlan, can be increased by introducing ionic groups (*e.g.*, by sulfation, particularly at O-6 in curdlan). Such modifications may be used with the invention, but are ideally avoided as they may alter the glucan's antigenicity.

When glucans are isolated from natural sources, they may be isolated in combination with contaminants. For example, the inventors have found that glucans may be contaminated with phlorotannin, which is identifiable by ultraviolet-visible (UV/VIS) spectroscopy. This problem is particularly common when the glucan is isolated from a brown alga or seaweed. For example, the UV spectrum of a commercially- available laminarin extracted from *Laminaria digitata* includes an absorption peak resulting from the presence of phlorotannin contamination. Similarly, glucans extracted from *Artie laminarialis*, *Saccorhiza dermatodea* and *Alaria esculenta* have UV spectra that include an absorption peak resulting from phlorotannin contamination.

The presence of phlorotannin in a sample of glucan may affect the biological properties of the glucan. Accordingly, it may be desirable to remove phlorotannin from the sample, especially when the glucan is for medical use numerous aspects of the present invention. Exemplary methods of removing phlorotannins from β-glucans may be found in WO2009/077854.

N. meningitidis: In certain embodiments, the conjugate compositions may include capsular saccharides from at least two of serogroups A, C, W135 and Y of *Neisseria meningitidis*. In other embodiments, such compositions further comprise an antigen from one or more of the following: (a) N. meningitidis; (b) *Haemophilus influenzae* type B; Staphylococcus aureus, groups A and B streptococcus, pathogenic E. coli, and/or (c) Streptococcus pneumoniae.

In certain embodiments the conjugate compositions include capsular polysaccharides from serogroups C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from serogroups A, C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from *H. influenzae* type B and serogroups C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from *H. influenzae* type B and serogroups A, C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from *S. pneumoniae* and serogroups C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from *S. pneumoniae* and serogroups A, C, W135 & Y of *N. meningitidis*. In certain embodiments the conjugate compositions include capsular polysaccharides from *H. influenzae* type B, *S. pneumoniae* and serogroups C, W135 & Y of *N. meningitidis*. In certain embodiments the

conjugate compositions include capsular polysaccharides from *H. influenzae* type B, *S. pneumoniae* and serogroups A, C, W135 & Y of *N. meningitidis*.

Streptococcus pneumoniae: *Streptococcus pneumoniae* polysaccharide conjugates may include a saccharide (including a polysaccharide or an oligosaccharide) and optionally one or more proteins from *Streptococcus pneumoniae*. Saccharide antigens maybe selected from serotypes 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 1OA, HA, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, and 33F. Optional protein antigens may be selected from a protein identified in WO98/18931, WO98/18930, U.S. Pat. No. 6,699,703, U.S. Pat. No. 6,800,744, WO97/43303, and WO97/37026. *Streptococcus pneumoniae* proteins may be selected from the Poly Histidine Triad family (PhtX), the Choline Binding Protein family (CbpX), CbpX truncates, LytX family, LytX truncates, CbpX truncate-LytX truncate chimeric proteins, pneumolysin (Ply), PspA, PsaA, Spl28, SpIOl, Spl30, Spl25 or Spl33.

Staphylococcus aureus: *Staphylococcus aureus* polysaccharide conjugates may include *S. aureus* type 5, 8 and 336 capsular polysaccharides and fragments thereof and optionally protein antigens derived from surface proteins, invasins (leukocidin, kinases, hyaluronidase), surface factors that inhibit phagocytic engulfment (capsule, Protein A), carotenoids, catalase production, Protein A, coagulase, clotting factor, and/or membrane-damaging toxins (optionally detoxified) that lyse eukaryotic cell membranes (hemolysins, leukotoxin, leukocidin). Exemplary depolymerization methods of generating fragments of *S. aureus* capsular polysaccharides may be found in U.S. Ser. No. 61/247,518, titled "Conjugation of Staphylococcus Aureus Type 5 and Type 8 Capsular Polysaccharides," filed Sept. 30, 2009, from page 5, line 6 through page 6, line 23, which is hereby incorporated by reference for its teaching of such depolymerization techniques.

***Haemophilus influenzae* B (Hib)**: Hib polysaccharide conjugates may include Hib saccharide antigens.

***Streptococcus agalactiae* (Group B Streptococcus)**: Group B Streptococcus polysaccharide conjugates may include saccharide antigens derived from serotypes Ia, Ib, Ia/c, II, III, IV, V, VI, VII and VIII as identified in WO04/041157 and optionally one or more protein antigens including, without limitation as identified in WO02/34771, WO03/093306, WO04/041157, or WO05/002619 (including by way of example proteins GBS 80, GBS 104, GBS 276 and GBS 322).

Vibrio cholerae: *V. cholerae* polysaccharide conjugates may include LPS, particularly lipopolysaccharides of Vibrio cholerae II, O1 Inaba O-specific polysaccharides.

Salmonella typhi (typhoid fever): Polysaccharide conjugates may include capsular polysaccharides such as Vi.

Carrier polypeptides

A number of promising antigens from *E. coli* were tested for their efficacy as carrier polypeptides. One such antigen is annotated as Bacterial Ig-like domain (group 1) protein (also as 'orf405', SEQ IDs 809 & 810 in reference 1), which is also known as: 'orf284' from *E. coli* NMEC strain IHE3034, 'c0415' from *E. coli* strain CFT073 and ecp_0367 from *E. coli* strain 536. Yet another such antigen is annotated as Flu antigen 43 protein (also as 'orf1364', SEQ IDs 2727 & 2728 in reference 1), which is also known as: 'orf1109' from *E. coli* NMEC strain IHE3034, 'c1273' from *E. coli* strain CFT073 and ecp_3009 from *E. coli* strain 536. Yet another such antigen is annotated as accessory colonization factor D (AcfD) precursor protein (also as 'orf3526', SEQ IDs 7051 & 7052 in reference 1), which is also known as: 'ECP_3050' from *E. coli* UPEC strain 536, 'yghJ' from *E. coli* commensal strain W3110, 'EcE24377A_3432' from *E. coli* ETEC strain E24377A, and 'EcHS_A3142' from *E. coli* commensal strain HS. Yet another such antigen is annotated as Fimbrial protein (also as 'orf3613', SEQ IDs 7225 & 7226 in reference 1), which is also known as: 'orf3431' from *E. coli* NMEC strain IHE3034 and 'c3791' from *E. coli* strain CFT073. Still another such antigen is annotated as Sell repeat-containing protein (also as 'upec-5211', disclosed in reference 2 SEQ ID 577) is also known as: 'c5321' from CFT073; 'ECED1_5081' from ED1a and 'EFER_4303' from *E. fergusonii* ATCC 35469. In particular, AcfD precursor (orf3526), Flu antigen 43 protein (orf1364), and Sell repeat-containing protein (upec-5211) were all demonstrated to be conjugable, soluble as conjugates and effective in enhancing the immune response to the conjugated polysaccharide.

Carrier polypeptides used with the invention can take various forms (*e.g.* native, fusions, glycosylated, non-glycosylated, lipidated, non-lipidated, phosphorylated, non-phosphorylated, myristoylated, non-myristoylated, monomeric, multimeric, particulate, denatured, *etc.*). For instance, a polypeptide of the invention may have a lipidated N-terminal cysteine.

Carrier polypeptides used with the invention can be prepared by various means (*e.g.* recombinant expression, purification from cell culture, chemical synthesis, *etc.*). Recombinantly-expressed proteins are preferred.

Carrier polypeptides used with the invention are preferably provided in purified or substantially purified form, *i.e.*, substantially free from other polypeptides (*e.g.* free from naturally-occurring polypeptides), particularly from other *E. coli* or host cell polypeptides, and are generally at least about 50% pure (by weight), and usually at least about 90% pure, *i.e.*, less than about 50%, and more preferably less than about 10% (*e.g.*, 5%) of a composition is made up of other expressed polypeptides. Thus the antigens in the compositions are separated from the whole organism with which the molecule is expressed. For the avoidance of doubt, when a purified or substantially carrier protein conjugated to a polysaccharide is used as a component of a vaccine, the carrier

protein is still "purified" or "substantially purified" despite the presence of other protein antigens and cellular components (*e.g.*, other polysaccharides, outer membrane vesicles, *etc.*).

The carrier polypeptides used with the invention are preferably *E. coli* polypeptides. Such polypeptides may be further selected from NMEC, APEC, UPEC, EAEC, EIEC, EPEC and ETEC *E. coli* polypeptides.

The term "polypeptide" refers to amino acid polymers of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, *etc.*), as well as other modifications known in the art. Polypeptides can occur as single chains or associated chains.

The invention also includes carrier polypeptides comprising a sequence -P-Q- or -Q-P-, wherein: -P- is an amino acid sequence as defined above and -Q- is not a sequence as defined above *i.e.* the invention provides fusion proteins. Where the N-terminus codon of -P- is not ATG, but this codon is not present at the N-terminus of a polypeptide, it will be translated as the standard amino acid for that codon rather than as a Met. Where this codon is at the N-terminus of a polypeptide, however, it will be translated as Met. Examples of -Q- moieties include, but are not limited to, histidine tags (*i.e.*, His_n where n = 3, 4, 5, 6, 7, 8, 9, 10 or more), a maltose-binding protein, or glutathione-S-transferase (GST).

The invention also includes oligomeric proteins comprising a carrier polypeptide of the invention. The oligomer may be a dimer, a trimer, a tetramer, *etc.* The oligomer may be a homo-oligomer or a hetero-oligomer. Polypeptides in the oligomer may be covalently or non-covalently associated.

Comparison of the immune response raised in a subject by the carrier polypeptide with the immune response raised by the full length protein may be carried out use by any means available to one of skill in the art. One simple method as used in the examples below involves immunization of a model subject such as mouse and then challenge with a lethal dose of *E. coli*. For proper comparison, one of skill in the art would naturally select the same adjuvant such as Freund's complete adjuvant. In such a test the carrier polypeptide fragments of the present invention will raise a substantially similar immune response in a subject (*i.e.*, will provide substantially the same protection against the lethal challenge) if, for example, the polypeptide provides at least 70% of the protection provided by the full length protein, at least 80% of the protection provided by the full length protein, at least 85% of the protection provided by the full

length protein, at least 90% of the protection provided by the full length protein, at least 95% of the protection provided by the full length protein, at least 97% of the protection provided by the full length protein, at least 98% of the protection provided by the full length protein, or at least 99% of the protection provided by the full length protein.

The invention also provides a process for producing a polypeptide of the invention, comprising the step of culturing a host cell transformed with nucleic acid of the invention under conditions which induce polypeptide expression. The polypeptide may then be purified, *e.g.*, from culture supernatants.

The invention provides an *E. coli* cell, containing a plasmid that encodes a carrier polypeptide of the invention. The chromosome of the *E. coli* cell may include a homolog of the carrier polypeptide, or such a homolog may be absent, but in both cases the polypeptide of the invention can be expressed from the plasmid. The plasmid may include a gene encoding a marker, *etc.* These and other details of suitable plasmids are given below.

Although expression of the carrier polypeptides of the invention may take place in an *E. coli* strain, the invention will usually use a heterologous host for expression. The heterologous host may be prokaryotic (*e.g.*, a bacterium) or eukaryotic. Suitable hosts include, but are not limited to, *Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonella typhimurium, Neisseria lactamica, Neisseria cinerea, Mycobacteria* (*e.g. M. tuberculosis*), yeasts, *etc.*

The invention provides a process for producing a carrier polypeptide of the invention, comprising the step of synthesising at least part of the polypeptide by chemical means.

Any and all of the foregoing proteins, polypeptides, hybrid polypeptides, epitopes and immunogenic fragments may be in any one of a number of forms including, without limitation, recombinant, isolated or substantially purified (from materials co-existing with such proteins, polypeptides, hybrid polypeptides, epitopes and immunogenic fragments in their natural state).

Orf1364 polypeptide

Flu antigen 43 protein is referred to herein as 'orf1364.' 'orf1364' polypeptide from *E. coli* NMEC is disclosed in reference 1 (SEQ IDs 2727 & 2728) is also known as: 'orf1109' from *E. coli* NMEC strain IHE3034, 'c1273' from CFT073 and ecp_3009 from 536.

When used according to the present invention, orf1364 polypeptide may take various forms. Preferred orf1364 sequences have 50% or more identity (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 87.5%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NOs 1-22. This includes variants (*e.g.*, allelic variants, homologs, orthologs, paralogs, mutants *etc.* Alternatively, the orf1364 sequences when aligned with any of SEQ ID NOs 1-22 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus (such that for an alignment that extends to p amino acids, where p>x, there are p-x+1 such windows) has at

least x•y identical aligned amino acids, where: x is selected from 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200; y is selected from 0.50, 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99; and if x•y is not an integer then it is rounded up to the nearest integer.

The preferred pairwise alignment algorithm is the Needleman-Wunsch global alignment algorithm (3), using default parameters (*e.g.*, with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the *needle* tool in the EMBOSS package (4).

Orf1364 polypeptide sequences may have 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 (or more) single amino acid alterations (deletions, insertions, substitutions), which may be at separate locations or may be contiguous, as compared to SEQ ID NOs 1-22.

Orf1364 polypeptide sequences may, compared to any one of SEQ ID NOs 1-22, include one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, *etc.*) amino acid substitutions, such as conservative substitutions (*i.e.*, substitutions of one amino acid with another which has a related side chain). Genetically encoded amino acids are generally divided into four families: (1) acidic, *i.e.*, aspartate, glutamate; (2) basic, *i.e.*, lysine, arginine, histidine; (3) non-polar, *i.e.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar, *i.e.*, glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In general, substitution of single amino acids within these families does not have a major effect on the biological activity.

Orf1364 polypeptide sequences may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, etc.) single amino acid deletions relative to any one of SEQ ID NOs 1-22. Similarly, a polypeptides may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, *etc.*) insertions (*e.g.*, each of 1, 2, 3, 4 or 5 amino acids) relative to any one of SEQ ID NOs 1-22.

Other preferred orf1364 sequences comprise at least *n* consecutive amino acids from SEQ ID NOs 1-22, wherein *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). Preferred fragments comprise an epitope or immunogenic fragment from orf1364. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 52 or more) from the C-terminus and/or the N-terminus of SEQ ID NOs 1-22. An exemplary fragment is indicated by # under the alignment below. The exemplary fragment may comprise less than 950 amino acids, less than 900 amino acids, less than 850 amino acids, less than 800 amino acids, less than 750 amino acids, less than 700 amino acids, less than 650 amino acids, less than 600 amino acids, less than 550 amino acids, less than 500 amino acids, less than 450 amino acids, less than 440 amino acids, or less than 430 amino acids of the flu antigen 43 (orf1364) protein. Further, the exemplary fragment may not comprise at least the first 10 N-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 20 N-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first

30 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 40 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the first 50 N-terminal amino acids as compared to E. coli Flu antigen 43 protein (orf1364 polypeptide), or at least the first 52 N-terminal amino acids as compared to the E. coli AcfD (orf3526) protein. Finally, the exemplary fragment may not comprise at least the last 50 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 100 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 150 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 200 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 250 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 300 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), at least the last 325 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide), or at least the last 328 C-terminal amino acids as compared to the E. coli Flu antigen 43 protein (orf1364 polypeptide).

Strain E110019 (SEQ ID NO: 1)
Group A: strain Sakai, EDL933, EC508, EC869, EC4024, EC4042, EC4045, EC4076, EC4113, EC4115, EC4196, EC4206, EC4401, EC4486, EC4501 and TW14588 (SEQ ID NO: 2)
strain B171 (SEQ ID NO: 3)
strain E22 (SEQ ID NO: 4)
strain B171 (SEQ ID NO: 5)
strain B171 (SEQ ID NO: 6)
strain E24377A and O42 (SEQ ID NO: 7)
strain E24377A (SEQ ID NO: 8)
Group B: strain UTI89, RS218 and IHE3034 (SEQ ID NO: 9)
strain E110019 (SEQ ID NO: 10)
strain E22 (SEQ ID NO: 11)
strain H10407 (SEQ ID NO: 12)
strain F11 and 536 (SEQ ID NO: 13)
strain SECEC (SEQ ID NO: 14)
strain H10407 (SEQ ID NO: 15)
strain W3110 and DH10B (SEQ ID NO: 16)
strain MG1655 (SEQ ID NO: 17)
strain O42 (SEQ ID NO: 18)
strain B7A (SEQ ID NO: 19)
strain CFT073 (SEQ ID NO: 20)
strain O42 (SEQ ID NO: 21)
strain CFT073 (SEQ ID NO: 22)

```
                          1                                                  50
strain E110019            MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
Group A                   MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain B171               MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain E22                MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain B171               MKRHLNTSYR LVWNHITGTL VVASELARSR GKRTGVAVAL SLAAVTSVPV
strain B171               MKRHLNTSYR LVWNHITGTL VVASELARSR GKRTGVAVAL SLAAVTSVPV
strain E24377A and O42    MKRHLNTSYR LVWNHITGTL VVASELARSR GKRTGVAVAL SLAAVTSVPV
strain E24377A            MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAIAL SLAAVTSVPA
Group B                   MKRHLNTSYR LVWNHITGTL VVASELARSR GKGAGVAVAL SLAAVTSVPA
strain E110019            MKRHLNTSYR LVWNHITGTL VVASELARSR GKRTGVAVAL SLAAVTSVPV
```

```
strain E22              MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain H10407           MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAIAL SLAAVTSVPA
strain F11 and 536      MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain SECEC            MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTPVPA
strain H10407           MKRHLNTSYR LVWNHITGTL VVASELARSR GKRTGVAVAL SLATATSVPA
strain W3110 and DH10B  MKRHLNTCYR LVWNHMTGAF VVASELARAR GKRGGVAVAL SLAAVTSLPV
strain MG1655           MKRHLNTCYR LVWNHMTGAF VVASELARAR GKRGGVAVAL SLAAVTSLPV
strain O42              MKRHLNTCYR LVWNHITGAF VVASELARAR GKRGGVAVAL SLAAVTSLPV
strain B7A              MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA
strain CFT073           MKRHLNTSYR LVWNHITGAF VVASELARAR GKRAGVAVAL SLAAATSLPA
strain O42              MKRHLNTCYR LVWNHITGAF VVASELARAR GKRGGVAVAL SLAAVTSLPV
strain CFT073           MKRHLNTCYR LVWNHITGAF VVASELARAR GKRGGVAVAL SLAAVTPLPV
            Consensus   MKRHLNT-YR LVWNH-TG-- VVASELAR-R GK--GVA-AL SLA--T--P-

                        51                                                  100
strain E110019          LAADKVVQAG ETVNDGTLTN HDNQIVFGTA NGMTISTGLE LGPDSEENTG
Group A                 LAADKVVQAG ETVNDGTLTN HDNQIVFGTA NGMTISTGLE LGPDSEENTG
strain B171             LAADTVVQAG ETVNGGTLTN HDNQIVLGTA NGMTISTGLE LGPDSEENTG
strain E22              LAADTVVQAG ETVNGGTLTN HDNQIVLGTA NGMTISTGLE LGPDSEENTG
strain B171             LAADTVVQAG ETVSGGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
strain B171             LAADTVVQAG ETVSGGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
strain E24377A and O42  LAADTVVQAG ETVSGGTLTN HDNQIVFGTA NGMTISTGLE YGPDNEANTG
strain E24377A          LAADTVVQAG ETVNDGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
Group B                 LAADTVVQAG ETVNGGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
strain E110019          LAADTVVQAG ETVSGGTLTN HDNQIVFGTA NGMTISTGLE YGPDNEANTG
strain E22              LAADTVVQAG ETVSGGTLVN HDNQIVFGTA NGMTISTGLE YGPDNEANTG
strain H10407           LAADTVVQAG ETVSGGTLTN HDNQIVFGTA NGMTISSGLE YGPDNEANTG
strain F11 and 536      LAADTVVQAG ETVNDGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
strain SECEC            LAADTVVEAG ETVNGGTLTN HDNQIVFGTT NGMTISTGLE YGTDNEANTG
strain H10407           LAADSVVQAG ETVSGGTLEN HDNQIVFGTT NGITISTGLE YGPDNEANTG
strain W3110 and DH10B  LAADIVVHPG ETVNGGTLAN HDNQIVFGTT NGMTISTGLE YGPDNEANTG
strain MG1655           LAADIVVHPG ETVNGGTLAN HDNQIVFGTT NGMTISTGLE YGPDNEANTG
strain O42              LAADIVVHPG ETVNGGTLAN HDNQIVFGTT NGMTISTGLE YGPDNEANTG
strain B7A              LAADKVVQAG ETVNDGTLTN HDNQIVLGTA NGMTISTGLE YGPDNEANTG
strain CFT073           LAADSVVPAG ETVNGGTLIN HDRQFVSGTA DGMTVSTGLE LGADSDNNTG
strain O42              LAADIVVHPG ETVNGGTLVN HDNQFVSGTA DGVTVSTGLE LGPDSDDNTG
strain CFT073           LSADIVVHPG ETVNGGTLVN HDNQFVSGTA NGVTVSTGLE LGPDSDENTG
            Consensus   L-AD-VV--G ETV--GTL-N HD-Q-V-GT- -G-T-S-GLE -G-D---NTG
                          ######## ########## ########## ########## ##########

                        101                                                 150
strain E110019          GQWIQNGGIA GNTTVTTNGR QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT
Group A                 GQWIQNGGIA GNTTVTTNGR QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT
strain B171             GQWIQNGGIA GNTTVTTNGR QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT
strain E22              GQWIQNGGIA GNTTVTTNGR QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT
strain B171             GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain B171             GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain E24377A and O42  GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain E24377A          GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
Group B                 GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain E110019          GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain E22              GQWIQNGGTA NNTTVTGGGL QRVNTGGSVS DTVISAGGGQ SLQGQAVNTT
strain H10407           GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain F11 and 536      GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain SECEC            GQWVQDGGTA SNTTISSGGL QFVGAGGKAT DTIINEGGGQ SLKGLALNTT
strain H10407           GQWVQDGGTA SNTTISSGGL QFVGAGGKAT DTIINEGGGQ SLKGLALNTT
strain W3110 and DH10B  GQWVQDGGTA NKTTVTSGGL QRVNPGGSVS DTVISAGGGQ SLQGRAVNTT
strain MG1655           GQWVQDGGTA NKTTVTSGGL QRVNPGGSVS DTVISAGGGQ SLQGRAVNTT
strain O42              GQWVQDGGTA NKTTVTSGGL QRVNPGGSVS DTVISAGGGQ SLQGRAVNTT
strain B7A              GQWIQNGGIA NNTTVTGGGL QRVNAGGSVS DTVISAGGGQ SLQGQAVNTT
strain CFT073           GQQIARGGTA RNTRVTANGL QDVMAGGSTS DTVISTGGGQ NLRGKASGTV
strain O42              GQQIARGGTA RNTTVTANGL QDVMAGGSAT DTVISAGGGQ NLRGQAYGTV
strain CFT073           GQWIKAGGTG RNTTVTANGR QIVQAGGTAS DTVIRDGGGQ SLNGLAVNTT
            Consensus   GQ----GG-- --T-----G- Q-V--GG--- DT-I--GGGQ -L-G-A--T-
```

```
                          ########## ########## ########## ########## ##########

                          151                                                200
strain E110019            LNNRGEQWVH EGGVATGTII NRDGYQSVKS GGLATGTIIN TGAEGGPDSD
Group A                   LNNRGEQWVH EGGVATGTII NRDGYQSVKS GGLATGTIIN TGAEGGPDSD
strain B171               LNNRGEQWVH EGGVATGTII NRDGYQSVKS GGLATGTIIN TGAEGGPDSD
strain E22                LNNRGEQWVH EGGVATGTII NRDGYQSVKS GGLATGTIIN TGAEGGPDSD
strain B171               LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain B171               LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain E24377A and O42    LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain E24377A            LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
Group B                   LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain E110019            LNG.GEQWVH EGGIATVTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAD
strain E22                LNG.GEQWVH EGGIATGTVI NEKGWQAIKS GAVATDTVVN TGAEGGPDAE
strain H10407             LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain F11 and 536        LNG.GEQWVH EGGIATGTVI NEKGWQAVKS GAMATDTVVN TGAEGGPDAE
strain SECEC              LNG.GEQWMH EGAIATGTVI NDKGWQVVKP GAVATDTVVN TGAEGGPDAE
strain H10407             LNG.GEQWMH EGAIATGTVI NDKGWQVVKP GAVATDTVVN TGAEGGPDAE
strain W3110 and DH10B    L.NGGEQWMH EGAIATGTVI NDKGWQVVKP GTVATDTVVN TGAEGGPDAE
strain MG1655             L.NGGEQWMH EGAIATGTVI NDKGWQVVKP GTVATDTVVN TGAEGGPDAE
strain O42                L.NGGEQWMH EGAIATGTVI NDKGWQVVKP GTVATDTVVN TGAEGGPDAE
strain B7A                L.NGGEQWVH EGGIATGTVI NEKGWQAIKS GAVATDTVVN TGAEGGPDAE
strain CFT073             L.NGGDQWIH AGGRASGTVI NQDGYQTIKH GGLVTGTIVN TGAEGGPDSE
strain O42                L.NGGEQWIH AGGSASGTVI NQSGYQTIKH GGQATGTIVN TGAEGGPESE
strain CFT073             LDNRGEQWVH GGGKAAGTII NQDGYQTIKH GGLATGTIVN TGAEGGPESE
            Consensus     L---G-QW-H -G--A--T-I N--G-Q--K- G---T-T--N TGAEGGP---
                          ########## ########## ########## ########## ##########

                          201                                                250
strain E110019            NSYTGQKVQG TAESTTINKN GRQIILFSGL ARDTLIYAGG DQSVHGRALN
Group A                   NSYTGQKVQG TAESTTINKN GRQIILFSGL ARDTLIYAGG DQSVHGRALN
strain B171               NSYTGQKVQG TAESTTINKN GRQIILFSGI ARDTLIYAGG DQSVHGRALN
strain E22                NSYTGQKVQG TAESTTINKN GRQIILFSGI ARDTLIYAGG DQSVHGRALN
strain B171               NGDTGQFVRG NAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain B171               NGDTGQFVRG NAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain E24377A and O42    NGDTGQFVRG NAVRTTINEN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain E24377A            NGDTGQFVRG NAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
Group B                   NGDTGQTVYG DAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain E110019            NGDTGQFVRG NAVRTTINKN GRQIVAVEGT ANTTVVYAGG DQTVHGHALD
strain E22                NGDTGQTVYG DAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain H10407             NGDTGQFVRG NAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain F11 and 536        NGDTGQFVRG NAVRTTINEN GRQIVAAEGT ANTTVVYAGG DQTVHGYALD
strain SECEC              NGDTGQFVRG NAVRTTINKN GRQIVTVEGT ANTTVVYAGG DQTVHGHALD
strain H10407             NADTGQFVRG DAVRTTINKN GRQIVVATGV ANTTVVYAGG DQTVHGYALD
strain W3110 and DH10B    NGDTGQFVRG DAVRTTINKN GRQIVRAEGT ANTTVVYAGG DQTVHGHALD
strain MG1655             NGDTGQFVRG DAVRTTINKN GRQIVRAEGT ANTTVVYAGG DQTVHGHALD
strain O42                NGDTGQFVRG DAVRTTINKN GRQIVRAEGT ANTTVVYAGG DQTVHGHALD
strain B7A                NGDTGQTVYG DAVRTTINKN GRQIVAAEGT ANTTVVYAGG DQTVHGHALD
strain CFT073             NVSTGQMVGG IAESTTINKN GRQVIWSSGI ARDTLIYTGG DQTVHGEAHN
strain O42                NVSSGQMVGG TAESTTINKN GRQVIWSSGM ARDTLIYAGG DQTVHGEAHN
strain CFT073             NVSSGQMVGG TAESTTINKN GRQVIWSSGM ARDTLIYAGG DQTVHGEAHN
            Consensus     N---GQ-V-G -A--TTIN-N GRQ-----G- A--T--Y-GG DQ-VHG-A-
                          ########## ########## ########## ########## ##########

                          251                                                300
strain E110019            TTLNGGYQYV HRDGLALNTV INEGGWQVVK AGGAAGNTTI NQNGELRVHA
Group A                   TTLNGGYQYV HRDGLALNTV INEGGWQVVK AGGAAGNTTI NQNGELRVHA
strain B171               TTLNGGYQYV HKDGLALNTV INEGGWQVVK AGGAVGNTTI NQNGELRVHA
strain E22                TTLNGGYQYV HKDGLALNTV INEGGWQVVK AGGAVGNTTI NQNGELRVHA
strain B171               TTLNGGYQYV HNGGTASGTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain B171               TTLNGGYQYV HNGGTASGTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain E24377A and O42    TTLNGGYQYV HNGGTASDTV VNSDGWQIVK EGGLADFTTV NQKGKLQVNA
strain E24377A            TTLNGGYQYV HNGGTASGTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
Group B                   TTLNGGYQYV HNGGTASDTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
```

```
strain E110019          TTLNGGYQYV HNGGTASDTV VNSDGWQIVK EGGLADFTTV NQKGKLQVNA
strain E22              TTLNGGYQYV HNGGTASGTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain H10407           TTLNGGYQYV HNGGTASGTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain F11 and 536      TTLNGGNQYV HNGGTASGTV VNSDGWQIVK EGGLADFTIV NQKGKLQVNA
strain SECEC            TTLNGGNQYV HNGGTTSDTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain H10407           TTLNGGNQYV HNGGTASDTV VNSDGWQIIK EGGLADFTTV NQKGKLQVNA
strain W3110 and DH10B  TTLNGGYQYV HNGGTASDTV VNSDGWQIVK NGGVAGNTTV NQKGRLQVDA
strain MG1655           TTLNGGYQYV HNGGTASDTV VNSDGWQIVK NGGVAGNTTV NQKGRLQVDA
strain O42              TTLNGGYQYV HNGGTASDTV VNSDGWQIVK NGGVAGNTTV NQKGRLQVDA
strain B7A              TTLNGGYQYV HNGGTASGTV VNSDGWQIVK NGGVAGNTTV NQKGRLQVDA
strain CFT073           TRLEGGNQYV HKYGLALNTV INEGGWQVVK AGGTAGNTTI NQNGELRVHA
strain O42              TRLEGGNQYV HKYGLALNTV INEGGWQVIK EGGTTAHTTI NQKGKLQVNA
strain CFT073           TRLEGGNQYV HNGGTATETL INRDGWQVIK EGGTAAHTTI NQKGKLQVNA
            Consensus   T-L-GG-QYV H--G----T- -N--GWQ--K -GG----T-- NQ-G-L-V-A
                        ########## ########## ########## ########## ##########

                        301                                                350
strain E110019          GGEATAVTQN TGGALVTSTA ATVIGTNRLG NFTVENGKAD GVVLESGGRL
Group A                 GGEATAVTQN TGGALVTSTA ATVIGTNRLG NFTVENGKAD GVVLESGGRL
strain B171             GGEATAVTQN TGGALVTSTA ATVTGANRLG HFSVGNGMAD NVVLENGGRL
strain E22              GGEATAVTQN TGGALVTSTA ATVTGANRLG HFSVGNGMAD NVVLENGGRL
strain B171             GGTATHVTLK QGGALVTSTA ATVLGSNRLG NFTVENGKAD GVVLESGGRL
strain B171             GGTATHVTLK QGGALVTSTA ATVLGSNRLG NFTVENGKAD GVVLESGGRL
strain E24377A and O42  GGTATNVTLK QGGALVTSTA ATVTGSNRLG NFTVENGNAD GVVLESGGRL
strain E24377A          GGTATNVTLK QGGALVTSTA ATVTGSNRLG NFTVENGKAD GVVLESGGRL
Group B                 GGTATNVTLT QGGALVTSTA ATVTGSNRLG NFTVENGNAD GVVLESGGRL
strain E110019          GGTATNVTLK QGGALVTSTA ATVTGSNRLG NFTVENGNAD GVVLESGGRL
strain E22              GGTATNVTLK QGGALVTSTA ATVLGSNRLG NFTVENGKAD GVVLESGGRL
strain H10407           GGTATHVTLK QGGALVTSTA ATVLGSNRLG NFTVENGKAD GVVLESGGRL
strain F11 and 536      GGTATNVTLK QGGALVTSTA ATVTGSNRLG NFTVENGNAD GVVLESGGRL
strain SECEC            GGTATNVTLK QGGALVTSTA ATVTGSNRLG NFAVENGKAD GVVLESGGRL
strain H10407           GGTATNVTLK QGGALVTSTA ATVLGSNRLG NFTVENGKAD GVVLESGGRL
strain W3110 and DH10B  GGTATNVTLK QGGALVTSTA ATVTGINRLG AFSVVEGKAD NVVLENGGRL
strain MG1655           GGTATNVTLK QGGALVTSTA ATVTGINRLG AFSVVEGKAD NVVLENGGRL
strain O42              GGTATNVTLK QGGALVTSTA ATVTGINRLG AFSVVEGKAD NVVLENGGRL
strain B7A              GGTATNVTLK QGGALVTSTA ATVTGINRLG AFSVVEGKAD NVVLENGGRL
strain CFT073           GGEASDVTQN TGGALVTSTA ATVTGTNRLG AFSVVEGKAD NVVLENGGRL
strain O42              GGKASDVTQN TGGALVTSTA ATVTGTNRLG AFSVLAGKAD NVVLENGGRL
strain CFT073           GGKASDVTQN TGGALVTSTA ATVTGTNRLG AFSVVAGKAD NVVLENGGRL
            Consensus   GG-A--VT-- -GGALVTSTA ATV-G-NRLG -F-V--G-AD -VVLE-GGRL
                        ########## ########## ########## ########## ##########

                        351                                                400
strain E110019          DVLESHSAQN TLVDDGGTLA VSAGGKATSV TITSGGALIA DSGATVEGTN
Group A                 DVLESHSAQN TLVDDGGTLA VSAGGKATSV TITSGGALIA DSGATVEGTN
strain B171             DVLEGHSAQN TLVDDGGTLA VSAGGKATDV TMTSGGALIA DSGATVEGTN
strain E22              DVLEGHSAQN TLVDDGGTLA VSAGGKATDV TMTSGGALIA DSGATVEGTN
strain B171             DVLEGHSAQK TRVDDGGTLA VSAGGKATDV TMTSGSALIA DSGATVEGTN
strain B171             DVLEGHSAQK TRVDDGGTLA VSAGGKATDV TMTSGSALIA DSGATVEGTN
strain E24377A and O42  DVLEGHSAWK TLVDDGGTLA VSAGGKATDV TMTSGSALIA DSGATVEGTN
strain E24377A          DVLEGHSAWK TLVDDGGTLA VSAGGKATDV TMTSGGALIA DSGATVEGTN
Group B                 DVLEGHSAWK TLVDDGGTLA VSAGGKATDV TMTSGGALIA DSGATVEGTN
strain E110019          DVLEGHSAWK TRVDDGGTLA VSAGGKATGV TMTSGGALIA DSGATVEGTN
strain E22              DVLEGHSAWK TLVDDGGTLA VSAGGKATGV TMTSGGALIA DSGATVEGTN
strain H10407           DVLEGHSAQK TRVDDGGTLA VSAGGKATGV TMTSGGALIA DSGATVEGTN
strain F11 and 536      DVLEGHSAWK TLVDDGGTLA VSAGGKATDV TMTSGGALIA DSGATVEGTN
strain SECEC            DVLEGHSAQK TRVDDGGTLA VSAGGKATGV TMTSGGALIA DSGATVEGTN
strain H10407           DVLEGHSAWK TLVDDGGILA VSAGGKATDV TMTSGGALIA DSGATVEGTN
strain W3110 and DH10B  DVLTGHTATN TRVDDGGTLD VRNGGTATTV SMGNGGVLLA DSGAAVSGTR
strain MG1655           DVLTGHTATN TRVDDGGTLD VRNGGTATTV SMGNGGVLLA DSGAAVSGTR
strain O42              DVLTGHTATN TRVDDGGTLD VRNGGTATTV SMGNGGVLLA DSGAAVSGTR
strain B7A              DVLTGHTATN TRVDDGGTLD VRNGGTATTV SMGNGGVLLA DSGAAVSGTR
strain CFT073           DVLSGHTATR TLVDDGGTLD VRNGGTATAV SMGNGGVLLA DSGAAVSGTR
strain O42              DVLSGHTATN TRVDDGGTLD VRNGGAATTV SMGNGGVLLA DSGAAVSGTR
```

```
strain CFT073            DVLSGHTATN TRVDDGGTLD IRNGGAATTV SMGNGGVLLA DSGAAVSGTR
              Consensus  DVL--H-A-- T-VDDGG-L- ---GG-AT-V ----G--L-A DSGA-V-GT-
                         ########## ########## ########## ########## ##########

                         401                                                450
strain E110019           ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
Group A                  ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain B171              ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain E22               ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain B171              ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGLASNTTVG HRGTLTLAAG
strain B171              ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGLASNTTVG HRGTLTLAAG
strain E24377A and O42   ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGLASNTTVG HRGTLTLAAG
strain E24377A           ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGLASNTTVG HRGTLTLAAG
Group B                  ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGLASNTTVG HRGTLTLAAG
strain E110019           ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQASNTTVG HRGTLMLAAG
strain E22               ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQASNTTVG HRGTLMLAAG
strain H10407            ASGK.FSIDG TSGQASGLLL ENGGSFTVNA GGQASNTTVG HRGTLMLAAG
strain F11 and 536       ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain SECEC             ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain H10407            ASGK.FSIDG ISGQASGLLL ENGGSFTVNA GGQAGNTTVG HRGTLTLAAG
strain W3110 and DH10B   SDGKAFSIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain MG1655            SDGKAFSIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain O42               SDGKAFSIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain B7A               SDGKAFSIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain CFT073            SDGTAFRIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain O42               SDGTAFRIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
strain CFT073            SDGKAFSIGG ..GQADALML EKGSSFTLNA GDTATDTTV. .NGGLFTARG
              Consensus  --G--F-I-G --GQA--L-L E-G-SFT-NA G--A--TTV- --G-L--A-G
                         ########## ########## ########## ########## ##########

                         451                                                500
strain E110019           GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
Group A                  GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain B171              GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain E22               GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain B171              GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain B171              GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain E24377A and O42   GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain E24377A           GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
Group B                  GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain E110019           GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain E22               GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain H10407            GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain F11 and 536       GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain SECEC             GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain H10407            GSLSGRTQLS KGASMVLNGD VVST...... .GDIV..... ..........
strain W3110 and DH10B   GTLAGTTTLN NGAILTLSGK TVNNDTLTIR EGDALLQGGS LTGNGSVEKS
strain MG1655            GTLAGTTTLN NGAILTLSGK TVNNDTLTIR EGDALLQGGS LTGNGSVEKS
strain O42               GTLAGTTTLN NGAILTLSGK TVNNDTLTIR EGDALLQGGS LTGNGSVEKS
strain B7A               GTLAGTTTLN NGAILTLSGK TVNNDTLTIR EGDALLQGGS LTGNGSVEKS
strain CFT073            GSLAGTTTLN NGATFTLAGK TVNNDTLTIR EGDALLQGGA LTGNGRVEKS
strain O42               GSLAGTTTLN NGATLTLSGK TVNNDTLTIR EGDALLQGGA LTGNGRVEKS
strain CFT073            GTLAGTTTLN NGAILTLSGK TVNNDTLTIR EGDALLQGGS LTGNGSVEKS
              Consensus  G-L-G-T-L- -GA---L-G- -V-------- -GD------- ----------
                         ########## ########## ####        ####

                         501                                                550
strain E110019           NAGEIRFDNQ T.TPNAA.LS R.AVAKSNSP VTFH...... ...KLTTT..
Group A                  NAGEIRFDNQ T.TPNAA.LS R.AVAKSNSP VTFH...... ...KLTTT..
strain B171              NAGEIRFDNQ T.TQDAV.LS R.AVAKGDSP VTFH...... ...KLTTN..
strain E22               NAGEIRFDNQ T.TQDAV.LS R.AVAKGDSP VTFH...... ...KLTTN..
strain B171              NAGEIRFDNQ T.TQDAV.LS R.AVAKGDSP VTFH...... ...KLTTS..
strain B171              NAGEIRFDNQ T.TQDAV.LS R.AVAKGDSP VTFH...... ...KLTTS..
strain E24377A and O42   NAGEIRFDNQ T.TPDAA.LS R.AVAKGDSP VTFH...... ...KLTTS..
```

```
strain E24377A           NAGEIRFDNQ T.TPDAV.LS R.AVAKGDSP VTFH...... ...KLTTS..
Group B                  NAGEIRFDNQ T.TPDAA.LS R.AVAKGDSP VTFH...... ...KLTTS..
strain E110019           NAGEIYFDNQ T.TPDAV.LS R.AVAKGNAP VTFH...... ...KLTTS..
strain E22               NAGEIYFDNQ T.TPDAV.LS R.AVAKGNAP VTFH...... ...KLTTS..
strain H10407            NAGEIHFDNQ T.TQDAV.LS R.AVAKSNSP VTFH...... ...KLTTT..
strain F11 and 536       NAGEIHFDNQ T.TPDAA.LS R.AVAKGDSP VTFH...... ...KLTTS..
strain SECEC             NAGEIRFDNQ T.TQDAV.LS R.AVAKGDAP VTFH...... ...KLTTS..
strain H10407            NAGEIHFDNQ T.TQDAV.LS R.AVAKSNSP VTFH...... ...KLTTT..
strain W3110 and DH10B   GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
strain MG1655            GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
strain O42               GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
strain B7A               GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
strain CFT073            GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDIIAHRG TALKLTGSTV
strain O42               GSGTLTVSNT TLTQKTVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
strain CFT073            GSGTLTVSNT TLTQKAVNLN EGTLTLNDST VTTDVIAQRG TALKLTGSTV
                Consensus --G-----N- T-T-----L- ---------- VT-------- ---KLT----
                           ########## ########## ########## ####         #####

                           551                                                600
strain E110019           .......... .......... .......... .......... ..........
Group A                  .......... .......... .......... .......... ..........
strain B171              .......... .......... .......... .......... ..........
strain E22               .......... .......... .......... .......... ..........
strain B171              .......... .......... .......... .......... ..........
strain B171              .......... .......... .......... .......... ..........
strain E24377A and O42   .......... .......... .......... .......... ..........
strain E24377A           .......... .......... .......... .......... ..........
Group B                  .......... .......... .......... .......... ..........
strain E110019           .......... .......... .......... .......... ..........
strain E22               .......... .......... .......... .......... ..........
strain H10407            .......... .......... .......... .......... ..........
strain F11 and 536       .......... .......... .......... .......... ..........
strain SECEC             .......... .......... .......... .......... ..........
strain H10407            .......... .......... .......... .......... ..........
strain W3110 and DH10B   LNGAIDPTNV TLASGATWNI PDNATVQSVV DDLSHAGQIH FTSTRTGKFV
strain MG1655            LNGAIDPTNV TLASGATWNI PDNATVQSVV DDLSHAGQIH FTSTRTGKFV
strain O42               LNGAIDPTNV TLASGATWNI PDNATVQSVV DDLSHAGQIH FTSTRTGKFV
strain B7A               LNGAIDPTNV TLASGATWNI PDNATVQSVV DDLSHAGQIH FTSTRTGKFV
strain CFT073            LNGAIDPTNV TLTSGATWNI PDNATVQSVV DDLSHAGQIH FTSARTGKFV
strain O42               LNGAIDPTNV TLTSGATWNI PDNATVQSVV DDLSHAGQIH FTSTRTGKFV
strain CFT073            LNGAIDPTNV TLASDATWNI PDNATVQSVV DDLSHAGQIH FTSSRTGTFV
                Consensus ---------- ---------- ---------- ---------- ----------

                           601                                                650
strain E110019           .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
Group A                  .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain B171              .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain E22               .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain B171              .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
strain B171              .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
strain E24377A and O42   .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
strain E24377A           .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
Group B                  .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain E110019           .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain E22               .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
strain H10407            .......NLT GQGGTINMRV SLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain F11 and 536       .......NLT GQGGTINMRV RLD.GSNTSD QLVINGGQAT GKTWLAFTNV
strain SECEC             .......NLT GQGGTINMRV RLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain H10407            .......NLT GQGGTINMRV SLD.GSNASD QLVINGGQAT GKTWLAFTNV
strain W3110 and DH10B   PATLKVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
strain MG1655            PATLKVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
strain O42               PATLKVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
strain B7A               PATLKVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
strain CFT073            PTTLQVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
```

```
strain O42              PATLQVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
strain CFT073           PATLKVKNLN GQNGTISLRV RPDMAQNNAD RLVIDGGRAT GKTILNLVNA
              Consensus -------NL- GQ-GTI--RV --D---N--D -LVI-GG-AT GKT-L---N-
                               ### ########## ########## ########## ##########

                        651                                                700
strain E110019          GNSNLGVATT GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
Group A                 GNSNLGVATT GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain B171             GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain E22              GNSNLGVATS GQGIRVVDAQ NGATTEESAF ALSRPLHAGA FNYTLNRDSD
strain B171             GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain B171             GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain E24377A and O42  GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain E24377A          GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
Group B                 GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain E110019          GNSNLGVATT GQGIRVVDAQ NGATTEEGVF ALSRPLQAGA FNYTLNRDSD
strain E22              GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain H10407           GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain F11 and 536      GNSNLGVATT GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain SECEC            GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain H10407           GNSNLGVATS GQGIRVVDAQ NGATTEEGAF ALSRPLQAGA FNYTLNRDSD
strain W3110 and DH10B  GNSASGLATS GKGIQVVEAI NGATTEEGAF VQGNRLQAGA FNYSLNRDSD
strain MG1655           GNSASGLATS GKGIQVVEAI NGATTEEGAF VQGNRLQAGA FNYSLNRDSD
strain O42              GNSASGLATS GKGIQVVEAI NGATTEEGAF VQGNRLQAGA FNYSLNRDSD
strain B7A              GNSASGLATS GKGIQVVEAI NGATTEEGAF IQGNKLQAGA FNYSLNRDSD
strain CFT073           GNSGTGLATT GKGIQVVEAI NGATTEEGAF VQGNMLQAGA FNYTLNRDSD
strain O42              GNSGTGLATT GKGIQVVEAI NGATTEEGAF VQGNMLQAGA FNYTLNRDSD
strain CFT073           GNSASGLATS GKGIQVVEAI NGATTEEGAF VQGNRLQAGA FNYSLNRDSD
              Consensus GNS--G-AT- G-GI-VV-A- NGATTEE--F -----L-AGA FNY-LNRDSD
                        ########## ########## ########## ########## ##########

                        701                                                750
strain E110019          EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
Group A                 EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain B171             EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain E22              EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQS GVSGENNSVR
strain B171             EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain B171             EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain E24377A and O42  EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain E24377A          EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVSGENNSVR
Group B                 EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQS GVSGENNSVR
strain E110019          EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain E22              EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQS GVSGENNSVR
strain H10407           EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain F11 and 536      EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSFR
strain SECEC            EDWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain H10407           EDWYLRSENA YRAEVPLYTS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain W3110 and DH10B  ESWYLRSENA YRAEVPLYAS MLTQAMDYDR IVAGSRSHQT GVNGENNSVR
strain MG1655           ESWYLRSENA YRAEVPLYAS MLTQAMDYDR IVAGSRSHQT GVNGENNSVR
strain O42              ESWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVSGENNSVR
strain B7A              ESWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVSGENNSVR
strain CFT073           ESWYLRSEER YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain O42              ESWYLRSEER YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
strain CFT073           ESWYLRSENA YRAEVPLYAS MLTQAMDYDR ILAGSRSHQT GVNGENNSVR
              Consensus E-WYLRSE-- YRAEVPLY-S MLTQAMDYDR I-AGSRSHQt GVnGENNS-R
                        ########## ######

                        751                                                800
strain E110019          LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
Group A                 LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain B171             LSIQGGHLGH DNNGGIARGA TPESSGSYGL VRLEGDLLRT EVAGMSLTTG
strain E22              LSIQGGHLGH DNNGGIARGA TPESNGSYGF VRLEGDLLRT EVAGMSLTTG
strain B171             LSIQGGHLGH DNNGGIARGA TPESNGSYGF VRLEGDLLRT EVAGMSLTTG
strain B171             LSIQGGHLGH DNNGGIARGA TPESNGSYGF VRLEGDLLRT EVAGMSLTTG
```

```
strain E24377A and O42   LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain E24377A           LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
Group B                  LSIQGGHLGH DNNGGIARGA TPESNGSYGF VRLEGDLLRT EVAGMSLTTG
strain E110019           LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain E22               LSIQGGHLGH DNNGGIARGA TPESNGSYGF VRLEGDLLRT EVAGMSLTTG
strain H10407            LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain F11 and 536       LSIQGGHLGH VNNGGIARGA TPESSGSYGL VRLEGDLLRT EVAGMSLTTG
strain SECEC             LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLESDLLRT EVAGMSVTAG
strain H10407            LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSVTAG
strain W3110 and DH10B   LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLMRT EVAGMSVTAG
strain MG1655            LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLMRT EVAGMSVTAG
strain O42               LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain B7A               LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSVTAG
strain CFT073            LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain O42               LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT EVAGMSLTTG
strain CFT073            LSIQGGHLGH DNNGGIARGA TPESSGSYGF VRLEGDLLRT DVAGMSVTAG
                Consensus LSIQGGHLGH -NNGGIARGA TPES-GSYG- VRLE-DL-RT -VAGMS-T-G

                          801                                             850
strain E110019           VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
Group A                  VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain B171              VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain E22               VYGAAGHSSV DVKNDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain B171              VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain B171              VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain E24377A and O42   VHGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain E24377A           VYGAAGHSSV DVKDDDGSRA GTARDDAGSL GGYLNLVHTS SGLWADIVAQ
Group B                  VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLHLVHTS SGLWADIVAQ
strain E110019           VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain E22               VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain H10407            VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain F11 and 536       VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain SECEC             VYSAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIMAQ
strain H10407            VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain W3110 and DH10B   VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLVHTS SGLWADIVAQ
strain MG1655            VYGAAGHSSV DVKDDDGSRA GTVRDDAGCL GGYLNLVHTS SGLWADIVAQ
strain O42               VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLTHTS SGLWADIVAQ
strain B7A               VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYLNLIHNA SGLWADIVAQ
strain CFT073            VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYMNLTHTS SGLWADIVAQ
strain O42               VYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYMNLTHTS SGLWADIVAQ
strain CFT073            IYGAAGHSSV DVKDDDGSRA GTVRDDAGSL GGYMNLTHTS SGLWADIVAQ
                Consensus ---AAGHSSV DVK-DDGSRA GT-RDDAG-L GGY--L-H-- SGLWADI-AQ

                          851                                             900
strain E110019           GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
Group A                  GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain B171              GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain E22               GTHHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain B171              GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNVML EPQLQYTWQG
strain B171              GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain E24377A and O42   GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain E24377A           GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
Group B                  GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain E110019           GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain E22               GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain H10407            GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLHYTWQG
strain F11 and 536       GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain SECEC             GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain H10407            GTRHSMKAST DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain W3110 and DH10B   GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain MG1655            GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain O42               GTRHSMKASS DNNDFRARGW GWLGSLETGL PFSITDNLML EPQLHYTWQG
strain B7A               GTRHSMKASS DNNDFRVRGW GWLGSLETGL PFSITDNLML EPQLQYTWQG
strain CFT073            GTRHSMKASS DNNDFRARGR GWLGSLETGL PFSITDNLML EPRLQYTWQG
```

```
strain O42              GTRHSMKASS GNNDFRARGW GWLGSLETGL PFSITDNLML EPRLQYTWQG
strain CFT073           GTRHSMKASS GNNDFRARGR GWLGSLETGL PFSITDNLML EPRLQYTWQG
                Consensus  GT-HSMKAS- -NNDFR-RG- GWLGSLETGL PFSITDN-ML EP-L-YTWQG

                           901                                             950
strain E110019          LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
Group A                 LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain B171             LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain E22              LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain B171             LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMSFGEG TSSRDTLRDS
strain B171             LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMSFGEG TSSRDTLRDS
strain E24377A and O42  LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMSFGEG TSSRDTLRDS
strain E24377A          LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMNFGKG TSSRDTLRDS
Group B                 LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMNFGKG TSSRDTLHDS
strain E110019          LSLDDGQDNA GYVKFGHGST QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain E22              LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMSFGEG TSSRDTLRDS
strain H10407           LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain F11 and 536      LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMNFGKG TSSRDTLRDS
strain SECEC            LSLDDGQDNA GYVKFGHGSA QHMRAGFRLG SHNDMSFGEG TSSRDTLRDS
strain H10407           LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
strain W3110 and DH10B  LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
strain MG1655           LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
strain O42              LSLDDGQDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRDTLRDS
strain B7A              LSLDDGQDNA SYVKFGHGSA QHVRAGFRLG SHHDMNFGKG TSSRDTLRGS
strain CFT073           LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
strain O42              LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
strain CFT073           LSLDDGKDNA GYVKFGHGSA QHVRAGFRLG SHNDMTFGEG TSSRAPLRDS
                Consensus  LSLDDG-DNA -YVKFGHGS- QH-RAGFRLG SH-DM-FG-G TSSR--L--S

                           951                                            1000
strain E110019          AKHSVSELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
Group A                 AKHSVSELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain B171             AKHSVSELPV NWWVQPSVIR TVSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain E22              AKHRVRELPV NWWVQPSVIR TVSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain B171             AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain B171             AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain E24377A and O42  AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain E24377A          AKHSVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
Group B                 AKHSVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain E110019          AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain E22              AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain H10407           TKHGVSELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain F11 and 536      AKHSVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSQNGTTL
strain SECEC            AKHRVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSQNGTSL
strain H10407           AKHSMRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSRNGTSL
strain W3110 and DH10B  AKHSVSELPV NWWVQPSVIR TFSSRGDMRV GTSTAGSGMT FSPSQNGTSL
strain MG1655           AKHSVSELPV NWWVQPSVIR TFSSRGDMRV GTSTAGSGMT FSPSQNGTSL
strain O42              TKHGVSELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSQNGTSL
strain B7A              AKHSVRELPV NWWVQPSVIR TFSSRGDMSM GTAAAGSNMT FSPSQNGTSL
strain CFT073           AKHSVRELPV NWWVQPSVIR TFSSRGDMRV GTSTAGSGMT FSPSQNGTSL
strain O42              AKHSVRELPV NWWVQPSVIR TFSSRGDMRV GTSTAGSGMT FSPSQNGTSL
strain CFT073           AKHSVRELPV NWWVQPSVIR TFSSRGDMRV GTSTAGSGMT FSPSQNGTSL
                Consensus  -KH---ELPV NWWVQPSVIR T-SSRGDM-- GT--AGS-MT FSPS-NGT-L

                           1001                                    1044
strain E110019          DLQAGLEARI RENITLGVQA GYAHSVSGSS AEGYNGQATL NMTF
Group A                 DLQAGLEARI RENITLGVQA GYAHSVSGSS AEGYNGQATL NMTF
strain B171             DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NMTF
strain E22              DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NMTF
strain B171             DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain B171             DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain E24377A and O42  DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain E24377A          DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
Group B                 DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
```

```
strain E110019          DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain E22              DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain H10407           DLQAGLEARV RENITLGVQA GYAHSVSGNS AEGYNGQATL NVTF
strain F11 and 536      DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain SECEC            DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain H10407           DLQAGLEARV RENITLGVQA GYAHSVIGSS AEGYNGQATL NVTF
strain W3110 and DH10B  DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain MG1655           DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain O42              DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
strain B7A              DLQAGLEARV RENITLGVQA GYVHSVSGSS AEGYNGQATL NVTF
strain CFT073           DLQAGLEARV RENITLGVQA GYAHSINGSS AEGYNSQATL NVTF
strain O42              DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNSQATL NVTF
strain CFT073           DLQAGLEARV RENITLGVQA GYAHSVSGSS AEGYNGQATL NVTF
              Consensus DLQAGLEAR- RENITLGVQA GY-HS--G-S AEGYN-QATL N-TF
```

Upec-5211 polypeptide

Sell repeat-containing protein is referred to herein as 'upec-5211.' 'upec-5211' polypeptide from *E. coli* is also known as: 'c5321' from CFT073; 'ECED1_5081' from ED1a and 'EFER_4303' from *E. fergusonii* ATCC 35469.

When used according to the present invention, upec-5211 polypeptide may take various forms. Preferred upec-5211 sequences have 50% or more identity (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 87.5%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NOs 23-25. This includes variants (*e.g.*, allelic variants, homologs, orthologs, paralogs, mutants *etc*).

Alternatively, the upec-5211 sequences when aligned with any of SEQ ID NOs 23-25 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus (such that for an alignment that extends to p amino acids, where p>x, there are p-x+1 such windows) has at least x•y identical aligned amino acids, where: x is selected from 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200; y is selected from 0.50, 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99; and if x•y is not an integer then it is rounded up to the nearest integer.

The preferred pairwise alignment algorithm is the Needleman-Wunsch global alignment algorithm (3), using default parameters (*e.g.*, with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the *needle* tool in the EMBOSS package (4).

Upec-5211 polypeptide sequences may have 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 (or more) single amino acid alterations (deletions, insertions, substitutions), which may be at separate locations or may be contiguous, as compared to SEQ ID NOs 23-25.

Upec-5211 polypeptide sequences may, compared to any one of SEQ ID NOs 23-25, include one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.) amino acid substitutions, such as conservative substitutions (*i.e.*, substitutions of one amino acid with another which has a related side chain). Genetically encoded amino acids are generally divided into four families: (1) acidic, *i.e.*, aspartate, glutamate; (2) basic, *i.e.*, lysine, arginine, histidine; (3) non-polar, *i.e.*, alanine, valine, leucine,

isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar, *i.e.*, glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In general, substitution of single amino acids within these families does not have a major effect on the biological activity.

Upec-5211 polypeptide sequences may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, etc.) single amino acid deletions relative to any one of SEQ ID NOs 23-25. Similarly, a polypeptides may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, *etc.*) insertions (*e.g.*, each of 1, 2, 3, 4 or 5 amino acids) relative to any one of SEQ ID NOs 23-25.

Other preferred upec-5211 sequences comprise at least *n* consecutive amino acids from SEQ ID NOs 23-25, wherein *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). Preferred fragments comprise an epitope or immunogenic fragment from upec-5211. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or the N-terminus of SEQ ID NOs 23-25.

```
Strains CFT073 and 83972  (SEQ ID NO: 23)
Strain ED1a  (SEQ ID NO: 24)
Escherichia fergusonii ATCC 35469  (SEQ ID NO: 25)

strain CFT073 and 83972  MKKSLLAVML TGLFALVSLP ALGNVNLEQL KQKAESGEAK AQLELGYRYF
strain ED1a              MKKSLLAVML TGLFALVSLP ALGNVNLEQL KQKAESGEAK AQLELGYRYF
E. fergusonii            MKKSLLAALL TGLFALVSLP ALGNVNFEQL KQKAERGEAK AQLELGYRYF
              Consensus  MKKSLLA +L TGLFALVSLP ALGNVN EQL KQKAE GEAK AQLELGYRYF
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  QGNETTKDLT QAMDWFRRAA EQGYTPAEYV LGLRYMNGEG VPQDYAQAVI
strain ED1a              QGNETTKDLT LAMDWFRRAA EQGYTPAEYV LGLRYMNGEG VPQDYAQAVI
E. fergusonii            QGNETTKDLT QAIDWFRRAA EQGYTPAEFV LGLRYMNGEG VPKDYAQAVI
              Consensus  QGNETTKDLT  A+DWFRRAA EQGYTPAE+V LGLRYMNGEG VP+DYAQAVI
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  WYKKAALKGL PQAQQNLGVM YHEGNGVKVD KAESVKWFRL AAEQGRDSGQ
strain ED1a              WYKKAALKGL PQAQQNLGVM YHEGNGVKVD KAESVKWFRL AAEQGRDSGQ
E. fergusonii            WYKKAALKGL PQAQQNLGVM YHDGKGVKID KAESVKWFRL AAEQGRDSGQ
              Consensus  WYKKAALKGL PQAQQNLGVM YH+G GVK+D KAESVKWFRL AAEQGRDSGQ
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  QSMGDAYFEG DGVTRDYVMA REWYSKAAEQ GNVWSCNQLG YMYSRGLGVE
strain ED1a              QSMGDAYFEG DGVTRDYVMA REWYSKAAEQ GNVWSCNQLG YMYSRGLGVE
E. fergusonii            QSMGDAYFEG DGVTRDYVMA REWYSKAAEQ GNVWSCNQLG YIYSKGLGVE
              Consensus  QSMGDAYFEG DGVTRDYVMA REWYSKAAEQ GNVWSCNQLG Y+YS+GLGVE
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  RNDAISAQWY RKSATSGDEL GQLHLADMYY FGIGVTQDYT QSRVLFSQSA
strain ED1a              RNDAISAQWY RKSATSGDEL GQLHLADMYY FGIGVTQDYT QSRVLFSQSA
E. fergusonii            KNDAISAQWY RKSATSGDEL GQLHLADMYY FGIGVTQDYT QSRILFTQSA
              Consensus  +NDAISAQWY RKSATSGDEL GQLHLADMYY FGIGVTQDYT QSR+LF+QSA
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  EQGNSIAQFR LGYILEQGLA GAKEPLKALE WYRKSAEQGN SDGQYYLAHL
strain ED1a              EQGNSIAQFR LGYILEQGLA GAKEPLKALE WYRKSAEQGN SDGQYYLAHL
E. fergusonii            EQGNAIAQYR LGYILEEGLA GAKEPLKALE WYRKSAEQGN AIGQYYLAEI
              Consensus  EQGN+IAQ+R LGYILE+GLA GAKEPLKALE WYRKSAEQGN + GQYYLA +
                         ########## ########## ########## ########## ##########
```

```
strain CFT073 and 83972  YDKGAEGVAK NREQAISWYT KSAEQGDATA QANLGAIYFR LGSEEEHKKA
strain ED1a              YDKGAEGVAK NREQAISWYT KSAEQGDATA QANLGAIYFR LGSEEEHKKA
E. fergusonii            YIRRAEGIPY NREQAIYWYT KSAEQGDTDA QVNLGALLYR HGSEEEQRRA
               Consensus Y + AEG+   NREQAI WYT KSAEQGD  A Q NLGA+ +R  GSEEE ++A
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  VEWFRKAAAK GEKAAQFNLG NALLQGKGVK KDEQQAAIWM RKAAEQGLSA
strain ED1a              VEWFRKAAAK GEKAAQFNLG NALLQGKGVK KDEQQAAIWM RKAAEQGLSA
E. fergusonii            VDWYRKAAEE GVAMAQFNLG NALLQGKGVK KDEQQAAIWM RKAAEQGFSS
               Consensus V+W+RKAA + G   AQFNLG NALLQGKGVK KDEQQAAIWM RKAAEQG S+
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  AQVQLGEIYY YGLGVERDYV QAWAWFDTAS TNDMNLFGTE NRNITEKKLT
strain ED1a              AQVQLGEIYY YGLGVERDYV QAWAWFDTAS TNDMNLFGTE NRNITEKKLT
E. fergusonii            AQVQLGEIYY YGLGVERDYV QAWAWFDTAS TNDMNLFGTE NRNITEKKLT
               Consensus AQVQLGEIYY YGLGVERDYV QAWAWFDTAS TNDMNLFGTE NRNITEKKLT
                         ########## ########## ########## ########## ##########

strain CFT073 and 83972  AKQLQQAELL SQQYIEKYAP EAWARMQKLK AQSAVKTGNK
strain ED1a              TKQLQQAELL SQQYIEKYAT EAWARMQKLK AQSAVKTGNK
E. fergusonii            AKQLQQAELL SQQYIEKYAP EAWARMQKLN ARSTVTTGNK
               Consensus  KQLQQAELL SQQYIEKYA  EAWARMQKL  A+S V TGNK
                         ########## ########## ########## ##########
```

Orf3526 polypeptide

The accessory colonization factor D (AcfD) precursor protein is referred to herein as 'orf3526.' 'orf3526' polypeptide from *E. coli* NMEC is disclosed in reference 1 (SEQ ID NOs: 7051 & 7052) is also known as: 'ECP_3050' from *E. coli* UPEC strain 536, 'yghJ' from *E. coli* commensal strain W3110, 'EcE24377A_3432' from *E. coli* ETEC strain E24377A, and 'EcHS_A3142' from *E. coli* commensal strain HS.

When used according to the present invention, orf3526 polypeptide may take various forms. Preferred orf3526 sequences have 50% or more identity (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 87.5%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NOs 26-40. This includes variants (*e.g.*, allelic variants, homologs, orthologs, paralogs, mutants *etc*).

Alternatively, the orf3526 sequences when aligned with any of SEQ ID NOs 26-40 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus (such that for an alignment that extends to p amino acids, where p>x, there are p-x+1 such windows) has at least x•y identical aligned amino acids, where: x is selected from 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200; y is selected from 0.50, 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99; and if x•y is not an integer then it is rounded up to the nearest integer.

The preferred pairwise alignment algorithm is the Needleman-Wunsch global alignment algorithm (3), using default parameters (*e.g.*, with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the *needle* tool in the EMBOSS package (4).

Orf3526 polypeptide sequences may have 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 (or more) single amino acid alterations (deletions, insertions, substitutions), which may be at separate locations or may be contiguous, as compared to SEQ ID NOs 26-40.

Orf3526 polypeptide sequences may, compared to any one of SEQ ID NOs 26-40, include one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.) amino acid substitutions, such as conservative substitutions (*i.e.*, substitutions of one amino acid with another which has a related side chain). Genetically encoded amino acids are generally divided into four families: (1) acidic, *i.e.*, aspartate, glutamate; (2) basic, *i.e.*, lysine, arginine, histidine; (3) non-polar, *i.e.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar, *i.e.*, glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In general, substitution of single amino acids within these families does not have a major effect on the biological activity.

Orf3526 polypeptide sequences may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, etc.) single amino acid deletions relative to any one of SEQ ID NOs 26-40. Similarly, a polypeptides may include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, *etc.*) insertions (*e.g.*, each of 1, 2, 3, 4 or 5 amino acids) relative to any one of SEQ ID NOs 26-40.

Other preferred orf3526 sequences comprise at least *n* consecutive amino acids from SEQ ID NOs 26-40, wherein *n* is 7 or more (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). Preferred fragments comprise an epitope or immunogenic fragment from orf3526. Other preferred fragments lack one or more amino acids (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or the N-terminus of SEQ ID NOs 26-40. The extent of the gly-ser region is indicated by "G" under the alignment below. The extent of the N-terminal proline-rich repeat is indicated by "P" under the alignment below. Three preferred fragments denoted orf3526A, orf3526B, and orf3526C are denoted by 'A', 'B' or 'C', respectively below the alignment.

```
EAEC strain 101-1 (GI: 83587587) (SEQ ID NO: 26)
UPEC strain 536 (GI: 110643204) (SEQ ID NO: 27)
EAEC strain O42 (SEQ ID NO: 28)
EPEC strain E2348/69  (SEQ ID NO: 29)
EIEC strain 53638 (GI: 75515237)  (SEQ ID NO: 30)
Commensal strain W3110 (GI: 89109748)  (SEQ ID NO: 31)
ETEC strain B7A (GI: 75227618)  (SEQ ID NO: 32)
EPEC strain E22 (GI: 75259912)  (SEQ ID NO: 33)
ETEC strain E24377A (GI: 157156747)  (SEQ ID NO: 34)
ETEC strain H10407  (SEQ ID NO: 35)
EPEC strain E110019 (GI: 75239450)  (SEQ ID NO: 36)
commensal strain HS (GI: 157162442)  (SEQ ID NO: 37)
antibiotic-resistant strain SECEC  (SEQ ID NO: 38)
NMEC strain IHE3034  (SEQ ID NO: 39)
UPEC strain F11 (GI: 75241179)  (SEQ ID NO: 40)
```

```
                |          |          |          |          |          |
strain 101-1    MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain 536      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain O42      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
str E2348/69    MNKKFKYKKS LLAAILSATL LAGCDGGGSG PSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain 53638    MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain W3110    MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain B7A      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain E22      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
str E24377A     MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
str H10407      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
str E110019     MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain HS       MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
strain SECEC    MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
str IHE3034     MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPSVDS GSGTLPEVKP DPTPTPEPTP
strain F11      MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPSVDS GSGTLPEVKP DPTPTPEPTP
  Consensus     MNKKFKYKKS LLAAILSATL LAGCDGGGSG SSSDTPPVDS GTGSLPEVKP DPTPNPEPTP
N-TERM REG      GGGGGGGGGG GGGGGGGGGG GGGGGGGGGG GGGPPPPPPP PPPPPPPPPP PPPPPPPPPP
                ########## ########## ########## ########## ########## ##########
                                        AAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

                                                   CCCCCCC CCCCCCCCCC CCCCCCCCCC

                |          |          |          |          |          |
strain 101-1    EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SQRITG-ATC NGESSDGFTF
strain 536      EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SQRITG-ATC NGESSDGFTF
strain O42      EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SQRITG-ATC NGESSDGFTF
str E2348/69    EPTPDPEPTP DPTPDP--EP TPEPEPEPVP TKTGYLTLGG SQRITG-ATC NGESSDGFTF
strain 53638    EPTPDPEPTP EPIPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
strain W3110    EPTPDPEPTP EPIPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
strain B7A      EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SLRVTGDITC NDESSDGFTF
strain E22      EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SLRVTGDITC NDESSDGFTF
str E24377A     EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SLRVTGDITC NDESSDGFTF
str H10407      EPTPDPEPTP EPIPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
str E110019     EPTPDPEPTP EPTPDP--EP TPEPEPEPVP TKTGYLTLGG SLRVTGDITC NDESSDGFTF
strain HS       EPTPDPEPTP EPIPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
strain SECEC    EPTPDPEPTP EPIPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
str IHE3034     EPTPDPEPTP DPTPDP--EP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
strain F11      EPTPDPEPTP DPTPDPDPEP TPEPEPEPVP TKTGYLTLGG SQRVTG-ATC NGESSDGFTF
 Consensus      EPTPDPEPTP EPTPDPDPEP TPEPEPEPVP TKTGYLTLGG SQRVTGDATC NGESSDGFTF
N-TERM REG      PPPPPPPPPP PPPPPPPPPP PPPPPPPPPP P
                ########## ########## ########## ########## ########## ##########
                AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

                CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

                |          |          |          |          |          |
strain 101-1    TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAASDD KKSNAVSLVT
strain 536      TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAASDD KKSNAVSLVT
strain O42      TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAASDD KKSNAVSLVT
str E2348/69    TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAASDD KKSNAVSLVT
strain 53638    KPGEDVTCVA G-NTTIATFN TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
strain W3110    KPGEDVTCVA G-NTTIATFN TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
strain B7A      TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAGSDN KKSNALSLVT
strain E22      TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAGSDN KKSNALSLVT
str E24377A     TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAGSDN KKSNALSLVT
str H10407      KPGEDVTCVA G-NTTIATFN TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
str E110019     TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAGSDN KKSNALSLVT
strain HS       KPGEDVTCVA G-NTTIATFN TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
strain SECEC    KPGEDVTCVA G-NTTIATFN TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
str IHE3034     TPGNTVSCVV G-STTIATFN TQSEAARSLR AVDKVSFSLE DAQELANSEN KKTNAISLVT
strain F11      TPGNTVSCVV G-STTIATFN TQSEAARSLR AVDKVSFSLE DAQELANSEN KKTNAISLVT
```

```
Consensus   TPGDKVTCVA GNNTTIATFD TQSEAARSLR AVEKVSFSLE DAQELAGSDD KKSNAVSLVT
            ########## ########## ########## ########## ########## ##########
            AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

            CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

            |          |          |          |          |          |
strain 101-1 SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain 536   SSNSCPADTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain O42   SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
str E2348/69 SSNSCPADTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain 53638 SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain W3110 SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain B7A   SMNSCPANTE QVCLEFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain E22   SMNSCPANTE QVCLEFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
str E24377A  SMNSCPANTE QVCLEFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
str H10407   SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
str E110019  SMNSCPANTE QVCLEFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain HS    SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
strain SECEC SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
str IHE3034  SSDSCPADAE QLCLTFSSVV DRARFEKLYK QIDLATDNFS KLVNEEVENN AATDKAPSTH
strain F11   SSDSCPADAE QLCLTFSSVV DRARFEKLYK QIDLATDNFS KLVNEEVENN AATDKAPSTH
Consensus   SSNSCPANTE QVCLTFSSVI ESKRFDSLYK QIDLAPEEFK KLVNEEVENN AATDKAPSTH
            ########## ########## ########## ########## ########## ##########
            AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

            CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

            |          |          |          |          |          |
strain 101-1 TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSMGNG VVGVNYYTSS
strain 536   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSMGNG VVGVNYYTSS
strain O42   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSMGNG VVGVNYYTSS
str E2348/69 TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
strain 53638 TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
strain W3110 TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
strain B7A   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPSEIILSE GRLVDSQGYG VAGVNYYTNS
strain E22   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPSEIILSE GRLVDSQGYG VAGVNYYTNS
str E24377A  TSPVVPATTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGDG VVGVNYYTNS
str H10407   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
str E110019  TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPSEIILSE GRLVDSQGYG VAGVNYYTNS
strain HS    TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
strain SECEC TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
str IHE3034  TSTVVPVTTE GTKPDLNASF VSANAEQFYQ YQPTEIILSE GQLVDSLGNG VAGVDYYTNS
strain F11   TSTVVPVTTE GTKPDLNASF VSANAEQFYQ YQPTEIILSE GQLVDSLGNG VAGVDYYTNS
Consensus   TSPVVPVTTP GTKPDLNASF VSANAEQFYQ YQPTEIILSE GRLVDSQGYG VAGVNYYTNS
            ########## ########## ########## ########## ########## ##########
            AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

            CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

            |          |          |          |          |          |
strain 101-1 GRGVTGENGK FNFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain 536   GRGVTGENGK FNFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain O42   GRGVTGENGK FNFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
str E2348/69 GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain 53638 GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain W3110 GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain B7A   GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain E22   GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
str E24377A  GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
str H10407   GRGVTGENGE FSFSWGEAIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
str E110019  GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain HS    GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain SECEC GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
```

```
str IHE3034   GRGVTDENGK FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
strain F11    GRGVTDENGK FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
 Consensus    GRGVTGENGE FSFSWGETIS FGIDTFELGS VRGNKSTIAL TELGDEVRGA NIDQLIHRYS
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  QAGKNDEREV PDVVRKVFAE YPNVINEIIN LSLSNGEALS EGDQTFERTN EFLEQFESGQ
strain 536    QAGKNDEREV PDVVRKVFAE YPNVINEIIN LSLSNGEALS EGDQTFERTN EFLEQFESGQ
strain O42    QAGKNDEREV PDVVRKVFAA YPNVINEIIN LSLSNGEALS EGDQTFERTN EFLEQFESGQ
str E2348/69  TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLD EGDQNVVLPN EFIEQFKTGQ
strain 53638  TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFNTGQ
strain W3110  TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFNTGQ
strain B7A    TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLD EGEQVVNLPN EFIEQFKTGQ
strain E22    TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLD EGEQVVNLPN EFIEQFKTGQ
str E24377A   KAGQNHTRVV PDEVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFKTGQ
str H10407    TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFNTGQ
str E110019   TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFKTGQ
strain HS     TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFNTGQ
strain SECEC  TTGQNNTRVV PEDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFNTGQ
str IHE3034   TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLD EGDQNVVLPN EFIEQFKTGQ
strain F11    TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLD EGDQNVVLPN EFIEQFKTGQ
 Consensus    TTGQNNTRVV PDDVRKVFAE YPNVINEIIN LSLSNGATLG EGEQVVNLPN EFIEQFKTGQ
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  AKEIDTAICD SLGGCNSQRW FSLTARNVNE GQIQGVINKL WGVDKDYKSV TKFHVFHDST
strain 536    AKEIDTAICD SLGGCNSQRW FSLTARNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain O42    AKEIDTAICD SLGGCNSQRW FSLTARNVNE GQIQGVINKL WGVDKDYKSV TKFHVFHDST
str E2348/69  AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDKDYKSV TKFHVFHDST
strain 53638  AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain W3110  AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain B7A    AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain E22    AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
str E24377A   AKEIDTAICA KTDGCNEARW FSLTTRNVND GKIQGVINKL WGVDTNYKSV SKFHVFHDST
str H10407    AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
str E110019   AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain HS     AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
strain SECEC  AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
str IHE3034   AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYQSV SKFHVFHDST
strain F11    AKEIDTAICA KTNGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYQSV SKFHVFHDST
 Consensus    AKEIDTAICA KTDGCNEARW FSLTTRNVND GQIQGVINKL WGVDTNYKSV SKFHVFHDST
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
strain 536    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
strain O42    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
str E2348/69  NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVQP
strain 53638  NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
strain W3110  NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
strain B7A    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
strain E22    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSLVEP
str E24377A   NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVRP
str H10407    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVRP
str E110019   NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNDLAY ITEAPSIVRP
```

```
strain HS     NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVRP
strain SECEC  NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVRP
str IHE3034   NFYGSTGNAR GQAVVNISNS AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVQP
strain F11    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVQP
 Consensus    NFYGSTGNAR GQAVVNISNA AFPILMARND KNYWLAFGEK RAWDKNELAY ITEAPSIVEP
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWEGGVD KNGQCTRNSD
strain 536    ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWEGGVD KNGQCTRNSD
strain O42    ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWEGGVD KNGQCTRNSD
str E2348/69  ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLSGD
strain 53638  ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD x
strain W3110  ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
strain B7A    ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
strain E22    ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
str E24377A   ENVTRETATF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
str H10407    ENVTRETASF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
str E110019   ENVTRETATF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
strain HS     ENVTRETATF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
strain SECEC  ENVTRETATF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
str IHE3034   ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWGGGVN SKGECTLSGD
strain F11    ENVTRDTATF NLPFISLGQV GDGKLMVIGN PHYNSILRCP NGYSWGGGVN SKGECTLSGD
 Consensus    ENVTRDTATF NLPFISLGQV GEGKLMVIGN PHYNSILRCP NGYSWNGGVN KDGQCTLNSD
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  SNDMKHFMQNV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA EFGFHPDFAG
strain 536    SNDMKHFMQNV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA EFGFHPDFAG
strain O42    SNDMKHFMQNV LRYLSNDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA EFGFHPDFAG
str E2348/69  SDDMKHFMQNV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA EFGFHPDFAG
strain 53638  SDDMKHFMQNV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFDFHPDFAG
strain W3110  PDDMKNFMENV LRYLSDDKW KPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFDFHPDFAG
strain B7A    PDDMKNFMENV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFDFHPDFAG
strain E22    PDDMKNFMENV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFDFHPDFAG
str E24377A   PDDMKNFMENV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFDFHPDFAG
str H10407    PDDMKNFMENV LRYLSNDRW LPDAKSSMTV GTNLETVYFK KHGQVLGNSA PFAFHKDFTG
str E110019   PDDMKNFMENV LRYLSNDRW LPDAKSSMTV GTNLDTVYFK KHGQVLGNSA PFAFHKDFTG
strain HS     PDDMKNFMENV LRYLSNDRW LPDAKSNMTV GTNLDTVYFK KHGQVTGNSA AFGFHPDFAG
strain SECEC  PDDMKNFMENV LRYLSNDRW LPDAKSNMTV GTNLDTVYFK KHGQVTGNSA AFGFHPDFAG
str IHE3034   SDDMKHFMQNV LRYLSNDIW QPNTKSIMTV GTNLENVYFK KAGQVLGNSA PFAFHEDFTG
strain F11    SDDMKHFMQNV LRYLSNDIW QPNTKSIMTV GTNLENVYFK KAGQVLGNSA PFAFHEDFTG
 Consensus    PDDMKNFMENV LRYLSDDKW TPDAKASMTV GTNLDTVYFK RHGQVTGNSA AFGFHPDFAG
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain 536    ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain O42    ISVEHLSSYG DLDPQKMPLL ILNGFEYVTQ VGGDPYAVPL RADTSKPKLS QQDVTDLIAY
str E2348/69  ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain 53638  ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain W3110  ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain B7A    ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain E22    ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
str E24377A   ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
```

```
str H10407    ITVKPMTSYG NLNPDEVPLL ILNGFEYVTQ WGSDPYSIPL RADTSKPKLT QQDVTDLIAY
str E110019   ITVKPMTSYG NLNPDEVPLL ILNGFEYVTQ WGSDPYSIPL RADTSKPKLT QQDVTDLIAY
strain HS     ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
strain SECEC  ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
str IHE3034   ITVKQLTSYG DLNPEEIPLL ILNGFEYVTQ WSGDPYAVPL RADTSKPKLT QQDVTDLIAY
strain F11    ITVKQLTSYG DLNPEEIPLL ILNGFEYVTQ WSGDPYAVPL RADTSKPKLT QQDVTDLIAY
 Consensus    ISVEHLSSYG DLDPQEMPLL ILNGFEYVTQ VGNDPYAIPL RADTSKPKLT QQDVTDLIAY
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
strain 536    LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
strain O42    LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRAT
str E2348/69  LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNTDPQGY PNRVRQQREK
strain 53638  LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
strain W3110  LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
strain B7A    LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
strain E22    LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
str E24377A   LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQRRST
str H10407    MNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRST
str E110019   MNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRST
strain HS     MNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRST
strain SECEC  MNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRST
str IHE3034   LNKGGSVLIM ENVMSNLKEE SASSFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQRRAT
strain F11    LNKGGSVLIM ENVMSNLKEE SASSFVRLLD AAGLSMALNK SVVNNDPQGY PDRVRQQREK
 Consensus    LNKGGSVLIM ENVMSNLKEE SASGFVRLLD AAGLSMALNK SVVNNDPQGY PNRVRQQRAT
              ########## ########## ########## ########## ########## ##########
              AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA
                                                          BBBBB BBBBBBBBBB
              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
strain 536    GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
strain O42    GIWVYERYPV VEGELP-YTI DSKTGKVTWK YQIDNKPDKK PKLEVASWQE EVDGKQVTQF
str E2348/69  GIWVYERYPA VDSAQPPYTI DPDTGKVTWK YQEEGKPDDK PKLEVASWQE DVDGKQVTRY
strain 53638  GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
strain W3110  GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
strain B7A    GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
strain E22    GIWVYERYPA VDGALP-YTI DSKTGEVKWK YQVENKPDDK PKLEVASWLE DVDGKQETRY
str E24377A   GIWVYERYPA VDGKPP-YTI DDTTKEVIWK YQQENKPDDK PKLEVASWQE EVEGKQVTQF
str H10407    GIWVYERYPA VDGKPP-YTI DDTTKEVIWK YQQENKPDDK PKLEVASWQE EVEGKQVTQF
str E110019   GIWVYERYPA VDGKPP-YTI DDTTKEVIWK YQQENKPDDK PKLEVASWQE EVEGKQVTQF
strain HS     GIWVYERYPA VDGKPP-YTI DDTTKEVIWK YQQENKPDDK PKLEVASWQE EVEGKQVTQF
strain SECEC  GIWVYERYPA VDGKPP-YTI DDTTKEVIWK YQQENKPDDK PKLEVASWQE EVEGKQVTQF
str IHE3034   GIWVYERYPA ADGAQPPYTI DPNTGEVTWK YQQDNKPDDK PKLEVASWQE EVEGKQVTRY
strain F11    GIWVYERYPF VDG-KPPYTI DETTKEVIWK YQQDNKPDDK PKLEVASWLE DVDGKQVKRY
 Consensus    GIWVYERYPA VDGALPPYTI DSKTGEV2WK YQQENKPDDK PKLEVASWQE DVDGKQVTRY
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  AFIDEADHKT EDSLKAAKAK IFEKFPGLKE CKDPTYHYEV NCLEYRPGTG VPVTGGMYVP
strain 536    AFIDEADHKT EDSLKAAKAK IFEKFPGLKE CKDPTYHYEV NCLEYRPGTG VPVTGGMYVP
strain O42    AFIDEADHKT TESLDAAKKK ILEKFKGLEE CKDSTYHYEI NCLEYRPGTN VPATGGMYVP
str E2348/69  AFIDEAEHST EESLEAAKAK IFEKFPGLQE CKDSTYHYEI NCLERRPGTD VPVTGGMYVP
strain 53638  AFIDEADHKT EDSLKAAKEK IFAAFPGLKE CTNPAYHYEV NCLEYRPGTG VPVTGGMYVP
strain W3110  AFIDEADHKT EDSLKAAKEK IFAAFPGLKE CTNPAYHYEV NCLEYRPGTG VPVTGGMYVP
strain B7A    AFIDEADHKT EDSLKAAKEK IFAAFPGLKE CTNPAYHYEV NCLEYRPGTG VPVTGGMYVP
```

```
strain E22    AFIDEADHKT EDSLKAAKEK IFAAFPGLKE CTNPAYHYEV NCLEYRPGTG VPVTGGMYVP
str E24377A   AFIDEADHKT PESLAAAKQR ILDAFPGLEV CKDSDYHYEV NCLEYRPGTD VPVTGGMYVP
str H10407    AFIDEADHKT PESLAAAKQR ILDAFPGLEV CKDSDYHYEV NCLEYRPGTD VPVTGGMYVP
str E110019   AFIDEADHKT PESLAAAKQR ILDAFPGLEV CKDSDYHYEV NCLEYRPGSG VPVTGGMYVP
strain HS     AFIDEADHKT PESLAAAKKR ILDAFPGLEE CKDSDYHYEV NCLEYRPGTG VPVTGGMYVP
strain SECEC  AFIDEADHKT PESLAAAKKR ILDAFPGLEE CKDSDYHYEV NCLEYRPGTG VPVTGGMYVP
str IHE3034   AFIDEAEYTT EESLEAAKAK IFEKFPGLQE CKDSTYHYEI NCLERRPGTD VPVTGGMYVP
strain F11    AFIDEAEHET NESLEAAKAK IIKAFPGLEE CKDPTYHYEV NCLEYRPGTN VPVTGGMYVP
 Consensus    AFIDEADHKT EESLKAAKAK IF2AFPGLEE CKDSTYHYEV NCLEYRPGTG VPVTGGMYVP
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain 536    QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain O42    RYTQLNLSAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
str E2348/69  RYTQLNLDAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain 53638  QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain W3110  QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain B7A    QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSCVDL ERLYQNMSVW
strain E22    QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
str E24377A   QYTQLDLSAD TAKAMLQAAD LGTNIQRLYQ HELYFRTNGR QGERLNSVDL ERLYQNMSVW
str H10407    QYTQLDLSAD TAKAMLQAAD LGTNIQRLYQ HELYFRTNGR QGERLNSVDL ERLYQNMSVW
str E110019   QYTQLDLGAD TAKAMLQAAD LGTNIQRLYQ HELYFRTNGR QGERLNSVDL ERLYQNMSVW
strain HS     QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
strain SECEC  QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
str IHE3034   RYTQLNLDAD TAKAMVQAAD LGTNIQRLYQ HELYFRTKGS KGERLNSVDL ERLYQNMSVW
strain F11    RYTQLNLSAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
 Consensus    QYTQLSLNAD TAKAMVQAAD LGTNIQRLYQ HELYFRTNGR KGERLSSVDL ERLYQNMSVW
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  LWNKIEYRYE NDKDDELGFK TFTEFLNCYA NDAYTGGTQC SDELKKSLVD NNMIYGEKSV
strain 536    LWNKIEYRYE NDKDDELGFK TFTEFLNCYA NDAYTGGTQC SDELKKSLVD NNMIYGEKSV
strain O42    LWNEIEYSYD SSKEDELGFK TFTEFLNCYA NDAYTGGTQC SDELKKSLVD NNMIYGEKSV
str E2348/69  LWNKIEYRYE NDKDDELGFK TFTEFLNCYA NNAYSEGTQC SADLKKSLVD NNMIYGDGSS
strain 53638  LWNDTSYRYE EGKNDELGFK TFTEFLNCYA NDAYAGGTKC SADLKKSLVD NNMIYGDGSS
strain W3110  LWNDTSYRYE EGKNDELGFK TFTEFLNCYA NDAYAGGTKC SADLKKSLVD NNMIYGDGSS
strain B7A    LWNDTSYRYE EGKNDELGFK TFTEFLNCYA NDAYAGGTKC SADLKKSLVD NNMIYGDGSS
strain E22    LWNDTSYRYE EGKNDELGFK TFTEFLNCYA NDAYAGGTKC SADLKKSLVD NNMIYGDGSS
str E24377A   LWNETKYRYE EGKEDELGFK TFTEFLNCYT NNAYVG-TQC SAELKKSLID NKMIYGEESS
str H10407    LWNETKYRYE EGKEDELGFK TFTEFLNCYT NNAYVG-TQC SAELKKSLID NKMIYGEESS
str E110019   LWNETKYRYE EGKEDELGFK TFTEFLNCYT NNAYVG-TQC SAELKKSLID NKMIYGEESS
strain HS     LWNKIEYRYE NDKDDELGFK TFTEFLNCYA NNAYDGGTQC SAELKQSLID NKMIYGE-GS
strain SECEC  LWNKIEYRYE NDKDDELGFK TFTEFLNCYA NNAYDGGTQC SAELKQSLID NKMIYGE-GS
str IHE3034   LWNDTKYRYE EGKEDELGFK TFTEFLNCYA NDAYAGGTKC SADLKKSLVD NNMIYGDGSS
strain F11    LWNEIEYSYD SSKEDELGFK TFTEFLNCYA NDAYTKGTLC SAELKQSLID NKMIYGEGS-
 Consensus    LWN3TEYRYE EGKEDELGFK TFTEFLNCYA NDAYAGGTQC SAELKKSLVD NNMIYGEGSS
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
              CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

              |          |          |          |          |          |
strain 101-1  NKAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSAEGEK VTETISLYSN
strain 536    NKAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSAEGEK VTETISLYSN
strain O42    NKAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQE VTESISLYSN
str E2348/69  -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSAEGEK VTETISLYSN
strain 53638  -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQN VTETISLYSN
```

```
strain W3110 -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQN VTETISLYSN
strain B7A   -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQN VTETISLYSN
strain E22   -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQN VTETISLYSN
str E24377A  -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGVVNTNGET VTQNINLYSA
str H10407   -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGVVNTNGET VTQNINLYSA
str E110019  -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSEEGQN VTETISLYSN
strain HS    -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSAEGEE VTETINLYSN
strain SECEC -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSAEGEE VTETINLYSN
str IHE3034  -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGSVSAKGES VTENISLYSN
strain F11   -KAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVSVGGEE VTETISLYSN
 Consensus   NKAGMMNPSY PLNYMEKPLT RLMLGRSWWD LNIKVDVEKY PGAVS2EGEN VTETISLYSN
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
             CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

             |          |          |          |          |          |
strain 101-1 PTKWFAGNMQ STGLWAPAQK EVTIESSASV PVTVTVALAD DLTGREKHEV ALNRPPKVTK
strain 536   PTKWFAGNMQ STGLWAPAQK EVTIESTASV AVTVTVALAD DLTGREKHEV ALNRPPKVTK
strain O42   PTKWFAGNMQ STGLWAPAQK EVTIKSNADV PVTVTVALAD DLTGREKHEV ALNRPPKVTK
str E2348/69 PTKWFAGNMQ STGLWAPAQQ EVTIESTASV PVTVTVALAD DLTGREKHEV ALNRPPKVTK
strain 53638 PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain W3110 PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain B7A   PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain E22   PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
str E24377A  PTKWFAGNMQ STGLWAPAQQ EVSIESKATV PVTVTVALAD DLTGREKHEV SLNRPPRVTK
str H10407   PTKWFAGNMQ STGLWAPAQQ EVSIESKSTV PVTVTVALAD DLTGREKHEV SLNRPPRVTK
str E110019  PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain HS    PTKWFAGNMQ STGLWAPAQQ EVSIKSNAKV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain SECEC PTKWFAGNMQ STGLWAPAQQ EVSIKSNAKV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
str IHE3034  PTKWFAGNMQ STGLWAPAQQ DVTIKSSASV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
strain F11   PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
 Consensus   PTKWFAGNMQ STGLWAPAQK EVTIKSNANV PVTVTVALAD DLTGREKHEV ALNRPPRVTK
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
             CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

             |          |          |          |          |          |
strain 101-1 TYELKANGEV KFTVPYGGLI YIKGNSPQN- ESAEFTFTGV VKAPFYKDGA WKNALNSPAP
strain 536   TYELKANGEV KFTVPYGGLI YIKGNSPQN- ESAEFTFTGV VKAPFYKDGA WKNALNSPAP
strain O42   TYELKANGEV KFTVPYGGLI YIKGNSKENN KSASFTFTGV VKAPFYKNGA WKNALNSPAP
str E2348/69 TYDLKANDKV TFKVPYGGLI YIKGNSPKN- ESAEFTFTGV VKAPFYKDGE WKNALNSPAP
strain 53638 TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
strain W3110 TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
strain B7A   TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
strain E22   TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
str E24377A  TYDLKANDKV TFKVPYGGLI YIKGDSKEV- QSADFTFTGV VKAPFYKDGK WQHDLNSPAP
str H10407   TYDLKANDKV TFKVPYGGLI YIKGDSKEV- QSADFTFTGV VKAPFYKDGK WQHDLNSPAP
str E110019  TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
strain HS    TYSLDASGTV KFKVPYGGLI YIKSDSKEE- KSANFTFTGV VKAPFYKDGK WKNDLKSPAP
strain SECEC TYSLDASGTV KFKVPYGGLI YIKSDSKEE- KSANFTFTGV VKAPFYKDGK WKNDLKSPAP
str IHE3034  TYTLEANGEV TFKVPYGGLI YIKGDSKDD- VSANFTFTGV VKAPFYKDGE WKNDLDSPAP
strain F11   TYSLDASGTV KFKVPYGGLI YIKGNSSTN- ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
 Consensus   TYSLDASGTV KFKVPYGGLI YIKGNS2TNN ESASFTFTGV VKAPFYKDGA WKNDLNSPAP
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
             CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

             |          |          |          |          |          |
strain 101-1 LGELESDAFV YTTPKKNLEA S---NFTGGV AEFAKDLDTF ASSMNDFYGR NDEDGKHRMF
strain 536   LGELESDAFV YTTPKKNLEA S---NYKGGQ EQFAEELDTF ASSMNDFYGR NDEDGKHRMF
strain O42   LGELESDAFV YTTPKKNLEA S---NFTGGV AEFAKDLDTF ASSMNDFYGR NDEDGKHRMF
```

```
str E2348/69 LGELESDSFV YTAPKNNLNA SNYSNYTDGV AEFAKELDTF ASSMNDFYGR DGESGNHRMF
strain 53638 LGELESDAFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFHGR DSEDGKHRMF
strain W3110 LGELESDAFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DSEDGKHRMF
strain B7A   LGELESDAFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DSEDGKHRMF
strain E22   LGELESDAFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DETSGKHRMF
str E24377A  LGELESASFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DSEDGKHRMF
str H10407   LGELESASFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DSEDGKHRMF
str E110019  LGELESDAFV YTTPKKNLNA S---NYTGGL EQFANDLDTF ASSMNDFYGR DSESGKHRMF
strain HS    LGELESASFV YTTPKKNLEA S---NYKGGL KQFAEDLDTF ASSMNDFYGR DGESGKHRMF
strain SECEC LGELESASFV YTTPKKNLEA S---NYKGGL KQFAEDLDTF ASSMNDFYGR DGESGKHRMF
str IHE3034  LGELESASFV YTTPKKNLEA S---NFTGGV AEFAKDLDTF ASSMNDFYGR NDEDGKHRMF
strain F11   LGELESASFV YTTPKKNLNA S---NYTGGL DQFAKDLDTF ASSMNDFYGR NDEDGKHRMF
 Consensus   LGELESDAFV YTTPKKNLNA SNYSNYTGGL EQFANDLDTF ASSMNDFYGR DSEDGKHRMF
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
             CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC

             |          |          |          |          |          |
strain 101-1 TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
strain 536   TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEAGHNAA
strain O42   TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
str E2348/69 TYKALTGHKH RFANDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
strain 53638 TYKNLPGHKH RFTNDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
strain W3110 TYKNLPGHKH RFTNDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
strain B7A   TYKNLPGHKH RFANDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
strain E22   TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
str E24377A  TYKNLPGHKH RFANDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
str H10407   TYKNLPGHKH RFANDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
str E110019  TYKNLTGHKH RFANDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
strain HS    TYEALTGHKH RFTNDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
strain SECEC TYEALTGHKH RFTNDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
str IHE3034  TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
strain F11   TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSTNSTT LPTTPLNDWL IWHEVGHNAA
 Consensus   TYKNLTGHKH RFTNDVQISI GDAHSGYPVM NSSFSPNSTT LPTTPLNDWL IWHEVGHNAA
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
             CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CCCCCCCCCC CC

             |          |          |          |          |          |
strain 101-1 ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
strain 536   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNNQA WARGGAGDRL
strain O42   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNNQA WARGGAGDRL
str E2348/69 ETPLNVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLDESNGQA WARGGAGDRL
strain 53638 ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNNQA WARGGAGDRL
strain W3110 ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNNQA WARGGAGDRL
strain B7A   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
strain E22   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
str E24377A  ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
str H10407   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
str E110019  ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
strain HS    ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
strain SECEC ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
str IHE3034  ETPLNVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLDESNGQA WARGGAGDRL
strain F11   ETPLNVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLDESNGQA WARGGAGDRL
 Consensus   ETPLTVPGAT EVANNVLALY MQDRYLGKMN RVADDITVAP EYLEESNGQA WARGGAGDRL
             ########## ########## ########## ########## ########## ##########

             BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB


             |          |          |          |          |          |
strain 101-1 LMYAQLKEWA EKNFDIKQWY PEGD-LPKFY SDREGMKGWN LFQLMHRKAR GDEVGKTKFG
```

```
strain 536    LMYAQLKEWA EKNFDITKWY PEGN-LPKFY SEREGMKGWN LFQLMHRKAR GDEVGKTKFG
strain O42    LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
str E2348/69  LMYAQLKEWA EENFDIKQWY PDGE-LPKFY SDRKGMKGWN LFQLMHRKAR GDDVSNDKFG
strain 53638  LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
strain W3110  LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
strain B7A    LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
strain E22    LMYAQLKEWA EKNFDIKKWY PEGE-LPKFF SDREGMKGWN LFQLMHRKAR GDDVGDKTFG
str E24377A   LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
str H10407    LMYAQLKEWA EKNFDIKKWY PDGTPLPEFY SEREGMKGWN LFQLMHRKAR GDEVSNDKFG
str E110019   LMYAQLKEWA EKNFDIKKWY PEGE-LPKFF SDREGMKGWN LFQLMHRKAR GDDVGNKTFG
strain HS     LMYAQLKEWA EKNFDIKQWY PEGS-LPAFY SEREGMKGWN LFQLMHRKAR GDDVGNDKFG
strain SECEC  LMYAQLKEWA EKNFDIKQWY PEGS-LPAFY SEREGMKGWN LFQLMHRKAR GDDVGNDKFG
str IHE3034   LMYAQLKEWA EENFDIKQWY PDGE-LPKFY SDRKGMKGWN LFQLMHRKAR GDDVGNSTFG
strain F11    LMYAQLKEWA EKNFDITKWY PDGK-LPAFY SEREGMKGWN LFQLMHRKAR GDDVGNSTFG
 Consensus    LMYAQLKEWA EKNFDIKKWY PDGTPLP2FY SEREGMKGWN LFQLMHRKAR GDEVGNDKFG
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB


              |          |          |          |          |          |
strain 101-1  ERNYCAESNG NAADKLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM NFEGGVSQSA
strain 536    ERNYCAESNG NAADTLMLCA SWVAQTDLSA FFKKWNPGAN AYQLPGASEM NFEGGVSQSA
strain O42    GRNYCAESNG NTADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGATEM SFEGGVSQSA
str E2348/69  GRNYCAESNG NAADTLMLCA SWVAQADLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
strain 53638  GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
strain W3110  GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
strain B7A    GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
strain E22    GKNYCAESNG NAADTLMLCA SWVAQTDLSA FFKKWNPGAN AYQLPGATEM SFEGGVSQSA
str E24377A   GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
str H10407    GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
str E110019   GKNYCAESNG NAADSLMLCA SWVAQTDLSA FFKKWNPGAN AYQLPGATEM SFEGGVSQSA
strain HS     NRNYCAESNG NAADTLMLCA SWVAQTDLSA FFKKWNPGAN AYQLPGATEM SFEGGVSQSA
strain SECEC  NRNYCAESNG NAADTLMLCA SWVAQTDLSA FFKKWNPGAN AYQLPGATEM SFEGGVSQSA
str IHE3034   GKNYCAESNG NAADTLMLCA SWVAQADLSE FFKKWNPGAS AYQLPGATEM SFQGGVSSSA
strain F11    GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGAAEM SFEGGVSSSA
 Consensus    GKNYCAESNG NAADTLMLCA SWVAQTDLSE FFKKWNPGAN AYQLPGASEM SFEGGVSQSA
              ########## ########## ########## ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB


              |          |          |          |          |          |
strain 101-1  YETLAALNLP KPQQGPETIN QVTEHKMSAE
strain 536    YETLAALNLP KPQQGPETIN KVTEYSMPAE
strain O42    YNTLASLDLP KPKQGPETIN KVTEYSMPAE
str E2348/69  YNTLAAMHLS KPEKGPETIN KVTEYSMPAE
strain 53638  YNTLASLDLP KPEQGPETIN QVTEHKMSAE
strain W3110  YNTLASLDLP KPEQGPETIN QVTEHKMSAE
strain B7A    YNTLASLKLP KPEQGPETIN KVTEHKMSVE
strain E22    YSTLASLKLP KPEQGPETIN KVTEHKMSLE
str E24377A   YNTLASLKLP KPEQGPETIN KVTEHKMSVE
str H10407    YNTLASLDLP KPEQGPETIN QVTEHKMSAE
str E110019   YSTLASLKLP KPEQGPETIN KVTEHKMSLE
strain HS     YNTLASLDLP KPKQGPETIN KVTEYSMPAE
strain SECEC  YNTLASLDLP KPEQGPETIN QVTEHKMSAE
str IHE3034   YSTLASLKLP KPEKGPETIN KVTEHKMSAE
strain F11    YSTLASLNLP KPEKGPETIN KVTEHKMSAE
 Consensus    YNTLASLDLP KPEQGPETIN KVTEHKMSAE
              ########## ########## ##########

              BBBBBBBBBB BBBBBBBBBB BBBBBBBBBB
```

In certain embodiments, the carrier polypeptide comprising an *E. coli* AcfD (orf3526) polypeptide will have a mutation relative to the *E. coli* AcfD (orf3526) protein which decreases the toxicity of the carrier polypeptide as compared to the *E. coli* AcfD (orf3526) protein.

Exemplary mutations that decrease the toxicity include a deletion of all or a portion of the zincin metalloprotease domain and a point mutation in zincin metalloprotease domain which reduces the protease activity. In certain cases, the point mutation is a mutation of a zinc binding residue or a mutation of a catalytic residue. A preferred point mutation is substitution of amino acid number 1305 based upon alignment with SEQ ID NO: 39.

Exemplary deletions such as for orf3526A include removal of at least the last 100 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 300 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 400 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 500 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 600 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 700 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 750 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the last 758 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein or such as for orf3526B does not comprise at least the first 100 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 200 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 300 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 400 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 500 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 600 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 700 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 750 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the first 760 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein.

Exemplary fragments such as orf3526C do not comprise at least the last 100 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 125 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 150 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 175 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 210 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the last 217 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein and optionally do not comprise at least the first 10 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 20 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 30 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the first 33 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein.

The foregoing carrier polypeptides may further contain a deletion relative to the *E. coli* AcfD (orf3526) protein which increases solubility of the carrier polypeptide as compared to the full length *E. coli* AcfD (orf3526) protein while the carrier polypeptide still raises a substantially similar immune response to the conjugated polysaccharide in a subject as the full length *E. coli* AcfD (orf3526) protein.

Exemplary deletions that increase the solubility include removal of substantially all of the N-terminal amino acids up to the gly-ser region, removal of all or a part of the N-terminal proline-rich repeat, or both. Furthermore, the deletion may include the removal of at least the first 20 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 20 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 30 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 33 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 40 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 50 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 60 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 70 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 80 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 90 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, or at least the first 94 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein.

In addition to improving the solubility and lowering the toxicity, the deletions are easier to purify after expression in commensal *E. coli* strains that have an endogenous AcfD (orf3526) protein.

Immunogenic compositions and medicaments

Polysaccharide conjugates of the invention are useful as active ingredients (immunogens) in immunogenic compositions, and such compositions may be useful as vaccines. Vaccines according to the invention may either be prophylactic (*i.e.* to prevent infection) or therapeutic (*i.e.* to treat infection), but will typically be prophylactic.

Immunogenic compositions will be pharmaceutically acceptable. They will usually include components in addition to the antigens *e.g.* they typically include one or more pharmaceutical carrier(s), excipient(s) and/or adjuvant(s). A thorough discussion of carriers and excipients is available in ref. 100. Thorough discussions of vaccine adjuvants are available in refs. 5 and 6.

Compositions will generally be administered to a mammal in aqueous form. Prior to administration, however, the composition may have been in a non-aqueous form. For instance, although some vaccines are manufactured in aqueous form, then filled and distributed and administered also in aqueous form, other vaccines are lyophilized during manufacture and are reconstituted into an aqueous form at the time of use. Thus a composition of the invention may be dried, such as a lyophilized formulation.

The composition may include preservatives such as thiomersal or 2-phenoxyethanol. It is preferred, however, that the vaccine should be substantially free from (*i.e.* less than 5μg/ml) mercurial material *e.g.* thiomersal-free. Vaccines containing no mercury are more preferred. Preservative-free vaccines are particularly preferred.

To improve thermal stability, a composition may include a temperature protective agent.

To control tonicity, it is preferred to include a physiological salt, such as a sodium salt. Sodium chloride (NaCl) is preferred, which may be present at between 1 and 20 mg/ml *e.g.* about 10±2mg/ml NaCl. Other salts that may be present include potassium chloride, potassium dihydrogen phosphate, disodium phosphate dehydrate, magnesium chloride, calcium chloride, *etc.*

Compositions will generally have an osmolality of between 200 mOsm/kg and 400 mOsm/kg, preferably between 240-360 mOsm/kg, and will more preferably fall within the range of 290-310 mOsm/kg.

Compositions may include one or more buffers. Typical buffers include: a phosphate buffer; a Tris buffer; a borate buffer; a succinate buffer; a histidine buffer (particularly with an aluminum hydroxide adjuvant); or a citrate buffer. Buffers will typically be included in the 5-20mM range.

The pH of a composition will generally be between 5.0 and 8.1, and more typically between 6.0 and 8.0 *e.g.* 6.5 and 7.5, or between 7.0 and 7.8.

The composition is preferably sterile. The composition is preferably non-pyrogenic *e.g.* containing <1 EU (endotoxin unit, a standard measure) per dose, and preferably <0.1 EU per dose. The composition is preferably gluten free.

The composition may include material for a single immunization, or may include material for multiple immunizations (*i.e.* a 'multidose' kit). The inclusion of a preservative is preferred in multidose arrangements. As an alternative (or in addition) to including a preservative in multidose compositions, the compositions may be contained in a container having an aseptic adaptor for removal of material.

Human vaccines are typically administered in a dosage volume of about 0.5ml, although a half dose (*i.e.* about 0.25ml) may be administered to children.

Immunogenic compositions of the invention may also comprise one or more immunoregulatory agents. Preferably, one or more of the immunoregulatory agents include one or more adjuvants. The adjuvants may include a TH1 adjuvant and/or a TH2 adjuvant, further discussed below.

Adjuvants which may be used in compositions of the invention include, but are not limited to:

A. Mineral-containing compositions

Mineral containing compositions suitable for use as adjuvants in the invention include mineral salts, such as aluminum salts and calcium salts (or mixtures thereof). Calcium salts include calcium phosphate (*e.g.* the "CAP" particles disclosed in ref. 7). Aluminum salts include hydroxides, phosphates, sulfates, *etc.*, with the salts taking any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*). Adsorption to these salts is preferred. The mineral containing compositions may also be formulated as a particle of metal salt (8).

The adjuvants known as aluminum hydroxide and aluminum phosphate may be used. These names are conventional, but are used for convenience only, as neither is a precise description of the actual chemical compound which is present (*e.g.* see chapter 9 of reference 5). The invention can use any of the "hydroxide" or "phosphate" adjuvants that are in general use as adjuvants. The adjuvants known as "aluminum hydroxide" are typically aluminum oxyhydroxide salts, which are usually at least partially crystalline. The adjuvants known as "aluminum phosphate" are typically aluminum hydroxyphosphates, often also containing a small amount of sulfate (*i.e.* aluminum hydroxyphosphate sulfate). They may be obtained by precipitation, and the reaction conditions and concentrations during precipitation influence the degree of substitution of phosphate for hydroxyl in the salt.

A fibrous morphology (*e.g.* as seen in transmission electron micrographs) is typical for aluminum hydroxide adjuvants. The pI of aluminum hydroxide adjuvants is typically about 11 *i.e.* the adjuvant itself has a positive surface charge at physiological pH. Adsorptive capacities of between 1.8-2.6 mg protein per mg Al^{+++} at pH 7.4 have been reported for aluminum hydroxide adjuvants.

Aluminum phosphate adjuvants generally have a PO_4/Al molar ratio between 0.3 and 1.2, preferably between 0.8 and 1.2, and more preferably 0.95±0.1. The aluminum phosphate will generally be amorphous, particularly for hydroxyphosphate salts. A typical adjuvant is amorphous aluminum hydroxyphosphate with PO_4/Al molar ratio between 0.84 and 0.92, included at 0.6mg Al^{3+}/ml. The aluminum phosphate will generally be particulate (*e.g.* plate-like morphology as seen in transmission electron micrographs). Typical diameters of the particles are in the range 0.5-20μm (*e.g.* about 5-10μm) after any antigen adsorption. Adsorptive capacities of between 0.7-1.5 mg protein per mg Al^{+++} at pH 7.4 have been reported for aluminum phosphate adjuvants.

The point of zero charge (PZC) of aluminum phosphate is inversely related to the degree of substitution of phosphate for hydroxyl, and this degree of substitution can vary depending on reaction conditions and concentration of reactants used for preparing the salt by precipitation. PZC is also altered by changing the concentration of free phosphate ions in solution (more phosphate = more acidic PZC) or by adding a buffer such as a histidine buffer (makes PZC more

basic). Aluminum phosphates used according to the invention will generally have a PZC of between 4.0 and 7.0, more preferably between 5.0 and 6.5 *e.g.* about 5.7.

Suspensions of aluminum salts used to prepare compositions of the invention may contain a buffer (*e.g.* a phosphate or a histidine or a Tris buffer), but this is not always necessary. The suspensions are preferably sterile and pyrogen-free. A suspension may include free aqueous phosphate ions *e.g.* present at a concentration between 1.0 and 20 mM, preferably between 5 and 15 mM, and more preferably about 10 mM. The suspensions may also comprise sodium chloride.

The invention can use a mixture of both an aluminum hydroxide and an aluminum phosphate. In this case there may be more aluminum phosphate than hydroxide *e.g.* a weight ratio of at least 2:1 *e.g.* ≥5:1, ≥6:1, ≥7:1, ≥8:1, ≥9:1, *etc.*

The concentration of Al^{+++} in a composition for administration to a patient is preferably less than 10mg/ml *e.g.* ≤5 mg/ml, ≤4 mg/ml, ≤3 mg/ml, ≤2 mg/ml, ≤1 mg/ml, *etc.* A preferred range is between 0.3 and 1mg/ml. A maximum of 0.85mg/dose is preferred.

B. Oil Emulsions

Oil emulsion compositions suitable for use as adjuvants in the invention include squalene-water emulsions, such as MF59(TM) (Chapter 10 of ref. 5; see also ref. 9) (5% Squalene, 0.5% TWEEN 80(TM), and 0.5% SPAN 85(TM), formulated into submicron particles using a microfluidizer). Complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA) may also be used.

Various oil-in-water emulsion adjuvants are known, and they typically include at least one oil and at least one surfactant, with the oil(s) and surfactant(s) being biodegradable (metabolisable) and biocompatible. The oil droplets in the emulsion are generally less than 5μm in diameter, and ideally have a sub-micron diameter, with these small sizes being achieved with a microfluidiser to provide stable emulsions. Droplets with a size less than 220nm are preferred as they can be subjected to filter sterilization.

The emulsion can comprise oils such as those from an animal (such as fish) or vegetable source. Sources for vegetable oils include nuts, seeds and grains. Peanut oil, soybean oil, coconut oil, and olive oil, the most commonly available, exemplify the nut oils. Jojoba oil can be used *e.g.* obtained from the jojoba bean. Seed oils include safflower oil, cottonseed oil, sunflower seed oil, sesame seed oil and the like. In the grain group, corn oil is the most readily available, but the oil of other cereal grains such as wheat, oats, rye, rice, teff, triticale and the like may also be used. 6-10 carbon fatty acid esters of glycerol and 1,2-propanediol, while not occurring naturally in seed oils, may be prepared by hydrolysis, separation and esterification of the appropriate materials starting from the nut and seed oils. Fats and oils from mammalian milk are metabolizable and may therefore be used in the practice of this invention. The procedures for separation,

purification, saponification and other means necessary for obtaining pure oils from animal sources are well known in the art. Most fish contain metabolizable oils which may be readily recovered. For example, cod liver oil, shark liver oils, and whale oil such as spermaceti exemplify several of the fish oils which may be used herein. A number of branched chain oils are synthesized biochemically in 5-carbon isoprene units and are generally referred to as terpenoids. Shark liver oil contains a branched, unsaturated terpenoids known as squalene, 2,6,10,15,19,23-hexamethyl-2,6,10,14,18,22-tetracosahexaene, which is particularly preferred herein. Squalane, the saturated analog to squalene, is also a preferred oil. Fish oils, including squalene and squalane, are readily available from commercial sources or may be obtained by methods known in the art. Other preferred oils are the tocopherols (see below). Mixtures of oils can be used.

Surfactants can be classified by their 'HLB' (hydrophile/lipophile balance). Preferred surfactants of the invention have a HLB of at least 10, preferably at least 15, and more preferably at least 16. The invention can be used with surfactants including, but not limited to: the polyoxyethylene sorbitan esters surfactants (commonly referred to as the Tweens), especially polysorbate 20 and polysorbate 80; copolymers of ethylene oxide (EO), propylene oxide (PO), and/or butylene oxide (BO), sold under the DOWFAX™ tradename, such as linear EO/PO block copolymers; octoxynols, which can vary in the number of repeating ethoxy (oxy-1,2-ethanediyl) groups, with octoxynol-9 (Triton X-100, or t-octylphenoxypolyethoxyethanol) being of particular interest; (octylphenoxy)polyethoxyethanol (IGEPAL CA-630/NP-40); phospholipids such as phosphatidylcholine (lecithin); nonylphenol ethoxylates, such as the Tergitol™ NP series; polyoxyethylene fatty ethers derived from lauryl, cetyl, stearyl and oleyl alcohols (known as Brij surfactants), such as triethyleneglycol monolauryl ether (Brij 30); and sorbitan esters (commonly known as the SPANs), such as sorbitan trioleate (SPAN 85(TM)) and sorbitan monolaurate. Non-ionic surfactants are preferred. Preferred surfactants for including in the emulsion are TWEEN 80(TM) (polyoxyethylene sorbitan monooleate), SPAN 85(TM) (sorbitan trioleate), lecithin and Triton X-100.

Mixtures of surfactants can be used *e.g.* TWEEN 80(TM)/SPAN 85(TM) mixtures. A combination of a polyoxyethylene sorbitan ester such as polyoxyethylene sorbitan monooleate (TWEEN 80(TM)) and an octoxynol such as t-octylphenoxypolyethoxyethanol (Triton X-100) is also suitable. Another useful combination comprises laureth 9 plus a polyoxyethylene sorbitan ester and/or an octoxynol.

Preferred amounts of surfactants (% by weight) are: polyoxyethylene sorbitan esters (such as TWEEN 80(TM)) 0.01 to 1%, in particular about 0.1 %; octyl- or nonylphenoxy polyoxyethanols (such as Triton X-100, or other detergents in the Triton series) 0.001 to 0.1 %, in particular 0.005 to 0.02%; polyoxyethylene ethers (such as laureth 9) 0.1 to 20 %, preferably 0.1 to 10 % and in particular 0.1 to 1 % or about 0.5%.

Preferred emulsion adjuvants have an average droplets size of <1μm *e.g.* ≤750nm, ≤500nm, ≤400nm, ≤300nm, ≤250nm, ≤220nm, ≤200nm, or smaller. These droplet sizes can conveniently be achieved by techniques such as microfluidisation.

Specific oil-in-water emulsion adjuvants useful with the invention include, but are not limited to:

- A submicron emulsion of squalene, TWEEN 80(TM), and SPAN 85(TM). The composition of the emulsion by volume can be about 5% squalene, about 0.5% polysorbate 80 and about 0.5% SPAN 85(TM). In weight terms, these ratios become 4.3% squalene, 0.5% polysorbate 80 and 0.48% SPAN 85(TM). This adjuvant is known as 'MF59(TM)' (10-11), as described in more detail in Chapter 10 of ref. 12 and chapter 12 of ref. 13. The MF59(TM) emulsion advantageously includes citrate ions *e.g.* 10mM sodium citrate buffer.
- An emulsion of squalene, a tocopherol, and TWEEN 80(TM). The emulsion may include phosphate buffered saline. It may also include SPAN 85(TM) (*e.g.* at 1%) and/or lecithin. These emulsions may have from 2 to 10% squalene, from 2 to 10% tocopherol and from 0.3 to 3% TWEEN 80(TM), and the weight ratio of squalene:tocopherol is preferably ≤1 as this provides a more stable emulsion. Squalene and TWEEN 80(TM) may be present volume ratio of about 5:2. One such emulsion can be made by dissolving TWEEN 80(TM) in PBS to give a 2% solution, then mixing 90ml of this solution with a mixture of (5g of DL-α-tocopherol and 5ml squalene), then microfluidising the mixture. The resulting emulsion may have submicron oil droplets *e.g.* with an average diameter of between 100 and 250nm, preferably about 180nm.
- An emulsion of squalene, a tocopherol, and a Triton detergent (*e.g.* Triton X-100). The emulsion may also include a 3d-MPL (see below). The emulsion may contain a phosphate buffer.
- An emulsion comprising a polysorbate (*e.g.* polysorbate 80), a Triton detergent (*e.g.* Triton X-100) and a tocopherol (*e.g.* an α-tocopherol succinate). The emulsion may include these three components at a mass ratio of about 75:11:10 (*e.g.* 750μg/ml polysorbate 80, 110μg/ml Triton X-100 and 100μg/ml α-tocopherol succinate), and these concentrations should include any contribution of these components from antigens. The emulsion may also include squalene. The emulsion may also include a 3d-MPL (see below). The aqueous phase may contain a phosphate buffer.
- An emulsion of squalane, polysorbate 80 and poloxamer 401 ("PLURONIC(TM) L121"). The emulsion can be formulated in phosphate buffered saline, pH 7.4. This emulsion is a useful delivery vehicle for muramyl dipeptides, and has been used with threonyl-MDP in the "SAF-1" adjuvant (14) (0.05-1% Thr-MDP, 5% squalane, 2.5% PLURONIC(TM) L121 and 0.2% polysorbate 80). It can also be used without the Thr-MDP, as in the "AF"

adjuvant (15) (5% squalane, 1.25% PLURONIC(TM) L121 and 0.2% polysorbate 80). Microfluidisation is preferred.

- An emulsion comprising squalene, an aqueous solvent, a polyoxyethylene alkyl ether hydrophilic nonionic surfactant (*e.g.* polyoxyethylene (12) cetostearyl ether) and a hydrophobic nonionic surfactant (*e.g.* a sorbitan ester or mannide ester, such as sorbitan monoleate or 'SPAN 80(TM)'). The emulsion is preferably thermoreversible and/or has at least 90% of the oil droplets (by volume) with a size less than 200 nm (16). The emulsion may also include one or more of: alditol; a cryoprotective agent (*e.g.* a sugar, such as dodecylmaltoside and/or sucrose); and/or an alkylpolyglycoside. Such emulsions may be lyophilized.

- An emulsion of squalene, poloxamer 105 and Abil-Care (17). The final concentration (weight) of these components in adjuvanted vaccines are 5% squalene, 4% poloxamer 105 (pluronic polyol) and 2% Abil-Care 85 (Bis-PEG/PPG-16/16 PEG/PPG-16/16 dimethicone; caprylic/capric triglyceride).

- An emulsion having from 0.5-50% of an oil, 0.1-10% of a phospholipid, and 0.05-5% of a non-ionic surfactant. As described in reference 18, preferred phospholipid components are phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, phosphatidylglycerol, phosphatidic acid, sphingomyelin and cardiolipin. Submicron droplet sizes are advantageous.

- A submicron oil-in-water emulsion of a non-metabolisable oil (such as light mineral oil) and at least one surfactant (such as lecithin, TWEEN 80(TM) or SPAN 80(TM)). Additives may be included, such as QuilA saponin, cholesterol, a saponin-lipophile conjugate (such as GPI-0100, described in reference 19, produced by addition of aliphatic amine to desacylsaponin via the carboxyl group of glucuronic acid), dimethyidioctadecylammonium bromide and/or N,N-dioctadecyl-N,N-bis (2-hydroxyethyl)propanediamine.

- An emulsion in which a saponin (*e.g.* QuilA or QS21) and a sterol (*e.g.* a cholesterol) are associated as helical micelles (20).

- An emulsion comprising a mineral oil, a non-ionic lipophilic ethoxylated fatty alcohol, and a non-ionic hydrophilic surfactant (*e.g.* an ethoxylated fatty alcohol and/or polyoxyethylene-polyoxypropylene block copolymer) (21).

- An emulsion comprising a mineral oil, a non-ionic hydrophilic ethoxylated fatty alcohol, and a non-ionic lipophilic surfactant (*e.g.* an ethoxylated fatty alcohol and/or polyoxyethylene-polyoxypropylene block copolymer) (21).

In some embodiments an emulsion may be mixed with antigen extemporaneously, at the time of delivery, and thus the adjuvant and antigen may be kept separately in a packaged or distributed

vaccine, ready for final formulation at the time of use. In other embodiments an emulsion is mixed with antigen during manufacture, and thus the composition is packaged in a liquid adjuvanted form,. The antigen will generally be in an aqueous form, such that the vaccine is finally prepared by mixing two liquids. The volume ratio of the two liquids for mixing can vary (*e.g.* between 5:1 and 1:5) but is generally about 1:1. Where concentrations of components are given in the above descriptions of specific emulsions, these concentrations are typically for an undiluted composition, and the concentration after mixing with an antigen solution will thus decrease.

Where a composition includes a tocopherol, any of the α, β, γ, δ, ε or ξ tocopherols can be used, but α-tocopherols are preferred. The tocopherol can take several forms *e.g.* different salts and/or isomers. Salts include organic salts, such as succinate, acetate, nicotinate, *etc.* D-α-tocopherol and DL-α-tocopherol can both be used. Tocopherols are advantageously included in vaccines for use in elderly patients (*e.g.* aged 60 years or older) because vitamin E has been reported to have a positive effect on the immune response in this patient group (22). They also have antioxidant properties that may help to stabilize the emulsions (23). A preferred α-tocopherol is DL-α-tocopherol, and the preferred salt of this tocopherol is the succinate. The succinate salt has been found to cooperate with TNF-related ligands *in vivo*.

C. Saponin formulations (chapter 22 of ref. 5)

Saponin formulations may also be used as adjuvants in the invention. Saponins are a heterogeneous group of sterol glycosides and triterpenoid glycosides that are found in the bark, leaves, stems, roots and even flowers of a wide range of plant species. Saponin from the bark of the *Quillaia saponaria* Molina tree have been widely studied as adjuvants. Saponin can also be commercially obtained from *Smilax ornata* (sarsaprilla), *Gypsophilla paniculata* (brides veil), and *Saponaria officianalis* (soap root). Saponin adjuvant formulations include purified formulations, such as QS21, as well as lipid formulations, such as ISCOMs. QS21 is marketed as Stimulon™.

Saponin compositions have been purified using HPLC and RP-HPLC. Specific purified fractions using these techniques have been identified, including QS7, QS17, QS18, QS21, QH-A, QH-B and QH-C. Preferably, the saponin is QS21. A method of production of QS21 is disclosed in ref. 24. Saponin formulations may also comprise a sterol, such as cholesterol (25).

Combinations of saponins and cholesterols can be used to form unique particles called immunostimulating complexs (ISCOMs) (chapter 23 of ref. 5). ISCOMs typically also include a phospholipid such as phosphatidylethanolamine or phosphatidylcholine. Any known saponin can be used in ISCOMs. Preferably, the ISCOM includes one or more of QuilA, QHA & QHC. ISCOMs are further described in refs. 25-26. Optionally, the ISCOMS may be devoid of additional detergent (27).

A review of the development of saponin based adjuvants can be found in refs. 28 & 29.

D. Virosomes and virus-like particles

Virosomes and virus-like particles (VLPs) can also be used as adjuvants in the invention. These structures generally contain one or more proteins from a virus optionally combined or formulated with a phospholipid. They are generally non-pathogenic, non-replicating and generally do not contain any of the native viral genome. The viral proteins may be recombinantly produced or isolated from whole viruses. These viral proteins suitable for use in virosomes or VLPs include proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, Qß-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in refs. 30-31. Virosomes are discussed further in, for example, ref. 32

E. Bacterial or microbial derivatives

Adjuvants suitable for use in the invention include bacterial or microbial derivatives such as non-toxic derivatives of enterobacterial lipopolysaccharide (LPS), Lipid A derivatives, immunostimulatory oligonucleotides and ADP-ribosylating toxins and detoxified derivatives thereof.

Non-toxic derivatives of LPS include monophosphoryl lipid A (MPL) and 3-O-deacylated MPL (3dMPL). 3dMPL is a mixture of 3 de-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. A preferred "small particle" form of 3 De-O-acylated monophosphoryl lipid A is disclosed in ref. 33. Such "small particles" of 3dMPL are small enough to be sterile filtered through a 0.22μm membrane (33). Other non-toxic LPS derivatives include monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives *e.g.* RC-529 (34,35).

Lipid A derivatives include derivatives of lipid A from *Escherichia coli* such as OM-174. OM-174 is described for example in refs. 36 & 37.

Immunostimulatory oligonucleotides suitable for use as adjuvants in the invention include nucleotide sequences containing a CpG motif (a dinucleotide sequence containing an unmethylated cytosine linked by a phosphate bond to a guanosine). Double-stranded RNAs and oligonucleotides containing palindromic or poly(dG) sequences have also been shown to be immunostimulatory.

The CpG's can include nucleotide modifications/analogs such as phosphorothioate modifications and can be double-stranded or single-stranded. References 38, 39 and 40 disclose possible analog substitutions *e.g.* replacement of guanosine with 2'-deoxy-7-deazaguanosine. The adjuvant effect of CpG oligonucleotides is further discussed in refs. 41-42.

The CpG sequence may be directed to TLR9, such as the motif GTCGTT or TTCGTT (43). The CpG sequence may be specific for inducing a Th1 immune response, such as a CpG-A ODN, or it may be more specific for inducing a B cell response, such a CpG-B ODN. CpG-A and CpG-B ODNs are discussed in refs. 44-45. Preferably, the CpG is a CpG-A ODN.

Preferably, the CpG oligonucleotide is constructed so that the 5' end is accessible for receptor recognition. Optionally, two CpG oligonucleotide sequences may be attached at their 3' ends to form "immunomers". See, for example, refs. 43 & 46-47.

A useful CpG adjuvant is CpG7909, also known as ProMune™ (Coley Pharmaceutical Group, Inc.). Another is CpG1826. As an alternative, or in addition, to using CpG sequences, TpG sequences can be used (48), and these oligonucleotides may be free from unmethylated CpG motifs. The immunostimulatory oligonucleotide may be pyrimidine-rich. For example, it may comprise more than one consecutive thymidine nucleotide (*e.g.* TTTT, as disclosed in ref. 48), and/or it may have a nucleotide composition with >25% thymidine (*e.g.* >35%, >40%, >50%, >60%, >80%, *etc.*). For example, it may comprise more than one consecutive cytosine nucleotide (*e.g.* CCCC, as disclosed in ref. 48), and/or it may have a nucleotide composition with >25% cytosine (*e.g.* >35%, >40%, >50%, >60%, >80%, *etc.*). These oligonucleotides may be free from unmethylated CpG motifs. Immunostimulatory oligonucleotides will typically comprise at least 20 nucleotides. They may comprise fewer than 100 nucleotides.

A particularly useful adjuvant based around immunostimulatory oligonucleotides is known as IC-31™ (49). Thus an adjuvant used with the invention may comprise a mixture of (i) an oligonucleotide (*e.g.* between 15-40 nucleotides) including at least one (and preferably multiple) CpI motifs (*i.e.* a cytosine linked to an inosine to form a dinucleotide), and (ii) a polycationic polymer, such as an oligopeptide (*e.g.* between 5-20 amino acids) including at least one (and preferably multiple) Lys-Arg-Lys tripeptide sequence(s). The oligonucleotide may be a deoxynucleotide comprising 26-mer sequence 5'-$(IC)_{13}$-3' (SEQ ID NO: 51). The polycationic polymer may be a peptide comprising 11-mer amino acid sequence KLKLLLLLKLK (SEQ ID NO: 52).

Bacterial ADP-ribosylating toxins and detoxified derivatives thereof may be used as adjuvants in the invention. Preferably, the protein is derived from *E. coli* (*E. coli* heat labile enterotoxin "LT"), cholera ("CT"), or pertussis ("PT"). The use of detoxified ADP-ribosylating toxins as mucosal adjuvants is described in ref. 50 and as parenteral adjuvants in ref. 51. The toxin or toxoid is preferably in the form of a holotoxin, comprising both A and B subunits. Preferably, the A subunit contains a detoxifying mutation; preferably the B subunit is not mutated. Preferably, the adjuvant is a detoxified LT mutant such as LT-K63, LT-R72, and LT-G192. The use of ADP-ribosylating toxins and detoxified derivatives thereof, particularly LT-K63 and LT-R72, as adjuvants can be found in refs. 52-53. A useful CT mutant is or CT-E29H (54). Numerical

reference for amino acid substitutions is preferably based on the alignments of the A and B subunits of ADP-ribosylating toxins set forth in ref. 55, specifically incorporated herein by reference in its entirety solely for the purpose of the alignment and amino acid numbering therein.

F. Human immunomodulators

Human immunomodulators suitable for use as adjuvants in the invention include cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 (56), *etc.*) (57), interferons (*e.g.* interferon-γ), macrophage colony stimulating factor, and tumor necrosis factor. A preferred immunomodulator is IL-12.

G. Bioadhesives and Mucoadhesives

Bioadhesives and mucoadhesives may also be used as adjuvants in the invention. Suitable bioadhesives include esterified hyaluronic acid microspheres (58) or mucoadhesives such as cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulose. Chitosan and derivatives thereof may also be used as adjuvants in the invention (59).

H. Microparticles

Microparticles may also be used as adjuvants in the invention. Microparticles (*i.e.* a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (*e.g.* a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, *etc.*), with poly(lactide-co-glycolide) are preferred, optionally treated to have a negatively-charged surface (*e.g.* with SDS) or a positively-charged surface (*e.g.* with a cationic detergent, such as CTAB).

I. Liposomes (Chapters 13 & 14 of ref. 5)

Examples of liposome formulations suitable for use as adjuvants are described in refs. 60-61.

J. Polyoxyethylene ether and polyoxyethylene ester formulations

Adjuvants suitable for use in the invention include polyoxyethylene ethers and polyoxyethylene esters (62). Such formulations further include polyoxyethylene sorbitan ester surfactants in combination with an octoxynol (63) as well as polyoxyethylene alkyl ethers or ester surfactants in combination with at least one additional non-ionic surfactant such as an octoxynol (64). Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether (laureth 9), polyoxyethylene-9-steoryl ether, polyoxytheylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether.

K. Phosphazenes

A phosphazene, such as poly(di(carboxylatophenoxy)phosphazene) ("PCPP") as described, for example, in references 65 and 66, may be used.

L. Muramyl peptides

Examples of muramyl peptides suitable for use as adjuvants in the invention include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), and N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE).

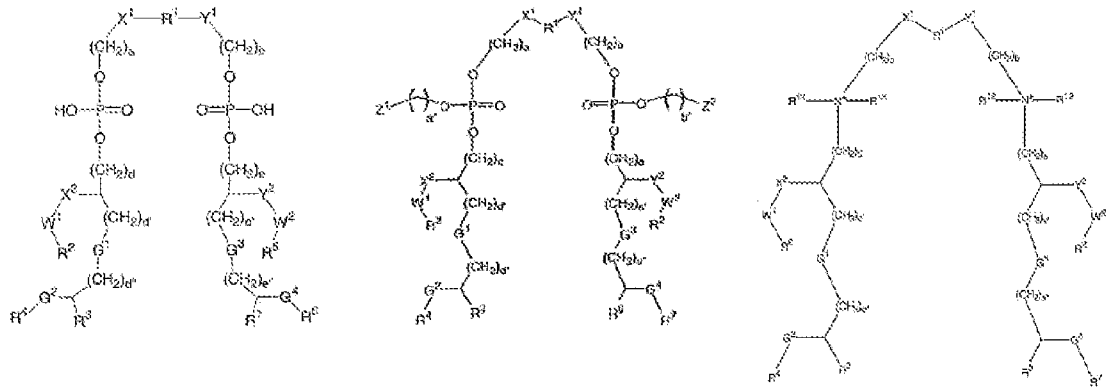
M. Imidazoquinolone Compounds.

Examples of imidazoquinolone compounds suitable for use adjuvants in the invention include Imiquimod ("R-837") (67,68), Resiquimod ("R-848") (69), and their analogs; and salts thereof (*e.g.* the hydrochloride salts). Further details about immunostimulatory imidazoquinolines can be found in references 70 to71.

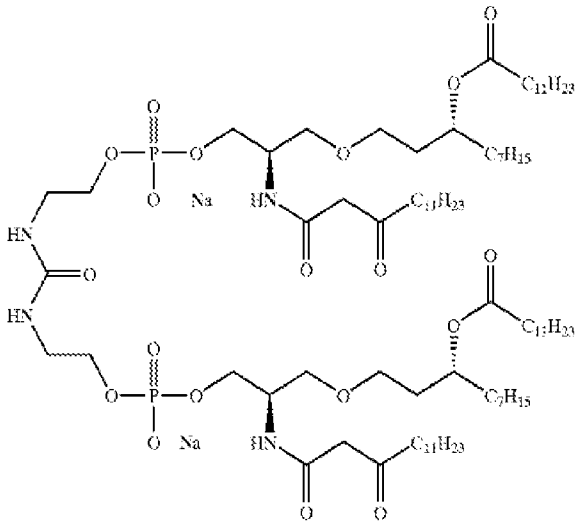
N. Substituted ureas

Substituted ureas useful as adjuvants include compounds of formula I, II or III, or salts thereof:

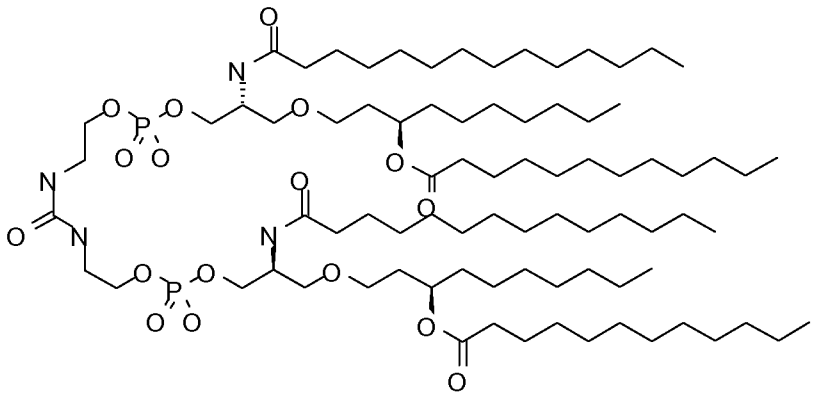
I II III



as defined in reference 72, such as 'ER 803058', 'ER 803732', 'ER 804053', ER 804058', 'ER 804059', 'ER 804442', 'ER 804680', 'ER 804764', ER 803022 or 'ER 804057' *e.g.*:



ER804057

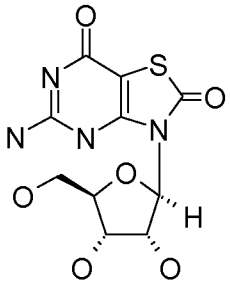


ER-803022:

O. Further adjuvants

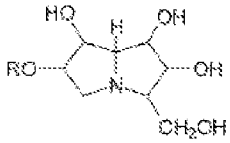
Further adjuvants that may be used with the invention include:

- An aminoalkyl glucosaminide phosphate derivative, such as RC-529 (73,74).
- A thiosemicarbazone compound, such as those disclosed in reference 75. Methods of formulating, manufacturing, and screening for active compounds are also described in reference 75. The thiosemicarbazones are particularly effective in the stimulation of human peripheral blood mononuclear cells for the production of cytokines, such as TNF-α.
- A tryptanthrin compound, such as those disclosed in reference 76. Methods of formulating, manufacturing, and screening for active compounds are also described in reference 76. The thiosemicarbazones are particularly effective in the stimulation of human peripheral blood mononuclear cells for the production of cytokines, such as TNF-α.
- A nucleoside analog, such as: (a) Isatorabine (ANA-245; 7-thia-8-oxoguanosine):



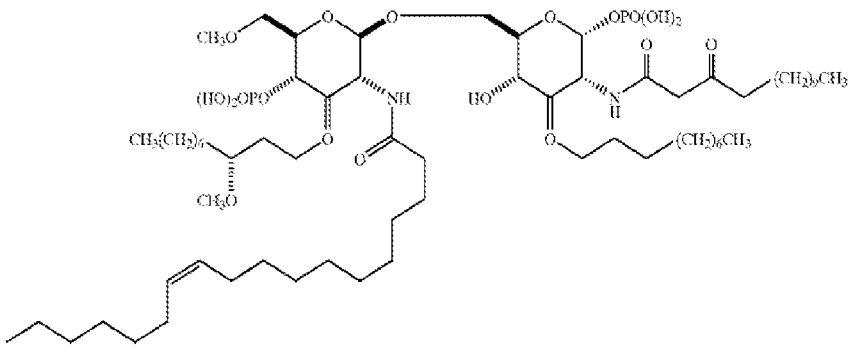
and prodrugs thereof; (b) ANA975; (c) ANA-025-1; (d) ANA380; (e) the compounds disclosed in references 77 to78Loxoribine (7-allyl-8-oxoguanosine) (79).

- Compounds disclosed in reference 80, including: Acylpiperazine compounds, Indoledione compounds, Tetrahydraisoquinoline (THIQ) compounds, Benzocyclodione compounds, Aminoazavinyl compounds, Aminobenzimidazole quinolinone (ABIQ) compounds (81,82), Hydrapthalamide compounds, Benzophenone compounds, Isoxazole compounds, Sterol compounds, Quinazilinone compounds, Pyrrole compounds (83), Anthraquinone compounds, Quinoxaline compounds, Triazine compounds, Pyrazalopyrimidine compounds, and Benzazole compounds (84).
- Compounds containing lipids linked to a phosphate-containing acyclic backbone, such as the TLR4 antagonist E5564 (85,86):
- A polyoxidonium polymer (87,88) or other N-oxidized polyethylene-piperazine derivative.
- Methyl inosine 5'-monophosphate ("MIMP") (89).
- A polyhydroxlated pyrrolizidine compound (90), such as one having formula:



where R is selected from the group comprising hydrogen, straight or branched, unsubstituted or substituted, saturated or unsaturated acyl, alkyl (*e.g.* cycloalkyl), alkenyl, alkynyl and aryl groups, or a pharmaceutically acceptable salt or derivative thereof. Examples include, but are not limited to: casuarine, casuarine-6-α-D-glucopyranose, 3-*epi*-casuarine, 7-*epi*-casuarine, 3,7-di*epi*-casuarine, *etc.*

- A CD1d ligand, such as an α-glycosylceramide (91-92) (*e.g.* α-galactosylceramide), phytosphingosine-containing α-glycosylceramides, OCH, KRN7000 ((2S,3S,4R)-1-O-(α-D-galactopyranosyl)-2-(N-hexacosanoylamino)-1,3,4-octadecanetriol), CRONY-101, 3''-O-sulfo-galactosylceramide, *etc.*
- A gamma inulin (93) or derivative thereof, such as algammulin.



Adjuvant combinations

The invention may also comprise combinations of aspects of one or more of the adjuvants identified above. For example, the following adjuvant compositions may be used in the invention: (1) a saponin and an oil-in-water emulsion (94); (2) a saponin (*e.g.* QS21) + a non-toxic LPS derivative (*e.g.* 3dMPL) (95); (3) a saponin (*e.g.* QS21) + a non-toxic LPS derivative (*e.g.* 3dMPL) + a cholesterol; (4) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) (96); (5) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (97); (6) SAF, containing 10% squalane, 0.4% TWEEN 80™, 5% PLURONIC(TM)-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion. (7) Ribi™ adjuvant system (RAS), (Ribi Immunochem) containing 2% squalene, 0.2% TWEEN 80(TM), and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); and (8) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dMPL).

Other substances that act as immunostimulating agents are disclosed in chapter 7 of ref. 5.

The use of an aluminum hydroxide and/or aluminum phosphate adjuvant is particularly preferred, and antigens are generally adsorbed to these salts. Calcium phosphate is another preferred adjuvant. Other preferred adjuvant combinations include combinations of Th1 and Th2 adjuvants such as CpG & alum or resiquimod & alum. A combination of aluminum phosphate and 3dMPL may be used.

The compositions of the invention may elicit both a cell mediated immune response as well as a humoral immune response. This immune response will preferably induce long lasting (*e.g.* neutralizing) antibodies and a cell mediated immunity that can quickly respond upon exposure to the pathogen immunized against.

Two types of T cells, CD4 and CD8 cells, are generally thought necessary to initiate and/or enhance cell mediated immunity and humoral immunity. CD8 T cells can express a CD8 co-receptor and are commonly referred to as Cytotoxic T lymphocytes (CTLs). CD8 T cells are able to recognized or interact with antigens displayed on MHC Class I molecules.

CD4 T cells can express a CD4 co-receptor and are commonly referred to as T helper cells. CD4 T cells are able to recognize antigenic peptides bound to MHC class II molecules. Upon interaction with a MHC class II molecule, the CD4 cells can secrete factors such as cytokines. These secreted cytokines can activate B cells, cytotoxic T cells, macrophages, and other cells that participate in an immune response. Helper T cells or CD4+ cells can be further divided into two functionally distinct subsets: TH1 phenotype and TH2 phenotypes which differ in their cytokine and effector function.

Activated TH1 cells enhance cellular immunity (including an increase in antigen-specific CTL production) and are therefore of particular value in responding to intracellular infections. Activated TH1 cells may secrete one or more of IL-2, IFN-γ, and TNF-β. A TH1 immune response may result in local inflammatory reactions by activating macrophages, NK (natural killer) cells, and CD8 cytotoxic T cells (CTLs). A TH1 immune response may also act to expand the immune response by stimulating growth of B and T cells with IL-12. TH1 stimulated B cells may secrete IgG2a.

Activated TH2 cells enhance antibody production and are therefore of value in responding to extracellular infections. Activated TH2 cells may secrete one or more of IL-4, IL-5, IL-6, and IL-10. A TH2 immune response may result in the production of IgG1, IgE, IgA and memory B cells for future protection.

An enhanced immune response may include one or more of an enhanced TH1 immune response and a TH2 immune response.

A TH1 immune response may include one or more of an increase in CTLs, an increase in one or more of the cytokines associated with a TH1 immune response (such as IL-2, IFN-γ, and TNF-β), an increase in activated macrophages, an increase in NK activity, or an increase in the production of IgG2a. Preferably, the enhanced TH1 immune response will include an increase in IgG2a production.

A TH1 immune response may be elicited using a TH1 adjuvant. A TH1 adjuvant will generally elicit increased levels of IgG2a production relative to immunization of the antigen without adjuvant. TH1 adjuvants suitable for use in the invention may include for example saponin formulations, virosomes and virus like particles, non-toxic derivatives of enterobacterial lipopolysaccharide (LPS), immunostimulatory oligonucleotides. Immunostimulatory oligonucleotides, such as oligonucleotides containing a CpG motif, are preferred TH1 adjuvants for use in the invention.

A TH2 immune response may include one or more of an increase in one or more of the cytokines associated with a TH2 immune response (such as IL-4, IL-5, IL-6 and IL-10), or an increase in

the production of IgG1, IgE, IgA and memory B cells. Preferably, the enhanced TH2 immune response will include an increase in IgG1 production.

A TH2 immune response may be elicited using a TH2 adjuvant. A TH2 adjuvant will generally elicit increased levels of IgG1 production relative to immunization of the antigen without adjuvant. TH2 adjuvants suitable for use in the invention include, for example, mineral containing compositions, oil-emulsions, and ADP-ribosylating toxins and detoxified derivatives thereof. Mineral containing compositions, such as aluminum salts are preferred TH2 adjuvants for use in the invention.

Preferably, the invention includes a composition comprising a combination of a TH1 adjuvant and a TH2 adjuvant. Preferably, such a composition elicits an enhanced TH1 and an enhanced TH2 response, i.e., an increase in the production of both IgG1 and IgG2a production relative to immunization without an adjuvant. Still more preferably, the composition comprising a combination of a TH1 and a TH2 adjuvant elicits an increased TH1 and/or an increased TH2 immune response relative to immunization with a single adjuvant (*i.e.*, relative to immunization with a TH1 adjuvant alone or immunization with a TH2 adjuvant alone).

The immune response may be one or both of a TH1 immune response and a TH2 response. Preferably, immune response provides for one or both of an enhanced TH1 response and an enhanced TH2 response.

The enhanced immune response may be one or both of a systemic and a mucosal immune response. Preferably, the immune response provides for one or both of an enhanced systemic and an enhanced mucosal immune response. Preferably the mucosal immune response is a TH2 immune response. Preferably, the mucosal immune response includes an increase in the production of IgA.

Pharmaceutical compositions

One aspect of the invention includes pharmaceutical compositions comprising (a) a polysaccharide conjugate as disclosed herein, (b) a pharmaceutically acceptable carrier, and optionally (c) an adjuvant as described in the preceding section.

The compositions of the invention may be prepared in various forms. For example, the compositions may be prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared (*e.g.* a lyophilized composition or a spray-freeze dried composition). The composition may be prepared for topical administration *e.g.* as an ointment, cream or powder. The composition may be prepared for oral administration *e.g.* as a tablet or capsule, as a spray, or as a syrup (optionally flavored). The composition may be prepared for pulmonary administration *e.g.* as an inhaler, using a fine powder or a spray. The composition may be prepared as a suppository

or pessary. The composition may be prepared for nasal, aural or ocular administration *e.g.* as drops. The composition may be in kit form, designed such that a combined composition is reconstituted just prior to administration to a patient. Such kits may comprise one or more antigens in liquid form and one or more lyophilized antigens.

Where a composition is to be prepared extemporaneously prior to use (*e.g.* where a component is presented in lyophilized form) and is presented as a kit, the kit may comprise two vials, or it may comprise one ready-filled syringe and one vial, with the contents of the syringe being used to reactivate the contents of the vial prior to injection.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of antigen(s) (*e.g.*, the polysaccharide and/or the carrier protein), as well as any other components, as needed. By 'immunologically effective amount', it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, age, the taxonomic group of individual to be treated (*e.g.* non-human primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

Methods of treatment, and administration of the vaccine

The invention also provides a method for raising an immune response in a mammal comprising the step of administering an effective amount of a composition of the invention. The immune response is preferably protective and preferably involves antibodies and/or cell-mediated immunity. The method may raise a booster response.

The invention also provides a polypeptide of the invention for use as a medicament *e.g.* for use in raising an immune response in a mammal.

The invention also provides the use of a polypeptide of the invention in the manufacture of a medicament for raising an immune response in a mammal.

The invention also provides a delivery device pre-filled with an immunogenic composition of the invention.

Because glucans (and β-glucans in particular) are an essential and principal polysaccharide constituent of almost all pathogenic fungi, particularly those involved in infections in immunocompromised subjects, and also in bacterial pathogens and protozoa, anti-glucan immunity may have efficacy against a broad range of pathogens and diseases. For example, anti-glucan serum raised after immunization with *S. cerevisiae* is cross-reactive with *C. albicans*.

Broad spectrum immunity is particularly useful because, for these human infectious fungal agents, chemotherapy is scanty, antifungal drug resistance is emerging and the need for preventative and therapeutic vaccines is increasingly recognized.

Therefore, where the polysaccharide immunogen of the polysaccharide conjugates disclosed herein are glucans, the uses and methods of the glucan polysaccharide conjugates disclosed herein are particularly useful for treating/protecting against infections of: Candida species, such as *C. albicans*; Cryptococcus species, such as *C. neoformans*; Enterococcus species, such as *E. faecalis*; Streptococcus species, such as *S. pneumoniae, S. mutans, S. agalactiae* and *S. pyogenes*; Leishmania species, such as *L. major*; Acanthamoeba species, such as *A. castellani*; Aspergillus species, such as *A. fumigatus* and *A. flavus*; Pneumocystis species, such as *P. carinii*; Mycobacterium species, such as *M. tuberculosis*; Pseudomonas species, such as *P. aeruginosa*; Staphylococcus species, such as *S. aureus*; Salmonella species, such as *S. typhimurium*; Coccidioides species such as *C. immitis*; Trichophyton species such as *T. verrucosum*; Blastomyces species such as *B. dermatidis*; Histoplasma species such as *H. capsulatum*; Paracoccidioides species such as *P. brasiliensis*; Pythium species such as *P. insidiosum*; and Escherichia species, such as *E. coli*.

The uses and methods are particularly useful for preventing/treating diseases including, but not limited to: candidiasis (including hepatosplenic candidiasis, invasive candidiasis, chronic mucocutaneous candidiasis and disseminated candidiasis); candidemia; aspergillosis, cryptococcosis, dermatomycoses, sporothrychosis and other subcutaneous mycoses, blastomycosis, histoplasmosis, coccidiomycosis, paracoccidiomycosis, pneumocystosis, thrush, tuberculosis, mycobacteriosis, respiratory infections, scarlet fever, pneumonia, impetigo, rheumatic fever, sepsis, septicaemia, cutaneous and visceral leishmaniasis, corneal acanthamoebiasis, cystic fibrosis, typhoid fever, gastroenteritis and hemolytic- uremic syndrome. Anti-*C. albicans* activity is particularly useful for treating infections in AIDS patients.

Efficacy of immunization can be tested by monitoring immune responses against β-glucan (*e.g.*, anti-β-glucan antibodies) after administration of the composition. Efficacy of therapeutic treatment can be tested by monitoring microbial infection after administration of the composition of the invention.

By raising an immune response in the mammal by these uses and methods, the mammal can be protected against infection by the pathogen from which the polysaccharide immunogen is derived (or mimics) as well as from *E. coli* infection owing to the *E. coli* carrier protein, including ExPEC and non-ExPEC strains. The invention is particularly useful for providing broad protection against pathogenic *E. coli*, including intestinal pathotypes such as EPEC, EAEC, EIEC, ETEC and DAEC pathotypes. Thus the mammal may be protected against diseases including, but not limited to peritonitis, pyelonephritis, cystitis, endocarditis, prostatitis,

urinary tract infections (UTIs), meningitis (particularly neonatal meningitis), sepsis (or SIRS), dehydration, pneumonia, diarrhea (infantile, travellers', acute, persistent, *etc.*), bacillary dysentery, hemolytic uremic syndrome (HUS), pericarditis, bacteriuria, *etc.*

The mammal is preferably a human, but may be, *e.g.*, a cow, a pig, a cat or a dog, as *E. coli* disease is also problematic in these species. While the specification refers to mammals and mammalian subjects, the polysaccharide conjugates disclosed herein are also useful for avian species such as chicken and duck and therefore wherever mammal or mammalian is recited herein, avian can also be included. Where the vaccine is for prophylactic use, the human is preferably a child (*e.g.* a toddler or infant) or a teenager; where the vaccine is for therapeutic use, the human is preferably a teenager or an adult. A vaccine intended for children may also be administered to adults *e.g.* to assess safety, dosage, immunogenicity, *etc.*

One way of checking efficacy of therapeutic treatment involves monitoring *E. coli* infection after administration of the compositions of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses, systemically (such as monitoring the level of IgG1 and IgG2a production) and/or mucosally (such as monitoring the level of IgA production), against the antigens in the compositions of the invention after administration of the composition. Typically, antigen-specific serum antibody responses are determined post-immunization but pre-challenge whereas antigen-specific mucosal antibody responses are determined post-immunization and post-challenge.

Another way of assessing the immunogenicity of the compositions of the present invention is to express the proteins recombinantly for screening patient sera or mucosal secretions by immunoblot and/or microarrays. A positive reaction between the protein and the patient sample indicates that the patient has mounted an immune response to the protein in question. This method may also be used to identify immunodominant antigens and/or epitopes within antigens.

The efficacy of vaccine compositions can also be determined *in vivo* by challenging animal models of *E. coli* infection, *e.g.*, guinea pigs or mice, with the vaccine compositions. A murine model of ExPEC and lethal sepsis is described in reference 98. A cotton rat model is disclosed in ref. 99.

Compositions of the invention will generally be administered directly to a patient. Direct delivery may be accomplished by parenteral injection (*e.g.* subcutaneously, intraperitoneally, intravenously, intramuscularly, or to the interstitial space of a tissue), or mucosally, such as by rectal, oral (*e.g.* tablet, spray), vaginal, topical, transdermal or transcutaneous, intranasal, ocular, aural, pulmonary or other mucosal administration. Novel direct delivery forms can also include transgenic expression of the polypeptides disclosed herein in foods, e.g., transgenic expression in a potato.

The invention may be used to elicit systemic and/or mucosal immunity, preferably to elicit an enhanced systemic and/or mucosal immunity.

Preferably the enhanced systemic and/or mucosal immunity is reflected in an enhanced TH1 and/or TH2 immune response. Preferably, the enhanced immune response includes an increase in the production of IgG1 and/or IgG2a and/or IgA.

Dosage can be by a single dose schedule or a multiple dose schedule. Multiple doses may be used in a primary immunization schedule and/or in a booster immunization schedule. In a multiple dose schedule the various doses may be given by the same or different routes *e.g.* a parenteral prime and mucosal boost, a mucosal prime and parenteral boost, *etc.* Multiple doses will typically be administered at least 1 week apart (*e.g.* about 2 weeks, about 3 weeks, about 4 weeks, about 6 weeks, about 8 weeks, about 10 weeks, about 12 weeks, about 16 weeks, *etc.*).

Vaccines of the invention may be used to treat both children and adults. Thus a human patient may be less than 1 year old, 1-5 years old, 5-15 years old, 15-55 years old, or at least 55 years old. Preferred patients for receiving the vaccines are the elderly (*e.g.* ≥50 years old, ≥60 years old, and preferably ≥65 years), the young (*e.g.* ≤5 years old), hospitalized patients, healthcare workers, armed service and military personnel, pregnant women, the chronically ill, or immunodeficient patients. The vaccines are not suitable solely for these groups, however, and may be used more generally in a population.

Vaccines of the invention are particularly useful for patients who are expecting a surgical operation, or other hospital in-patients. They are also useful in patients who will be catheterized. They are also useful in adolescent females (*e.g.* aged 11-18) and in patients with chronic urinary tract infections.

Vaccines of the invention may be administered to patients at substantially the same time as (*e.g.* during the same medical consultation or visit to a healthcare professional or vaccination centre) other vaccines *e.g.* at substantially the same time as a measles vaccine, a mumps vaccine, a rubella vaccine, a MMR vaccine, a varicella vaccine, a MMRV vaccine, a diphtheria vaccine, a tetanus vaccine, a pertussis vaccine, a DTP vaccine, a conjugated *H. influenzae* type b vaccine, an inactivated poliovirus vaccine, a hepatitis B virus vaccine, a meningococcal conjugate vaccine (such as a tetravalent A-C-W135-Y vaccine), a respiratory syncytial virus vaccine, *etc.*

General

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, *e.g.*, references 100-101, *etc.*

The term "comprising" encompasses "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "about" in relation to a numerical value *x* means, for example, $x \pm 10\%$.

"GI" numbering is used herein. A GI number, or "GenInfo Identifier", is a series of digits assigned consecutively to each sequence record processed by NCBI when sequences are added to its databases. The GI number bears no resemblance to the accession number of the sequence record. When a sequence is updated (*e.g.* for correction, or to add more annotation or information) then it receives a new GI number. Thus the sequence associated with a given GI number is never changed.

References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of ref. 102. A preferred alignment is determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is disclosed in ref. 103.

One of skill in the art would understand that "isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated" when in such living organism, but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated," as the term is used in this disclosure. Further, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method would be understood to be "isolated" even if it is still present in said organism, which organism may be living or non-living, except where such transformation, genetic manipulation or other recombinant method produces an organism that is otherwise indistinguishable from the naturally occurring organism.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 (A) shows an SDS-Page (4-12% gradient) with the following (from left to right): (1) orf405B before conjugation, (2) orf405B after conjugation, (3) upec-5211 before conjugation, and (4) upec-5211 after conjugation; and (B) shows an SDS-Page (4-12% gradient) with the following (from left to right): molecular weight markers, (1) upec-5211 before conjugation, (2) upec-5211 after conjugation, (3) orf3526 before conjugation, (4) orf3526 after conjugation, (5) orf1364 before conjugation, and (6) orf1364 after conjugation.

Figure 2 shows the results of the first carrier protein immunization study. The y-axis shows the anti-laminarin IgG induction after immunization as measured by ELSA. The x-axis shows the ELISA results for the following conjugates (from left to right): laminarin conjugated to CRM_{197}, laminarin conjugated to orf405B, and laminarin conjugated to upec-5211.

Figure 3 shows the results of the second carrier protein immunization study. The y-axis shows the anti-laminarin IgG induction after immunization as measured by ELSA. The x-axis shows the ELISA results for the following conjugates (from left to right): laminarin conjugated to CRM_{197}, laminarin conjugated to orf3526, and laminarin conjugated to orf1364.

BRIEF DESCRIPTION OF SEQUENCE LISTING

SEQ ID	Description
	SEQ ID NOs: 1-22 = orf1364 amino acid sequences
1	Orf1364 sequence from EPEC strain E110019
2	Orf1364 sequence from EHEC strains Sakai, EDL933, EC508, EC869, EC4024, EC4042, EC4045, EC4076, EC4113, EC4115, EC4196, EC4206, EC4401, EC4486, EC4501 and TW14588
3	Orf1364 sequence from EPEC strain B171
4	Orf1364 sequence from EPEC strain E22
5	Orf1364 sequence from EPEC strain B171
6	Orf1364 sequence from EPEC strain B171
7	Orf1364 sequence from ETEC strain E24377A and EAEC strain O42
8	Orf1364 sequence from ETEC strain E24377A
9	Orf1364 sequence from UPEC strainUTI89 and NMEC strains RS218 and IHE3034
10	Orf1364 sequence from EPEC strain E110019
11	Orf1364 sequence from EPEC strain E22
12	Orf1364 sequence from ETEC strain H10407
13	Orf1364 sequence from UPEC strains F11 and 536
14	Orf1364 sequence from antibiotic resistant strain SECEC
15	Orf1364 sequence from ETEC strain H10407
16	Orf1364 sequence from commensal strains W3110 and DH10B
17	Orf1364 sequence from commensal strain MG1655
18	Orf1364 sequence from EAEC strain O42
19	Orf1364 sequence from ETEC strain B7A
20	Orf1364 sequence from UPEC strain CFT073
21	Orf1364 sequence from EAEC strain O42

22	Orf1364 sequence from UPEC strain CFT073
	SEQ ID NOs: 23-25 = upec-5211 amino acid sequences
23	Upec-5211 sequence from UPEC strain CFT073 and from asymptomatic bacteriuria (ABU) strain 83972
24	Upec-5211 sequence from commensal strain ED1a
25	Upec-5211 sequence from *E. fergusonii* ATCC 35469
	SEQ ID NOs: 26-40 = orf3526 amino acid sequences
26	Orf3526 sequence from EAEC strain 101-1 (GI: 83587587)
27	Orf3526 sequence from UPEC strain 536 (GI: 110643204 - ECP_3050)
28	Orf3526 sequence from EAEC strain O42
29	Orf3526 sequence from EPEC strain E2348/69
30	Orf3526 sequence from EIEC strain 53638 (GI: 75515237)
31	Orf3526 sequence from commensal strain W3110 (GI: 89109748 - yghJ)
32	Orf3526 sequence from ETEC strain B7A (GI: 75227618)
33	Orf3526 sequence from EPEC strain E22 (GI: 75259912)
34	Orf3526 sequence from ETEC strain E24377A (GI: 157156747 - EcE24377A_3432)
35	Orf3526 sequence from ETEC strain H10407
36	Orf3526 sequence from EPEC strain E110019 (GI: 75239450)
37	Orf3526 sequence from commensal strain HS (GI: 157162442 - EcHS_A3142)
38	Orf3526 sequence from antibiotic-resistant strain SECEC
39	Orf3526 sequence from NMEC strain IHE3034
40	Orf3526 sequence from UPEC strain F11 (GI: 75241179)
41	pK1-0405B – Coding sequence for orf405B used in the Examples
42	pK1-0405B – Amino acid sequence of orf405B used in the Examples
43	pK1-1364 – Coding sequence for orf1364 used in the Examples
44	pK1-1364 – Amino acid sequence of orf1364 used in the Examples
45	Coding sequence for orf3613 used in the Examples
46	Amino acid sequence of orf3613 used in the Examples
47	Coding sequence for upec-5211 used in the Examples
48	Amino acid sequence of upec-5211 used in the Examples
49	Coding sequence for orf3526 used in the Examples
50	Amino acid sequence of orf3526 used in the Examples
51	dIdC polyoligonucleotide

52	Polycationic polymer peptide

EXAMPLES

Conjugation Process

Five different recombinant proteins from different strains of Extraintestinal Pathogenic *E. coli* were compared for their ability to act as carrier proteins for polysaccharide immunogen (using laminarin as an exemplary polysaccharide immunogen): orf3613 (SEQ ID NO: 46), upec-5211 (SEQ ID NO: 48), orf3526 (SEQ ID NO: 50), orf1364 (SEQ ID NO: 44) and orf405B (SEQ ID NO: 42). The conjugate of orf3613 proved to be insoluble, so it was not tested further. After conjugation with laminarin, each of the four remaining conjugates were compared with a laminarin-CRM_{197} conjugate given that CRM_{197} is well characterized in its use as a carrier protein for polysaccharide immunogens.

These conjugates were prepared starting from a suitable solution of each protein and activated laminarin, prepared by the method disclosed in Torosantucci *et al.* (2005) *J. Exp. Med.* 202(5):597-606. Briefly, prior to conjugation, the saccharide was activated with an N-hydroxysuccinimide diester of adipic acid. Then the conjugation reaction for upec-5211 and orf405B was carried out in 50 mM NaH2PO4, 150 mM NaCl, 6 M guanidine HCl at pH 7.6 using a molar ratio of polysaccharide ester groups to protein of 30 and a protein concentration of 1.2-1.4 mg/ml. The results of these conjugation reactions were run on a gradient SDS-Page (4-12%) as shown in Figure 1(A).

In a separate set of conjugation reactions, the conjugation reaction for upec-5211 (again), orf3526 and orf1364 was carried out in PBS using a molar ratio of polysaccharide ester groups to protein of 30 and a protein concentration of 5 mg/ml.

The reaction mixtures in each of the five reactions were mixed at room temperature and left overnight.

The resultant conjugates were purified by immobilized metal ion affinity chromatography (IMAC), a well known separation method for protein purification (see, for example, refs. 1 and 2). The purification was performed with HIS MULTITRAP HP (TM) plates (GE Healthcare), prepacked 96-well filter plates for small-scale purification of histidine-tagged proteins, with the use of a vacuum source.

The plates were pre-packed with IMAC matrix NI SEPHAROSE HIGH PERFORMANCE (TM) (GE Healthcare), which consists of 34-µm beads of highly cross-linked agarose to which iminodiacetic acid groups are covalently conjugated. The iminodiacetic acid groups were loaded with Ni^{2+} in order to provide a group with affinity for the carrier protein with exposed histidine

groups in the conjugate. The conjugates, prepared in solution as described above, were bound to the IMAC matrix. The IMAC matrix was then washed to remove the free saccharide. After washing, the purified conjugates were finally eluted.

Purification:

1. The matrix of the HIS MULTITRAP HP (TM) plate was washed with 0.5 ml of distilled water, and then equilibrated two times with 0.5 ml binding buffer, consisting of 6M Guanidine HCl, 20 mM $NaPO_4$, pH 7.6).

2. The conjugates (about 200 µl/well for 250 µg/well of protein) were loaded onto the plate and the plate was left on rocking platform for 1 hour at room temperature to allow the his-tagged protein to bind to the IMAC matrix. The plate was then washed two times with 0.5 ml binding buffer to remove free saccharide from the bound conjugates.

3. The conjugates were recovered by the addition to the matrix of with 0.2 ml elution buffer 6M Guanidine-HCl, 20 mM $NaPO_4$, 500 mM NaCl, 500 mM imidazole, pH 7.6. The plate was incubated for 10 minutes and the eluate containing the conjugates was collected. This elution procedure was repeated once. The two elution fractions were combined and dialyzed against 10 mM $NaPO_4$ pH 7.2 or 10 mM $NaPO_4$, 100 mM NaCl pH 7.2 or PBS with spectra Por membrane, 6-8 kDa cut-off.

The purified conjugates of the invention were characterized by SDS-Page 4-12% (See Figures 1(A) and 1(B)), by MicroBCA to quantify the protein amount, and by HPAEC-PAD to quantify the saccharide amount. Table 1 summarizes the conjugation efficiencies as determined by MicroBCA and HPAEC-PAD.

TABLE 1

Protein	CHO, mg/ml (HPAEC-PAD)	Protein, mg/ml (MicroBCA)	Sacch./Prot. w/w %	CHO mol / Protein mol
orf405B	0.04223	0.139	30.38	2.3
upec-5211	0.04844	0.119	40.71	3.7
upec-5211 (second preparation with PBS)	0.2258	0.285	79.30	5.9
orf3526	0.09179	0.471	19.53	4.9
orf1364	0.162	0.458	35.37	2.9

Carrier protein Immunization study – E. coli protein tested as carrier in Laminarin conjugates

Balb/c mice at 4-6 weeks old were immunized at days 0, 14 and 28 by subcutaneous injection with a 5 μg dose of polysaccharide immunogen in an injection volume of 150 μl. The mice were bleed on days 0, 27 and 42.

Immunizations were carried out in groups of six mice for the first immunization study and eight mice for the second immunization study.

In the first immunization study, the mice were injected with laminarin conjugated to: upec-5211 (prepared via the first protocol above), orf405B, and (as a strong positive control) CRM_{197}.

The results show that upec-5211 as a carrier protein induced a decent antibody response against the polysaccharide portion (approximately two-thirds of the response induced by CRM_{197}), while orf405B as a carrier protein induced no antibody response to the polysaccharide portion (See Figure 2).

In the second immunization study, the mice were injected with laminarin conjugated to: orf3526, orf1364, and (as a strong positive control) CRM_{197}.

The results show that orf3526 as a carrier protein induced a strong antibody response against the polysaccharide portion (comparable or lightly great than that of CRM_{197}) while orf1364 give a lower, but still decent response (approximately one third of the response induced by CRM_{197}) (See Figure 3).

It is especially noteworthy that the ability to induce an antibody response to a polysaccharide conjugate is not related to the ability of the carrier protein to induce a protective immune response to itself. Table 2 summarizes previous experiments using similar fragments (or in the case of upec-5211, orf1364, and orf405B, the exact same sequence) of the same protein to inoculate mice showed a very different response in the sepsis model. Each of upec-5211, orf1364 and orf405B provided relatively similar immune responses while each provides quite different degrees of induction of an immune response to a conjugated polysaccharide.

TABLE 2

Candidate	**Sepsis Animal Model**		
	Survival with vaccination (%)	Survival without vaccination (%)	P value
orf3526	**(75)**	**(0)**	
upec-5211	30/93 **(32)**	14/91 **(15)**	0.009
orf1364 (flu antigen 43 fragment)	21/77 **(27)**	8/84 **(9)**	0.004
orf405B (bacterial Ig-like domain	17/63 **(26)**	9/66 **(13)**	0.07

(group 1) protein fragment)			

Immunogenicity – orf3526A, orf3526B, and orf3526C

To confirm that the three preferred fragments described in the alignment above had substantially the same immunogenicity as the full length AcfD (orf3526), the fragments based upon the full length sequence of SEQ ID NO: 39 were purified. The purified fragments were used in immunization experiments in mice, adjuvanted with Freund's complete adjuvant (orf3526A and orf3526B) or with alum (orf3526C). Immunized mice were then challenged with a lethal dose of *E. coli*. Results of the challenge are shown in Table 3 below.

TABLE 3

Candidates	**Sepsis animal model**		
	Survival with vaccination (%)	**Survival without vaccination (%)**	**P value**
orf3526A/FCA	7/8 **(87.5)**	1/8 **(12.5)**	0.01
orf3526B/FCA	6/8 **(75)**	1/8 **(12.5)**	0.04
orf3526C 2 µg/alum	13/16 **(81)**	0/7 **(0)**	0.0005
orf3526C 20 µg/alum	15/16 **(93)**	0/7 **(0)**	0.0001

Thus, the fragments generate substantially similar immune responses in mice and will therefore be expected to provide substantially similar enhancement of the immune response to a conjugated polysaccharide as the full length orf3526.

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

REFERENCES (the contents of which are hereby incorporated in full)

[1] WO2006/089264.
[2] WO2006/091517.
[3] Needleman & Wunsch (1970) *J. Mol. Biol.* 48, 443-453.
[4] Rice *et al.* (2000) *Trends Genet* 16:276-277.
[5] *Vaccine Design: The Subunit and Adjuvant Approach* (eds. Powell & Newman) Plenum Press 1995 (ISBN 0-306-44867-X).
[6] *Vaccine Adjuvants: Preparation Methods and Research Protocols* (Volume 42 of *Methods in Molecular Medicine* series). ISBN: 1-59259-083-7. Ed. O'Hagan.
[7] US patent 6355271.
[8] WO00/23105.
[9] WO90/14837.
[10] WO90/14837.

[11] Podda (2001) *Vaccine* 19: 2673-2680.

[12] *Vaccine Design: The Subunit and Adjuvant Approach* (eds. Powell & Newman) Plenum Press 1995 (ISBN 0-306-44867-X).

[13] *Vaccine Adjuvants: Preparation Methods and Research Protocols* (Volume 42 of *Methods in Molecular Medicine* series). ISBN: 1-59259-083-7. Ed. O'Hagan.

[14] Allison & Byars (1992) *Res Immunol* 143:519-25.

[15] Hariharan *et al.* (1995) *Cancer Res* 55:3486-9.

[16] US-2007/014805.

[17] Suli *et al.* (2004) *Vaccine* 22(25-26):3464-9.

[18] WO95/11700.

[19] US patent 6,080,725.

[20] WO2005/097181.

[21] WO2006/113373.

[22] Han *et al.* (2005) *Impact of Vitamin E on Immune Function and Infectious Diseases in the Aged* at *Nutrition, Immune functions and Health* EuroConference, Paris, 9-10 June 2005.

[23] US- 6630161.

[24] US 5,057,540.

[25] WO96/33739.

[26] WO96/11711.

[27] WO00/07621.

[28] Barr *et al.* (1998) *Advanced Drug Delivery Reviews* 32:247-271.

[29] Sjolanderet *et al.* (1998) *Advanced Drug Delivery Reviews* 32:321-338.

[30] Niikura *et al.* (2002) *Virology* 293:273-280.

[31] WO03/024481.

[32] Gluck *et al.* (2002) *Vaccine* 20:B10-B16.

[33] EP-A-0689454.

[34] Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.

[35] Evans *et al.* (2003) *Expert Rev Vaccines* 2:219-229.

[36] Meraldi *et al.* (2003) *Vaccine* 21:2485-2491.

[37] Pajak *et al.* (2003) *Vaccine* 21:836-842.

[38] Kandimalla *et al.* (2003) *Nucleic Acids Research* 31:2393-2400.

[39] WO02/26757.

[40] WO99/62923.

[41] Krieg (2003) *Nature Medicine* 9:831-835.

[42] US 6,429,199.

[43] Kandimalla *et al.* (2003) *Biochemical Society Transactions* 31 (part 3):654-658.

[44] Blackwell *et al.* (2003) *J Immunol* 170:4061-4068.

[45] WO01/95935.

[46] Kandimalla *et al.* (2003) *BBRC* 306:948-953.

[47] WO03/035836.

[48] WO01/22972.

[49] Schellack *et al.* (2006) *Vaccine* 24:5461-72.

[50] WO95/17211.

[51] WO98/42375.

[52] Beignon *et al.* (2002) *Infect Immun* 70:3012-3019.

[53] Pine *et al.* (2002) *J Control Release* 85:263-270.

[54] Tebbey *et al.* (2000) *Vaccine* 18:2723-34.

[55] Domenighini *et al.* (1995) *Mol Microbiol* 15:1165-1167.
[56] WO99/40936.
[57] WO99/44636.
[58] Singh *et al]* (2001) *J Cont Release* 70:267-276.
[59] WO99/27960.
[60] US 6,090,406.
[61] EP-A-0626169.
[62] WO99/52549.
[63] WO01/21207.
[64] WO01/21152.
[65] Andrianov *et al.* (1998) *Biomaterials* 19:109-115.
[66] Payne *et al.* (1998) *Adv Drug Delivery Review* 31:185-196.
[67] US 4,680,338.
[68] US 4,988,815.
[69] WO92/15582.
[70] Stanley (2002) *Clin Exp Dermatol* 27:571-577.
[71] Jones (2003) *Curr Opin Investig Drugs* 4:214-218.
[72] WO03/011223.
[73] Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.
[74] Evans *et al.* (2003) *Expert Rev Vaccines* 2:219-229.
[75] WO2004/060308.
[76] WO2004/064759.
[77] US 6,924,271.
[78] US 5,658,731.
[79] US patent 5,011,828.
[80] WO2004/87153.
[81] US 6,605,617.
[82] WO02/18383.
[83] WO2004/018455.
[84] WO03/082272.
[85] Wong *et al.* (2003) *J Clin Pharmacol* 43(7):735-42.
[86] US2005/0215517.
[87] Dyakonova et al. (2004) Int Immunopharmacol 4(13):1615-23.
[88] FR-2859633.
[89] Signorelli & Hadden (2003) *Int Immunopharmacol* 3(8):1177-86.
[90] WO2004/064715.
[91] De Libero *et al*, *Nature Reviews Immunology,* 2005, 5: 485-496
[92] WO03/105769
[93] Cooper (1995) *Pharm Biotechnol* 6:559-80.
[94] WO99/11241.
[95] WO94/00153.
[96] WO98/57659.
[97] European patent applications 0835318, 0735898 and 0761231.
[98] Durant *et al.* (2007) *Infect Immun* 75:1916-25.
[99] WO02/081653.
[100] Gennaro (2000) *Remington: The Science and Practice of Pharmacy.* 20th edition, ISBN: 0683306472.

[101] *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag)

[102] *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30

[103] Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482-489.

CLAIMS

We claim:

1. A glucan polysaccharide conjugate comprising a glucan polysaccharide conjugated to a carrier polypeptide selected from the group consisting of an *E. coli* AcfD precursor protein (orf3526 polypeptide), an *E. coli* Flu antigen 43 protein (orf1364 polypeptide), and an *Escherichia* Sel1 repeat-containing protein (upec-5211 polypeptide).

2. The glucan polysaccharide conjugate of claim 1 wherein the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide) comprising:

(a) the amino acid sequence of SEQ ID NO 50;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NO: 50;

(c) at least 85% sequence identity to SEQ ID NO: 50;

(d) a fragment of at least 10 consecutive amino acids from SEQ ID NO: 50;

and/or

(e) when aligned with SEQ ID NO: 50 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

3. The glucan polysaccharide conjugate of claim 1 wherein the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide) comprising:

(a) the amino acid sequence of SEQ ID NO 44;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NO: 44;

(c) at least 85% sequence identity to any one of SEQ ID NO: 44;

(d) a fragment of at least 10 consecutive amino acids from SEQ ID NO: 44;

and/or

(e) when aligned with SEQ ID NO: 44 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

4. The glucan polysaccharide conjugate of claim 1 wherein the carrier polypeptide is the *Escherichia* Sel1 repeat-containing protein (upec-5211 polypeptide) comprising:

(a) the amino acid sequence of SEQ ID NO 48;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NO: 48;

(c) at least 85% sequence identity to SEQ ID NO: 48;

(d) a fragment of at least 10 consecutive amino acids from SEQ ID NO: 48;

and/or

(e) when aligned with SEQ ID NO: 48 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

5. The glucan polysaccharide conjugate of any one of claims 1 to 4 wherein the glucan polysaccharide contains β-1,3-linkages and/or β-1,6-linkages.

6. The glucan polysaccharide conjugate of any one of claims 1 to 5, wherein the glucan polysaccharide is a single molecular species.

7. The glucan polysaccharide conjugate of any one of claims 1 to 6, wherein the glucan polysaccharide is conjugated to the carrier protein directly.

8. The glucan polysaccharide conjugate of any one of claims 1 to 6, wherein the glucan polysaccharide is conjugated to the carrier protein via a linker.

9. The glucan polysaccharide conjugate of any one of claims 1 to 8, wherein the glucan polysaccharide has a molecular weight of less than 100 kDa (e.g. less than 80, 70, 60, 50, 40, 30, 25, 20, or 15 kDa).

10. The glucan polysaccharide conjugate of any one of claims 1 to 8, wherein the glucan polysaccharide has 60 or fewer glucose monosaccharide units.

11. The glucan polysaccharide conjugate of any one of claims 1 to 10, wherein the glucan polysaccharide is a β-1,3 glucan polysaccharide with some β-1,6 branching.

12. The glucan polysaccharide conjugate of claim 11, wherein the glucan polysaccharide is a laminarin.

13. The glucan polysaccharide conjugate of any one of claims 1 to 11, wherein the glucan polysaccharide comprises both β-1,3-linked glucose residues and β-1,6-linked glucose residues, with a ratio of β-1,3 linked glucose residues to β-1,6-linked residues of at least 8:1 and/or there are one or more sequences of at least five adjacent non-terminal residues linked to other residues only by β-1,3 linkages.

14. The glucan polysaccharide conjugate of any one of claims 1 to 11, wherein the glucan polysaccharide comprises both β-1,3-linked glucose residues and β-1,6-linked glucose residues, with a ratio of β-1,3 linked glucose residues to β-1,6-linked residues of at least 8:1.

15. The glucan polysaccharide conjugate of any one of claims 1 to 10, wherein the glucan polysaccharide has exclusively β-1,3 linkages.

16. The glucan polysaccharide conjugate of any one of claims 13 to 15, wherein the glucan polysaccharide is a curdlan.

17. The glucan polysaccharide conjugate of any one of claims 1-16 further comprising an adjuvant.

18. A vaccine component comprising the glucan polysaccharide conjugate of claims 1-16.

19. A vaccine comprising the vaccine component of claim 18.

20. The vaccine of claim 19 further comprising an adjuvant.

21. The vaccine of claim 19 or claim 20 further comprising an additional vaccine component selected from: a *Neisseria meningitidis* antigen, a *Streptococcus pneumoniae* antigen, a *Streptococcus pyogenes* antigen, a *Moraxella catarrhalis* antigen, a *Bordetella pertussis* antigen, a *Staphylococcus aureus* antigen, a *Staphylococcus epidermis* antigen, a *Clostridium tetani* antigen, a *Cornynebacterium diphtheriae* antigen, a *Haemophilus influenzae* type B (Hib) antigen, a *Pseudomonas aeruginosa* antigen, a *Legionella pneumophila* antigen, a *Streptococcus agalactiae* antigen, a *Neiserria gonorrhoeae* antigen, a *Chlamydia trachomatis* antigen, a *Treponema pallidum* antigen, a *Haemophilus ducreyi* antigen, an *Enterococcus faecalis* antigen, an *Enterococcus faecium* antigen, a *Helicobacter pylori* antigen, a *Staphylococcus saprophyticus* antigen, a *Yersinia enterocolitica* antigen, an additional *E. coli* antigen, a *Bacillus anthracis* antigen, a *Yersinia pestis* antigen, a *Mycobacterium tuberculosis* antigen, a *Rickettsia* antigen, a *Listeria monocytogenes* antigen, a *Chlamydia pneumoniae* antigen, a *Vibrio cholerae* antigen, a *Salmonella typhi* antigen, a *Borrelia burgdorferi* antigen, a *Porphyromonas gingivalis* antigen, a Shigella antigen and a *Klebsiella* antigen.

22. The method of inducing an enhanced immune response in a mammalian subject to polysaccharide comprising:

administering to the mammalian subject of the glucan polysaccharide conjugate of claims 1-16, the vaccine component of claim 18, or the vaccine of claims 19-21.

23. The use of the glucan polysaccharide conjugate of claims 1-16, the vaccine component of claim 18, or the vaccine of claims 19-21 to induce an enhanced immune response in a mammalian subject to the polysaccharide.

24. A polysaccharide conjugate comprising a polysaccharide conjugated to a carrier polypeptide selected from the group consisting of an *E. coli* AcfD precursor protein (orf3526 polypeptide), an *E. coli* Flu antigen 43 protein (orf1364 polypeptide), and an *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide).

25. The polysaccharide conjugate of claim 24 wherein the carrier polypeptide is the *E. coli* AcfD precursor protein (orf3526 polypeptide).

26. The polysaccharide conjugate of claim 25, wherein carrier polypeptide comprises a mutation reducing the toxicity and/or a deletion improving purification as compared to the *E. coli* AcfD precursor protein (orf3526 polypeptide) of SEQ ID NO: 39.

27. The polysaccharide conjugate of claim 26, wherein the mutation is selected from a deletion of all or a portion of the zincin metalloprotease domain and a point mutation in zincin metalloprotease domain which reduces the protease activity.

28. The polysaccharide conjugate of any one of claims 26-27, wherein the point mutation is a mutation of a zinc binding residue or a mutation of a catalytic residue.

29. The polysaccharide conjugate of any one of claims 26-28, wherein the zinc binding residue is amino acid number 1305 based upon alignment with SEQ ID NO: 39.

30. The polysaccharide conjugate of any one of claims 26-29, wherein the carrier polypeptide does not comprise at least the last 100 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 300 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 400 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 500 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 600 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 700 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the last 750 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the last 758 C-terminal amino acids of the *E. coli* AcfD (orf3526) protein or does not comprise at least the first 100 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 200 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 300 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 400 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 500 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 600 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 700 N-

terminal amino acids of the *E. coli* AcfD (orf3526) protein, at least the first 750 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein, or at least the first 760 N-terminal amino acids of the *E. coli* AcfD (orf3526) protein.

31. The polysaccharide conjugate of any one of claims 26-29, wherein the carrier polypeptide does not comprise at least the last 100 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 125 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 150 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 175 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 200 C-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the last 210 C-terminal amino acids of the E. coli AcfD (orf3526) protein, or at least the last 217 C-terminal amino acids of the E. coli AcfD (orf3526) protein and optionally do not comprise at least the first 10 N-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the first 20 N-terminal amino acids of the E. coli AcfD (orf3526) protein, at least the first 30 N-terminal amino acids of the E. coli AcfD (orf3526) protein, or at least the first 33 N-terminal amino acids of the E. coli AcfD (orf3526) protein.

32. The polysaccharide conjugate of any one of claims 25-31 wherein the carrier polypeptide comprises:

(a) the amino acid sequence selected from the group consisting of SEQ ID NOs 26-40;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NOs: 26-40;

(c) at least 85% sequence identity to any one of SEQ ID NOs: 26-40;

(d) a fragment of at least 10 consecutive amino acids from any one of SEQ ID NOs: 26-40;

and/or

(e) when aligned with any of SEQ ID NOs: 26-40 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

33. The polysaccharide conjugate of claims 25-32, wherein the carrier polypeptide further contains a deletion relative to the *E. coli* AcfD (orf3526) protein which increases solubility of the carrier polypeptide as compared to the *E. coli* AcfD (orf3526) protein.

34. The polysaccharide conjugate of claim 33, wherein the deletion is removal of substantially all of the N-terminal amino acids up to the gly-ser region, removal of all or a part of the N-terminal proline-rich repeat, or both.

35. The polysaccharide conjugate of claim 33, wherein the deletion is removal of at least the first 10 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 20 N-

terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 30 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 33 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 40 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 50 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 60 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 70 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 80 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, at least the first 90 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein, or at least the first 94 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein.

36. The polysaccharide conjugate of claim 24 wherein the carrier polypeptide is the *E. coli* Flu antigen 43 protein (orf1364 polypeptide).

37. The polysaccharide conjugate of claim 36 wherein the carrier polypeptide is a fragment of the *E. coli* Flu antigen 43 protein (orf1364 polypeptide) wherein the fragment contains a deletion relative to the full length *E. coli* Flu antigen 43 protein (orf1364 polypeptide) which deletion increases solubility of the fragment as compared to the full length protein.

38. The polysaccharide conjugate of claim 37, wherein the deletion comprises the carboxyl-terminal β-barrel domain.

39. The polysaccharide conjugate of claim 38, wherein the carrier polypeptide corresponds to the amino acid sequence of SEQ ID NO:44.

40. The polysaccharide conjugate of any of claims 37-39, wherein the fragment comprises less than 950 amino acids, less than 900 amino acids, less than 850 amino acids, less than 800 amino acids, less than 750 amino acids, less than 700 amino acids, less than 650 amino acids, less than 600 amino acids, less than 550 amino acids, less than 500 amino acids, less than 450 amino acids, less than 440 amino acids, or less than 430 amino acids of the flu antigen 43 (orf1364) protein.

41. The polysaccharide conjugate of any one of claims 36-40 wherein the carrier polypeptide comprises:

(a) the amino acid sequence selected from the group consisting of SEQ ID NOs 1-22;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NOs: 1-22;

(c) at least 85% sequence identity to any one of SEQ ID NOs: 1-22;

(d) a fragment of at least 10 consecutive amino acids from any one of SEQ ID NOs:1-22;

and/or

(e) when aligned with any of SEQ ID NOs: 1-22 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

42. The polysaccharide conjugate of claim 41, wherein the fragment does not comprise at least the first 10 N-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the first 20 N-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the first 30 N-terminal amino acids as compared to *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the first 40 N-terminal amino acids as compared to *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the first 50 N-terminal amino acids as compared to *E. coli* Flu antigen 43 protein (orf1364 polypeptide), or at least the first 52 N-terminal amino acids as compared to the *E. coli* AcfD (orf3526) protein.

43. The polysaccharide conjugate of claim 41 or claim 42, wherein the fragment does not comprise at least the last 50 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 100 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 150 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 200 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 250 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 300 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), at least the last 325 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide), or at least the last 328 C-terminal amino acids as compared to the *E. coli* Flu antigen 43 protein (orf1364 polypeptide).

44. The polysaccharide conjugate of claim 24 wherein the carrier polypeptide is the *Escherichia* Sell repeat-containing protein (upec-5211 polypeptide).

45. The polysaccharide conjugate of claim 44 wherein the carrier polypeptide comprises:

(a) the amino acid sequence selected from the group consisting of SEQ ID NOs 23-25;

(b) from 1 to 10 single amino acid alterations compared to SEQ ID NOs: 23-25;

(c) at least 85% sequence identity to any one of SEQ ID NOs: 23-25;

(d) a fragment of at least 10 consecutive amino acids from any one of SEQ ID NOs:23-25;

and/or

(e) when aligned with any of SEQ ID NOs: 23-25 using a pairwise alignment algorithm, each moving window of x amino acids from N terminus to C terminus has at least x•y identical aligned amino acids, where x is 30 and y is 0.75.

46. The polysaccharide conjugate of any of claims 24-45, wherein the polysaccharide is selected from the list comprising:

(a) a glucan,

(b) a capsular saccharide from at least one of serogroups A, C, W135 and Y of *Neisseria meningitidis*,

(c) a saccharide antigen from *Streptococcus pneumoniae*,

(d) a capsular polysaccharide from *Staphylococcus aureus*,

(e) a *Haemophilus influenzae* B polysaccharide,

(f) a saccharide antigen from *Streptococcus aglactiae*,

(g) a lipopolysaccharide from *Vibrio cholerae*, or

(h) a capsular polysaccharide from *Salmonella typhi*.

47. The polysaccharide conjugate of any of claims 24-46 further comprising an adjuvant.

48. A vaccine component comprising the polysaccharide conjugate of claims 24-46.

49. A vaccine comprising the vaccine component of claim 48.

50. The vaccine of claim 49 further comprising an adjuvant.

51. The vaccine of claim 49 or claim 50 further comprising an additional vaccine component selected from: a *Neisseria meningitidis* antigen, a *Streptococcus pneumoniae* antigen, a *Streptococcus pyogenes* antigen, a *Moraxella catarrhalis* antigen, a *Bordetella pertussis* antigen, a *Staphylococcus aureus* antigen, a *Staphylococcus epidermis* antigen, a *Clostridium tetani* antigen, a *Cornynebacterium diphtheriae* antigen, a *Haemophilus influenzae* type B (Hib) antigen, a *Pseudomonas aeruginosa* antigen, a *Legionella pneumophila* antigen, a *Streptococcus agalactiae* antigen, a *Neiserria gonorrhoeae* antigen, a *Chlamydia trachomatis* antigen, a *Treponema pallidum* antigen, a *Haemophilus ducreyi* antigen, an *Enterococcus faecalis* antigen, an *Enterococcus faecium* antigen, a *Helicobacter pylori* antigen, a *Staphylococcus saprophyticus* antigen, a *Yersinia enterocolitica* antigen, an additional *E. coli* antigen, a *Bacillus anthracis* antigen, a *Yersinia pestis* antigen, a *Mycobacterium tuberculosis* antigen, a *Rickettsia* antigen, a *Listeria monocytogenes* antigen, a *Chlamydia pneumoniae* antigen, a *Vibrio cholerae* antigen, a *Salmonella typhi* antigen, a

Borrelia burgdorferi antigen, a *Porphyromonas gingivalis* antigen, a Shigella antigen and a *Klebsiella* antigen.

52. The method of inducing an enhanced immune response in a mammalian subject to polysaccharide comprising:

administering to the mammalian subject of the polysaccharide conjugate of claims 24-46, the vaccine component of claim 48, or the vaccine of claims 49-51.

53. The use of the polysaccharide conjugate of claims 24-46, the vaccine component of claim 48, or the vaccine of claims 49-51 to induce an enhanced immune response in a mammalian subject to the polysaccharide.

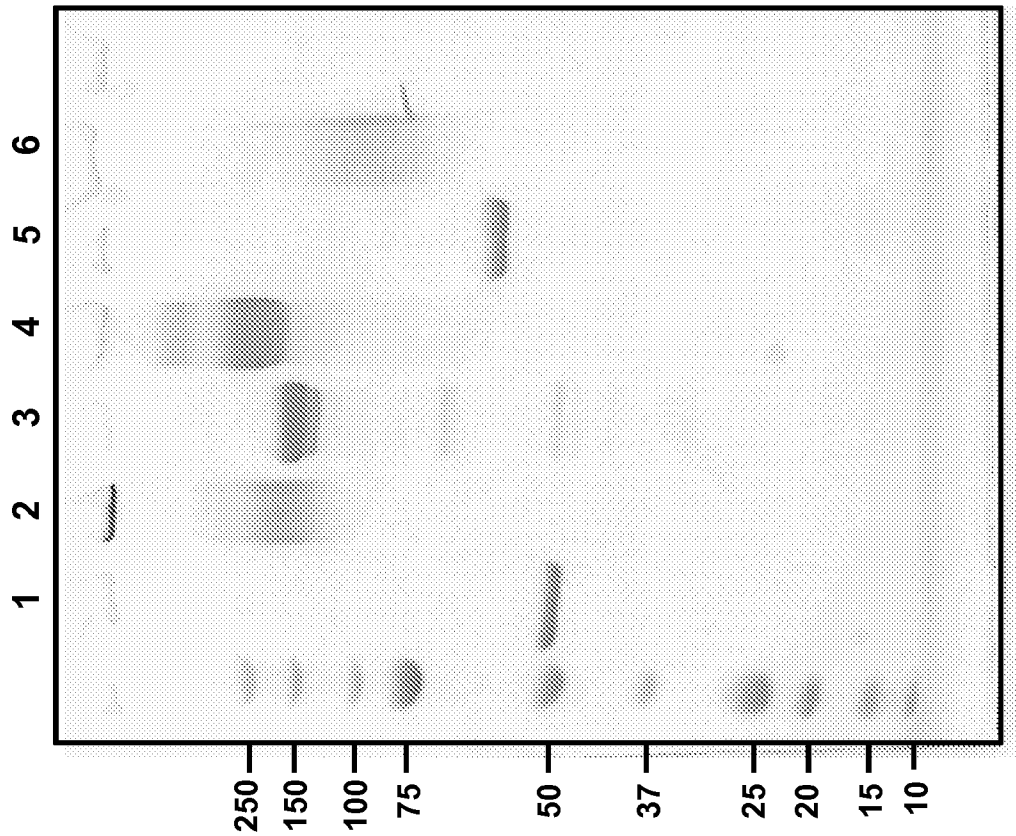


Fig. 1(b)

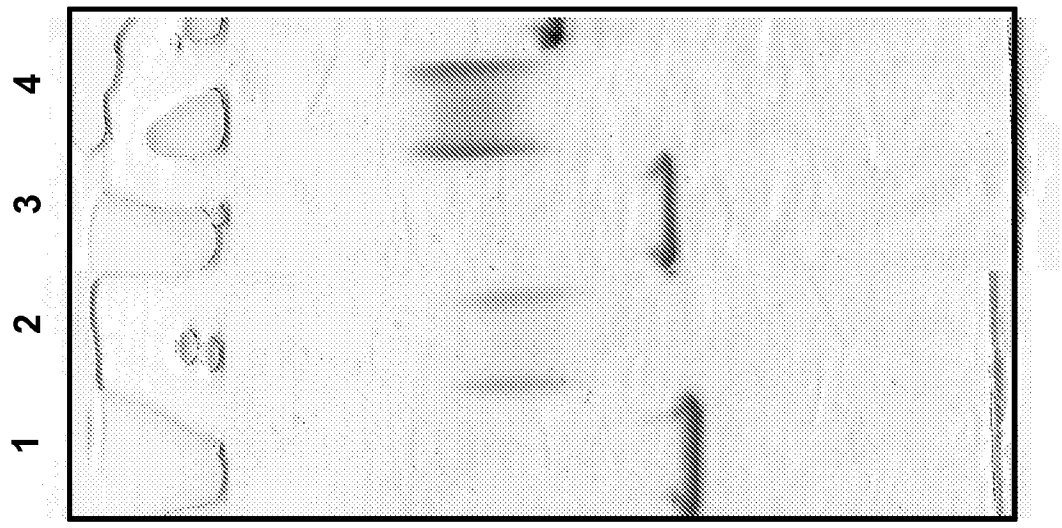


Fig. 1(a)

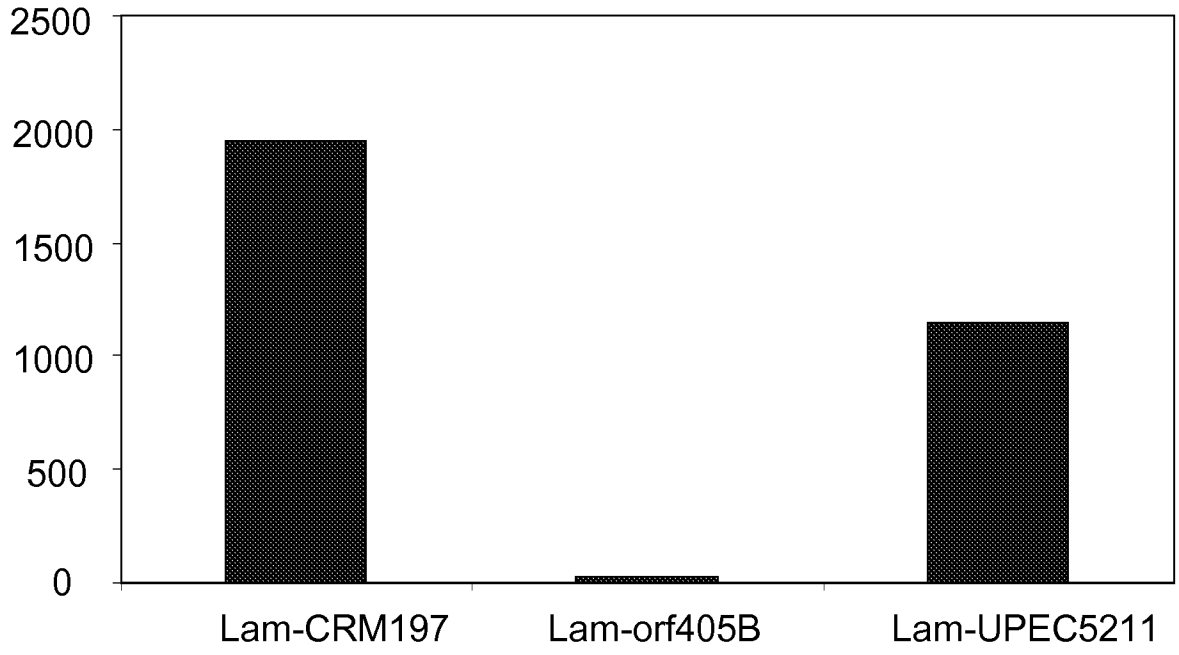

2500
2000
1500
1000
500
0
Lam-CRM197
Lam-orf405B
Lam-UPEC5211

Fig. 2

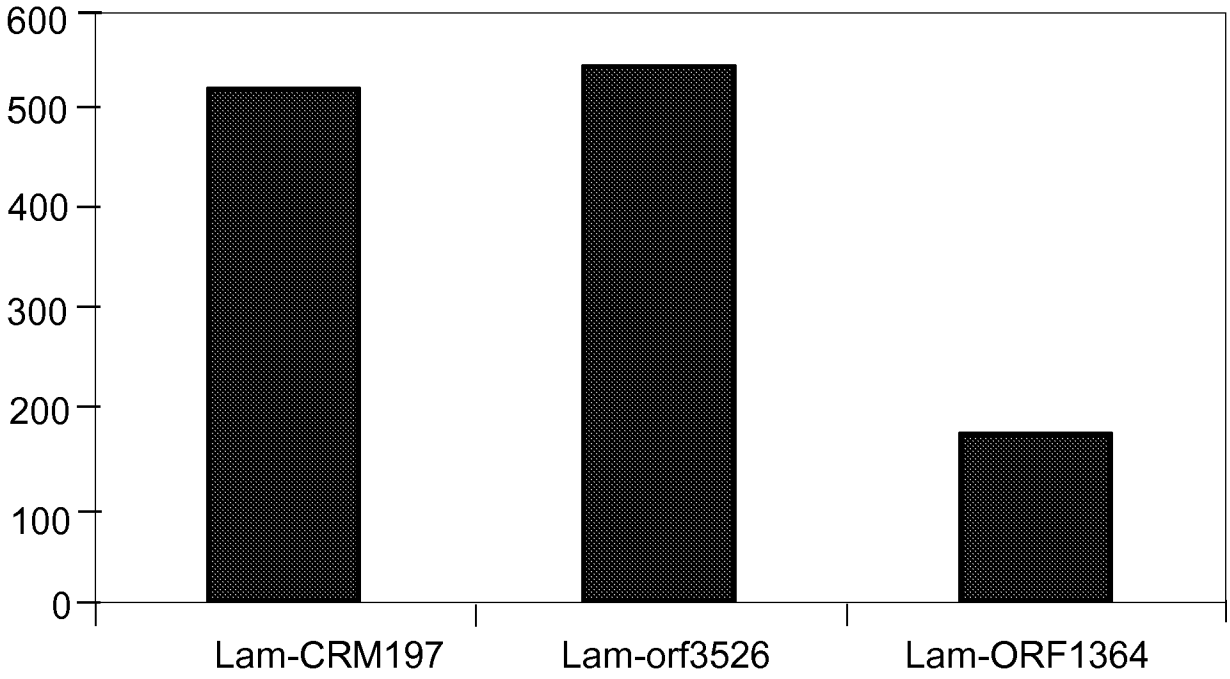

600
500
400
300
200
100
0
Lam-CRM197
Lam-orf3526
Lam-ORF1364

Fig. 3