

ESCHERICHIA COLI COMPOSITIONS AND METHODS THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefits of U.S. Provisional Application No. 63/254,195, filed October 11, 2021, U.S. Provisional Application No. 63/144,058, filed February 1, 2021, and
5 U.S. Provisional Application No. 63/106,077, filed October 27, 2020. The entire content of each of the foregoing applications is incorporated herein by reference.

REFERENCE TO SEQUENCE LISTING

This application is being filed electronically via EFS-Web and includes an electronically submitted sequence listing in .txt format. The .txt file contains a sequence listing entitled
10 "PC72671_ST25.txt" created on October 26, 2020 and having a size of 160 KB. The sequence listing contained in this .txt file is part of the specification and is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to new immunogenic compositions for use in a vaccine for
15 eliciting immune responses against *E. coli* and *Klebsiella pneumoniae* serotypes in a subject.

BACKGROUND OF THE INVENTION

Escherichia coli is one of the most common human bacterial pathogens with clinical presentations that include blood stream infections, urinary tract infections (catheter and non-catheter associated; surgical site infections, pneumonia, and serious food poisoning related
20 diarrhea. They are classified serologically by differences in the structure of the lipopolysaccharide-associated O-antigen (>180 known serotypes), the capsule polysaccharide K-antigen (> 100 serotypes), and the flagellar H-antigen (>50 serotypes).

Urinary tract infections (UTIs) most often present as a cystitis that in some individuals can recur repeatedly following resolution. Left untreated, they can progress to pyelonephritis and
25 blood stream infections. *E. coli* infections are associated with high levels of antibiotic resistance [Al-Hasan MN, et al. The Journal of antimicrobial chemotherapy 2009; 64:169-74] with many strains being resistant to multiple antibiotics including antibiotics of last resort such as carbapenems and polymyxins [Zowawi HM, et al. Nature reviews Urology 2015; 12:570-84]. In particular, O25b serotype multilocus sequence type (MLST) 131 has emerged as a worldwide
30 pandemic clone, causing predominantly community-onset infections with high rates of resistance to extended-spectrum cephalosporins (ESBLs) and fluoroquinolones [Rogers BA, et al. The Journal of antimicrobial chemotherapy 2011; 66:1-14; Nicolas-Chanoine M-H, et al.

Clinical Microbiology Reviews 2014; 27:543-74]. *E. coli* BSI and UTI infecting strains are also known as invasive Extra-intestinal Pathogenic *E. coli* (ExPEC) or uropathogenic *E. coli* (UPEC).

Second to *E. coli*, *Klebsiella* spp. (including *K. pneumoniae* and *K. oxytoca*) are the next most common Gram-negative pathogens associated with invasive infections including UTIs, pneumonia, intra-abdominal infection, and bloodstream infection (BSI) [Nicolas-Chanoine M-H, et al. Clinical Microbiology Reviews 2014; 27:543-74; Podschun R, et al. Clin Microbiol Rev 1998; 11:589-603; Yinnon AM, et al. QJM : monthly journal of the Association of Physicians 1996; 89:933-41; Anderson DJ, et al. PLoS One 2014; 9:e91713]. *Klebsiella* maintain a profound ability to acquire antibiotic resistance through horizontally transmissible ESBL and carbapenem resistance conferring genes [Chen L, et al. Trends Microbiol 2014; 22:686-96; Iredell J, et al. Bmj 2016; 352:h6420]. Accordingly, during the last decade the prevalence of ESBL-resistant *Klebsiella* producing extended-spectrum β -lactamases (ESBL) has increased dramatically globally. *Klebsiella* spp. can express up to 8 different O-types and >80 K-types. While there are a multitude of K-antigens associated with virulent *Klebsiella* strains, only four O-antigen serotypes account for $\geq$ 80% of *Klebsiella* clinical isolates irrespective of sample site (blood, urine, sputum), infection status (invasive versus non-invasive) or the nature of acquisition (community vs nosocomial) [Follador R, et al. Microbial Genomics 2016; 2:e000073].

The increased rate of invasive multidrug-resistant (MDR) *E. coli* and *Klebsiella* infections in the vulnerable newborn population and the elderly underscores the need for vaccine-based approaches as an alternative to standard-of-care antibiotics which are becoming less effective.

SUMMARY OF THE INVENTION

To meet these and other needs, the present invention relates to compositions and methods of use thereof for producing new immunogenic compositions for use in a vaccine for eliciting immune responses against *E. coli* and *Klebsiella pneumoniae* serotypes in a subject.

In some embodiments of the methods provided herein, the subject is a mammal, preferably a human. In some particular embodiments, the human is a child, such as an infant. In some other particular embodiments, the human is a woman, particularly a pregnant woman.

The composition may be administered to the subject with or without administration of an adjuvant. The effective amount administered to the subject is an amount that is sufficient to elicit an immune response against an *E. coli* or *K. pneumoniae* antigen in the subject. Subjects that can be selected for treatment include

those that are at risk for developing an *E. coli* or *K. pneumoniae* infection because of exposure or the possibility of exposure to *E. coli* or *K. pneumoniae*. Because humans may be infected with *E. coli* or *K. pneumoniae* by the age of 2, the entire birth cohort is included as a relevant population for immunization. This could be done, for example, by beginning an immunization regimen anytime from birth to 6 months of age, from 6 months of age to 5 years of age, in pregnant women (or women of child-bearing age) to protect their infants by passive transfer of antibody, infants still *in utero*, and subjects greater than 50 years of age.

In one aspect, the invention relates to a composition comprising a FimH polypeptide comprising amino acids having a sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 27, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113, or any combination thereof. In some embodiments, the composition further includes a saccharide selected from any saccharide having a Formula in **Table 1**, preferably Formula O1A, Formula O1B, Formula O2, Formula O6, and Formula O25B, wherein *n* is an integer from 1 to 100, preferably 31 to 100.

In one aspect, the invention relates to a composition that includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, and SEQ ID NO: 29, or any combination thereof.

In another aspect, the invention relates to a composition that includes a polypeptide having at least *n* consecutive amino acids from any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, and SEQ ID NO: 29, wherein *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). In some embodiments, the composition further includes a saccharide selected from any one Formula in **Table 1**, preferably Formula O1A, Formula O1B, Formula O2, Formula O6, and Formula O25B, wherein *n* is an integer from 1 to 100, preferably 31 to 100.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A-1H—depict amino acid sequences, including amino acid sequences for exemplary polypeptides derived from *E. coli* or fragments thereof; and amino acid sequences for exemplary wzzB sequences.

FIG. 2A-2T —depict maps of exemplary expression vectors.

FIG. 3 —depicts results from expression and purification.

FIG. 4 —depicts results from expression and purification.

FIG. 5 —depicts results from expression.

FIG. 6A-6C —depict pSB02083 and pSB02158 SEC pools and affinities, including yields.

- FIG. 7**– depicts results from expression of pSB2198 FimH dscG Lock Mutant construct.
- FIG. 8** – depicts results from expression of pSB2307 FimH dscG wild type.
- FIG. 9A-9C** – depict structures of O-antigens synthesized by the polymerase-dependent pathway with four or less residues in the backbone.
- 5 **FIG. 10A-10B** – **FIG. 10A** depicts structures of O-antigens synthesized by the polymerase-dependent pathway with five or six residues in the backbone; **FIG. 10B** depicts O-antigens believed to be synthesized by the ABC-transporter-dependent pathway.
- FIG. 11** – depicts computational mutagenesis scanning of Phe1 with other amino acids having aliphatic hydrophobic sidechains, e.g. Ile, Leu and Val, that may stabilize the FimH protein and
- 10 accommodate mannose binding.
- FIG. 12A-12B** –depict plasmids: a pUC replicon plasmid, 500-700x copies per cell, Chain length regulator (**FIG. 12A**); and P15a replicon plasmid, 10-12x copies per cell, O-antigen operon (**FIG. 12B**).
- FIG. 13A-13B** – depict modulation of O-antigen chain length in serotype O25a and O25b strains
- 15 by plasmid-based expression of heterologous *wzzB* and *fepE* chain length regulators. Genetic complementation of LPS expression in plasmid transformants of *wzzB* knockout strains O25K5H1 (O25a) and GAR2401 (O25b) is shown. On the left side of **FIG. 13A**, LPS profiles of plasmid transformants of O25a O25K5H1Δ*wzzB* are shown; and on the right, analogous profiles of O25b GAR 2401Δ*wzzB* transformants. An immunoblot of a replicate gel probed with O25-
- 20 specific sera (Statens Serum Institut) is shown in **FIG. 13B**. O25a Δ*wxxB* (Knock out) background associated with Lanes 1-7; O25b 2401 Δ*wzzB* (Knock out) background associated with Lanes 8-15.
- FIG. 14** – depicts long chain O-antigen expression conferred by *E. coli* and *Salmonella fepE* plasmids in host O25K5H1Δ*wzzB*.
- 25 **FIG. 15** – depicts that *Salmonella fepE* expression generates Long O-antigen LPS in a variety of clinical isolates.
- FIG. 16A-16B** – depict plasmid-mediated Arabinose-inducible Expression of O25b Long O-antigen LPS in O25b O-antigen knock-out host strain. Results from an SPS PAGE are shown in **FIG. 16A** and results from an O25 Immuno-Blot are shown in **FIG. 16B**, wherein Lane 1 is
- 30 from Clone 1, no arabinose; Lane 2 is from Clone 1, 0.2% arabinose; Lane 3 is from Clone 9, no Arabinose; Lane 4 is from Clone 9, 0.2% Arabinose; Lane 5 is from O55 *E. coli* LPS Standard; and Lane 6 is from O111 *E. coli* LPS Standard, in both **FIG. 16A** and in **FIG. 16B**.
- FIG. 17** – depicts plasmid-mediated Arabinose-inducible Expression of Long O-antigen LPS in common host strain.
- 35 **FIG. 18** – depicts expression of O25 O-antigen LPS in Exploratory Bioprocess strains.

FIG. 19A-19B – depict SEC profiles and properties of short (**FIG. 19A**, Strain 1 O25b wt 2831) and long O25b O-antigens (**FIG. 19B**, Strain 2 O25b 2401 Δ wzzB / LT2 FepE) purified from strains GAR2831 and '2401 Δ wzzB / fepE.

FIG. 20A-20B – depict vaccination schedules in rabbits: (**FIG. 20A**) Information regarding vaccination schedule for rabbit study 1 VAC-2017-PRL-EC-0723; (**FIG. 20B**) vaccination schedule for rabbit study 2 VAC-2018-PRL-EC-077.

FIG. 21A-21C – depict O25b Glycoconjugate IgG responses, wherein –●– represents results from Prebleed; –■– Bleed 1 (6 wk); –▲– Bleed 2 (8 wk); –◆– Bleed 3 (12 wk). **FIG. 21A** depicts results from Rabbit 1-3 (Medium Activation); **FIG. 21B** depicts results from Rabbit 2-3 (Low Activation); **FIG. 21C** depicts results from Rabbit 3-1 (High Activation).

FIG. 22A-22F – depict IgG responses to O25b Long O-antigen Glycoconjugate, i.e., Low activation O25b-CRM₁₉₇ conjugate (**FIG. 22D-22F**, wherein –●– represents results from Prebleed from Rabbit 2-1, –■– Week 12 Antisera from Rabbit 2-1) vs unconjugated polysaccharide, i.e., free O25b polysaccharide (**FIG. 22A-22C**, wherein –●– represents results from Prebleed from Rabbit A-1, –■– Week 6 Antisera from Rabbit A-1, –▲– Week 8 Antisera from Rabbit A-1). Note that MFIs are plotted on log scale to highlight differences between pre-immune and immune antibodies in the <1000 MFI range. **FIG. 22A** depicts results from Rabbit A-1 (Unconjugated Poly); **FIG. 22B** depicts results from Rabbit A-3 (Unconjugated Poly); **FIG. 22C** depicts results from Rabbit A-4 (Unconjugated Poly); **FIG. 22D** depicts results from Rabbit 2-1 (low activation); **FIG. 22E** depicts results from Rabbit 2-2 (low activation); and **FIG. 22F** depicts results from Rabbit 2-3 (low activation).

FIG. 23A-23C – depict surface expression of native vs long O25b O-antigen detected with O25b antisera. **FIG. 23A** depicts results wherein –●– represents results from O25b 2831 vs PD3 antisera; –■– represents results from O25b 2831 wt vs prebleed; –▲– represents results from O25b 2831 / fepE vs PD3 antisera; –▼– represents results from O25b 2831 / fepE vs prebleed. **FIG. 23B** depicts results wherein –●– represents results from O25b 2401 vs PD3 antisera; –■– represents results from O25b 2401 vs prebleed; –▲– represents results from O25b 2401 / fepE vs PD3 antisera; –▼– represents results from O25b 2401 / fepE vs prebleed. **FIG. 23C** depicts results wherein –●– represents results from *E. coli* K12 vs PD3 antisera; and –■– represents results from *E. coli* K12 vs prebleed.

FIG. 24 – depicts generalized structures of the carbohydrate backbone of the outer core oligosaccharides of the five known chemotypes. All glycoses are in the α -anomeric configuration unless otherwise indicated. The genes whose products catalyse formation of each linkage are indicated in dashed arrows. An asterisk denotes the residue of the core oligosaccharide to which attachment of O-antigen occurs.

- FIG. 25** – depicts that unconjugated free O25b polysaccharide is not immunogenic (dLIA), wherein –●– represents results from Week 18 (1wk = PD4) Antisera from 4-1; –■– represents results from Week 18 (1wk = PD4) Antisera from 4-2; –▲– represents results from Week 18 (1wk = PD4) Antisera from 5-1; –▼– represents results from Week 18 (1wk = PD4) Antisera from 5-2; –*– represents results from Week 18 (1wk = PD4) Antisera from 6-1; –▲– represents results from Week 18 (1wk = PD4) Antisera from 6-2.
- FIG. 26A-26C**– depict graphs illustrating the specificity of BRC Rabbit O25b RAC conjugate immune sera OPA titers. **FIG. 26A** shows OPA titers of Rabbit 2-3 pre-immune serum (–●–) and post-immune serum wk 13 (–■–). **FIG. 26B** shows OPA titers of Rabbit 1-2 pre-immune serum (–●–) and post-immune serum wk 19 (–■–). **FIG. 26C** shows Rabbit 1-2 wk 19 OPA Titer Specificity, in which OPA activity of Rabbit 1-2 immune serum is blocked by pre-incubation with 100µg/mL of purified unconjugated O25b long O-antigen polysaccharide, wherein –■– represents results from Rabbit 1-2 immune serum wk 19; and –▼– represents results from Rabbit 1-2 wk 19 w/R1 Long-OAg.
- FIG. 27A-27C** – **FIG. 27A** depicts an illustration of an exemplary administration schedule. **FIG. 27B** and **FIG. 27C** show graphs depicting O-antigen O25b IgG levels elicited by unconjugated O25b long O-antigen polysaccharide (**FIG. 27B**, O25b Free Poly (2µg)) and derived O25b RAC/DMSO long O-antigen glycoconjugate (**FIG. 27C**, O25b-CRM₁₉₇ RAC Long (2µg)), wherein –...– (dotted line) represents Naïve CD1 O25b IgG level.
- FIG. 28A-28B** – depict graphs showing OPA immunogenicity of RAC, eTEC O25b long glycoconjugates, and single end glycoconjugates post dose 2 (**FIG. 28A**) and post dose 3 (**FIG. 28B**), wherein –○– represents results from single end short 2 µg; –●– single end long 2µg; –▲– RAC/DMSO long 2µg; –▼– eTEC long 2µg; * Background control (n=20). †Responder rates are % mice with titers > 2x unvaccinated baseline.
- FIG. 29** – depicts graph showing OPA immunogenicity of eTEC chemistry and modified levels of polysaccharide activation. †Responder rates are % mice with titers > 2x unvaccinated baseline.
- FIG. 30A-30B** – depict an illustration of an exemplary administration schedule (**FIG. 30A**); and a graph depicting protection of mice immunized with doses of *E. coli* eTEC conjugates from lethal challenge with O25b isolate (**FIG. 30B**), wherein –◇– represents eTEC Long Chain 17% activation; –△– eTEC represents Long Chain 10% activation; –▽– represents eTEC Long Chain 4% activation; –□– represents O25b Polysaccharide; –○– represents unvaccinated controls.
- FIG. 31** – depicts a schematic illustrating an exemplary preparation of single-ended conjugates, wherein the conjugation process involves selective activation of 2-Keto-3-deoxyoctanoic acid

(KDO) with a disulfide amine linker, upon unmasking of a thiol functional group. The KDO is then conjugated to bromo activated CRM₁₉₇ protein as depicted in **FIG. 31** (Preparation of Single-Ended Conjugates).

- FIG. 32A-32B** – depict an exemplary process flow diagram for the activation (**FIG. 32A**) and conjugation (**FIG. 32B**) processes used in the preparation of *E. coli* glycoconjugate to CRM₁₉₇.
- FIG. 33** - depicts structures of the repeat unit (RU) of *E. coli* and *K. pneumoniae* polymannan O-antigens. *Legend*: Trimeric *E. coli* O8 and *K. pneumoniae* O5 are identical, as are the tetrameric *E. coli* O9A/ *K. pneumoniae* O3a and pentameric *E. coli* O9/ *K. pneumoniae* O3. Differentiation of the *K. pneumoniae* O3 subtypes at the level of biosynthetic enzyme sequences is described in Guachalla LM et al. (*Scientific Reports* 2017; 7:6635).
- FIG. 34A-34B** – depict *E. coli* serotype O8 immune sera is bactericidal against an invasive *K. pneumoniae* serotype O5 strain. *Legend*: Rabbit immune sera elicited by an *E. coli* serotype O8 O-antigen CRM₁₉₇ conjugate was evaluated in bactericidal assays with an *E. coli* O8 strain (**FIG. 34A**) and a *K. pneumoniae* O5 strain (**FIG. 34B**). Potent opsonophagocytic assay (OPA) activity against an *E. coli* O8 strain was observed after two vaccine doses (week 15) that was absent following preadsorption with unconjugated O8 polysaccharide (O8-OAg), or with matched pre-immune sera (week 0). The same rabbit immune serum showed antigen-specific serum bactericidal activity (SBA) against the *K. pneumoniae* O5 strain. BRC – baby rabbit complement, hC – IgG/IgM depleted human sera as complement source.
- FIG. 35A-35B** – depict *E. coli* serotype O9 O-antigen immune sera is bactericidal against an invasive *K. pneumoniae* O3 isolate. *Legend*: Rabbit immune sera elicited by an *E. coli* serotype O9a O-antigen CRM₁₉₇ conjugate was evaluated in opsonophagocytic assays (OPAs) with an *E. coli* O9a strain (**FIG. 35A**) and a *K. pneumoniae* O3b strain (**FIG. 35B**). OPA activity against the *E. coli* O9 strain was observed after two vaccine doses (week 15) that was absent following preadsorption with unconjugated O9 polysaccharide (O9-OAg), or with matched pre-immune sera (week 0). The same rabbit immune serum also showed potent antigen-specific serum bactericidal activity (SBA) against the *K. pneumoniae* O3b strain. BRC, baby rabbit complement; hC, IgG/IgM depleted human sera used as complement source.
- FIG. 36A-36B** – depict that the *E. coli* FimH_{LD} Antigen requires a potent adjuvant to elicit neutralizing antibodies. **FIG. 36A** depicts the dosing schedule of Study VAC-2019-PRL-EC-1369: between 12 and 20 CD-1 mice per group were vaccinated with 3 µg or 30 µg doses of *E. coli* FimH_{LD} antigen with or without 20 µg QS21/PS80 or 50 µg AlPO₄ adjuvant. **FIG. 36B** depicts the titers after dose 3 of individual mice vaccinated with wild-type FimH_{LD} or FimH_{LD} lock mutant antigens which are shown as closed symbols or open symbols, respectively. *p* values from t-tests (unpaired Welch's correction) of log transformed neutralization titer data are indicated.

- FIG. 37A-37B** – demonstrates that FimH-DSG variants are more immunogenic than FimH_{LD} constructs. **FIG. 37A.** depicts the dosing schedule of Study VAC-2019-PRL-EC-1438: 20 CD-1 mice per group were vaccinated with 10 µg or 30 µg of *E. coli* FimH_{LD} or FimH-DSG antigen variants with 20 µg QS21/PS80. **FIG. 37B** – depicts the titers after dose 3 of individual mice vaccinated with wild-type FimH_{LD} or FimH_{LD} lock mutant antigens which are shown as closed symbols or open symbols, respectively. *p* values from t-tests (unpaired Welch's correction) of log transformed neutralization titer data are indicated.
- FIG. 38** – depicts the schedule and dosing for a FimH-DSG and O-antigen combination and adjuvant formulation study (Study VAC-2020-PRL-EC-1679).
- FIG. 39A-39B** – depict Study VAC-2020-PRL-EC-1679 which demonstrates the influence of adjuvant and FimH-DSG and 4-valent O-antigen combinations on FimH neutralization. Yeast mannan assay binding neutralization titers at PD2 (**FIG. 39A**) and PD3 (**FIG. 39B**) timepoints. Closed symbols reflect mice vaccinated with O-antigens or FimH-DSG antigens only; open symbols are mice in groups vaccinated with FimH-DSG O-antigen combinations as labeled on x-axis.
- FIG. 40A-40B** – depict results of bladder assay with mouse serum pools corroborate yeast mannan assay data with individual mouse sera. Bladder cell binding inhibition curves from pooled mouse sera (n=10). Details are same as in the legend to **FIG. 39A-39B**. Rabbit anti-FimH positive control serum titrations are also shown, as an invariant assay reference standard.
- FIG. 41** – depicts the influence of adjuvant and FimH-DSG on O-Antigen specific serum IgG levels. Closed and open symbols represent PD2 and PD3 timepoint values, respectively. Ten pre-vaccinated CD-1 mice per group were pooled to determine the baseline IgG level for each antigen. The GMT of each pre-vaccinated pool was < the lower limit of quantitation (0.15 µg/mL, dotted line), determined from standard curve bias analysis. Responder rates are indicated as the percentage of mice per group that show greater than a 5-fold increase in titer over baseline. Differences in antibody titers between groups were analyzed by unpaired t test with Welch's correction **P < 0.05.
- FIG. 42** – depicts the influence of adjuvant and FimH-DSG on O-antigen-specific serum OPA titers. Details are same as for **FIG. 41**. Dotted lines reflect baseline titers from mouse pre-vaccination serum pools (1/2x LOD or 50, n=10). Closed and open symbols represent PD2 and PD3 timepoint values, respectively.

SEQUENCE IDENTIFIERS

- SEQ ID NO: 1 sets forth an amino acid sequence for a wild type type 1 fimbriae D-mannose specific adhesin [*Escherichia coli* FimH J96].
- SEQ ID NO: 2 sets forth an amino acid sequence for a fragment of FimH, corresponding to aa residues 22-300 of SEQ ID NO: 1 (mature FimH protein).

- SEQ ID NO: 3 sets forth an amino acid sequence for a FimH lectin domain.
- SEQ ID NO: 4 sets forth an amino acid sequence for a FimH pilin domain.
- SEQ ID NO: 5 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02198 - FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S N249Q / 7
- 5 AA linker / FimG A1..K14 / GGH8 in pcDNA3.1(+))
- SEQ ID NO: 6 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02307 - FimH mlgK signal pept / F22..Q300 J96 FimH N28S N91S N249Q / His8 in pcDNA3.1(+))
- SEQ ID NO: 7 sets forth an amino acid sequence for a fragment of a polypeptide derived from
- 10 *E. coli* FimH (pSB02083 FimH Lectin Domain Wild Type construct)
- SEQ ID NO: 8 sets forth an amino acid sequence for a fragment of a polypeptide derived from *E. coli* FimH (pSB02158 FimH Lectin Domain Lock Mutant)
- SEQ ID NO: 9 sets forth an amino acid sequence for a fragment of a polypeptide derived from *E. coli* FimG (FimG A1..K14)
- 15 SEQ ID NO: 10 sets forth an amino acid sequence for a fragment of a polypeptide derived from *E. coli* FimC.
- SEQ ID NO: 11 sets forth an amino acid sequence for a 4 aa linker.
- SEQ ID NO: 12 sets forth an amino acid sequence for a 5 aa linker.
- SEQ ID NO: 13 sets forth an amino acid sequence for a 6 aa linker.
- 20 SEQ ID NO: 14 sets forth an amino acid sequence for a 7 aa linker.
- SEQ ID NO: 15 sets forth an amino acid sequence for a 8 aa linker.
- SEQ ID NO: 16 sets forth an amino acid sequence for a 9 aa linker.
- SEQ ID NO: 17 sets forth an amino acid sequence for a 10 aa linker.
- SEQ ID NO: 18 sets forth an amino acid sequence for a FimH J96 signal sequence.
- 25 SEQ ID NO: 19 sets forth an amino acid sequence for the signal peptide of SEQ ID NO: 5 (pSB02198 - FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S N249Q / 7 AA linker / FimG A1..K14 / GGH8 in pcDNA3.1(+)).
- SEQ ID NO: 20 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH according to SEQ ID NO: 5 (mature protein of pSB02198 - FimH mlgK signal pept / F22..Q300
- 30 J96 FimH N28S V48C L55C N91S N249Q / 7 AA linker / FimG A1..K14 / GGH8 in pcDNA3.1(+)).
- SEQ ID NO: 21 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimG.
- SEQ ID NO: 22 sets forth an amino acid sequence for the signal peptide of SEQ ID NO: 6 (pSB02307 - FimH mlgK signal pept / F22..Q300 J96 FimH N28S N91S N249Q / His8 in
- 35 pcDNA3.1(+)).

- SEQ ID NO: 23 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH according to SEQ ID NO: 6 (mature protein of FimH mlgK signal pept / F22..Q300 J96 FimH N28S N91S N249Q / His8 in pcDNA3.1(+)).
- 5 SEQ ID NO: 24 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH according to SEQ ID NO: 7 (mature protein of pSB02083 FimH Lectin Domain Wild Type construct).
- SEQ ID NO: 25 sets forth an amino acid sequence for a His-tag.
- SEQ ID NO: 26 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH according to SEQ ID NO: 8 (mature protein of pSB02158 FimH Lectin Domain Lock Mutant)
- 10 SEQ ID NO: 27 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB01878).
- SEQ ID NO: 28 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (K12).
- SEQ ID NO: 29 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (UTI89).
- 15 SEQ ID NO: 30 sets forth a O25b 2401 WzzB amino acid sequence.
- SEQ ID NO: 31 sets forth a O25a:K5:H1 WzzB amino acid sequence.
- SEQ ID NO: 32 sets forth a O25a ETEC ATCC WzzB amino acid sequence.
- SEQ ID NO: 33 sets forth a K12 W3110 WzzB amino acid sequence.
- 20 SEQ ID NO: 34 sets forth a Salmonella LT2 WzzB amino acid sequence.
- SEQ ID NO: 35 sets forth a O25b 2401 FepE amino acid sequence.
- SEQ ID NO: 36 sets forth a O25a:K5:H1 FepE amino acid sequence.
- SEQ ID NO: 37 sets forth a O25a ETEC ATCC FepE amino acid sequence.
- SEQ ID NO: 38 sets forth a O157 FepE amino acid sequence.
- 25 SEQ ID NO: 39 sets forth a Salmonella LT2 FepE amino acid sequence.
- SEQ ID NO: 40 sets forth a primer sequence for LT2wzzB_S.
- SEQ ID NO: 41 sets forth a primer sequence for LT2wzzB_AS.
- SEQ ID NO: 42 sets forth a primer sequence for O25bFepE_S.
- SEQ ID NO: 43 sets forth a primer sequence for O25bFepE_A.
- 30 SEQ ID NO: 44 sets forth a primer sequence for wzzB P1_S.
- SEQ ID NO: 45 sets forth a primer sequence for wzzB P2_AS.
- SEQ ID NO: 46 sets forth a primer sequence for wzzB P3_S.
- SEQ ID NO: 47 sets forth a primer sequence for wzzB P4_AS.
- SEQ ID NO: 48 sets forth a primer sequence for O157 FepE_S.
- 35 SEQ ID NO: 49 sets forth a primer sequence for O157 FepE_AS.
- SEQ ID NO: 50 sets forth a primer sequence for pBAD33_adaptor_S.

- SEQ ID NO: 51 sets forth a primer sequence for pBAD33_adaptor_AS.
- SEQ ID NO: 52 sets forth a primer sequence for JUMPSTART_r.
- SEQ ID NO: 53 sets forth a primer sequence for gnd_f.
- SEQ ID NO: 54 sets forth an amino acid sequence for a mouse IgK signal sequence.
- 5 SEQ ID NO: 55 sets forth an amino acid sequence for a human IgG receptor FcRn large subunit p51 signal peptide.
- SEQ ID NO: 56 sets forth an amino acid sequence for a human IL10 protein signal peptide.
- SEQ ID NO: 57 sets forth an amino acid sequence for a human respiratory syncytial virus A (strain A2) fusion glycoprotein F0 signal peptide.
- 10 SEQ ID NO: 58 sets forth an amino acid sequence for an influenza A hemagglutinin signal peptide.
- SEQ ID NOs: 59-101 set forth amino acid and nucleic acid sequences for a nanostructure-related polypeptide or fragment thereof.
- SEQ ID NOs: 102-109 set forth SignalP 4.1 (DTU Bioinformatics) sequences from various
- 15 species used for signal peptide predictions.
- SEQ ID NO: 110 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02083 -- FimH_{LD} (mlgK signal pept, N28S, N91S)).
- SEQ ID NO: 111 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02158 -- FimH_{LD}-LM (mlgK signal pept, N28S N91S V48C L55C)).
- 20 SEQ ID NO: 112 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02307 -- FimH-DSG (mlgK signal pept, N28S N91S N249Q 7aa linker FimG A1..K14)).
- SEQ ID NO: 113 sets forth an amino acid sequence for a polypeptide derived from *E. coli* FimH (pSB02198 -- FimH-DSG-LM (mlgK signal pept, N28S N91S 249Q V48C L55C 7aa linker FimG A1..K14)).

25

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure relates to compositions comprising *E. coli* FimH polypeptides and O-antigen glycoconjugates, methods for producing and purifying the compositions, and methods of using said compositions.

- 30 In one aspect, the invention includes a composition including the FimH polypeptide or fragment thereof described herein. The composition may include a polypeptide or fragment thereof that is suitable for in vivo administration. For example, the polypeptide or fragment thereof in such a composition may have a purity of at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at
- 35 least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, by mass. In an embodiment the polypeptide in such a composition may have a purity of at least 95% by mass.

In an embodiment the polypeptide in such a composition may have a purity of at least 97% by mass. In an embodiment the polypeptide in such a composition may have a purity of at least 98% by mass. In an embodiment the polypeptide in such a composition may have a purity of at least 99% by mass. The composition may further comprise an adjuvant.

In a further aspect, the invention includes a composition for use in inducing an immune response against *E. coli* or *E.coli* infection. Use of the composition described herein for inducing an immune response against *E. coli* or *E.coli* infection and use of the composition described herein in the manufacture of a medicament for inducing an immune response against *E. coli* or *E.coli* infection, are also disclosed.

Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it was individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as"), provided herein is intended merely to further illustrate the disclosure and does not pose a limitation on the scope of the claims. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the disclosure.

Several documents are cited throughout the text of this disclosure. Each of the documents cited herein (including all patents, patent applications, scientific publications, manufacturer's specifications, instructions, etc.), whether supra or infra, are hereby incorporated by reference in their entirety. Nothing herein is to be construed as an admission that the present disclosure was not entitled to antedate such disclosure.

Definitions

As used herein the term "about" means approximately or nearly, and in the context of a numerical value or range set forth herein in one embodiment means $\pm 20\%$, $\pm 10\%$, $\pm 5\%$, or $\pm 3\%$ of the numerical value or range recited or claimed.

The terms "a" and "an" and "the" and similar reference used in the context of describing the disclosure (especially in the context of the claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context.

"Fragment", with reference to an amino acid sequence (peptide or protein), relates to a part of an amino acid sequence, i.e. a sequence which represents the amino

acid sequence shortened at the N-terminus and/or C-terminus. A fragment shortened at the C-terminus (N-terminal fragment) is obtainable e.g. by translation of a truncated open reading frame that lacks the 3'-end of the open reading frame. A fragment shortened at the N-terminus (C-terminal fragment) is obtainable e.g. by translation of a truncated open reading frame that lacks the 5'-end of the open reading frame, as long as the truncated open reading frame comprises a start codon that serves to initiate translation. A fragment of an amino acid sequence comprises e.g. at least 50 %, at least 60 %, at least 70 %, at least 80%, at least 90% of the amino acid residues from an amino acid sequence. A fragment of an amino acid sequence preferably comprises at least 6, in particular at least 8, at least 12, at least 15, at least 20, at least 30, at least 50, or at least 100 consecutive amino acids from an amino acid sequence.

As used herein, the term "wild type" or "WT" or "native" refers to an amino acid sequence that is found in nature, including allelic variations. A wild type amino acid sequence, peptide or protein has an amino acid sequence that has not been intentionally modified.

As used herein, "variants" of an amino acid sequence (peptide, protein or polypeptide), or "mutants", or reference to a "mutated" polypeptide, comprise amino acid insertion variants/mutants, amino acid addition variants/mutants, amino acid deletion variants/mutants and/or amino acid substitution variants/mutants. The term "variant" or "mutant" includes all mutants, splice variants, posttranslationally modified variants, conformations, isoforms, allelic variants, species variants, and species homologs, in particular those which are naturally occurring. The term "variant" or "mutant" includes, in particular, fragments of an amino acid sequence.

Amino acid insertion variants comprise insertions of single or two or more amino acids in a particular amino acid sequence. In the case of amino acid sequence variants having an insertion, one or more amino acid residues are inserted into a particular site in an amino acid sequence, although random insertion with appropriate screening of the resulting product is also possible. Amino acid addition variants comprise amino- and/or carboxy-terminal fusions of one or more amino acids, such as 1, 2, 3, 5, 10, 20, 30, 50, or more amino acids. Amino acid deletion variants are characterized by the removal of one or more amino acids from the sequence, such as by removal of 1, 2, 3, 5, 10, 20, 30, 50, or more amino acids. The deletions may be in any position of the protein. Amino acid deletion variants that comprise the deletion at the N-terminal and/or C-terminal end of the protein are also called N-terminal and/or C-terminal truncation variants. Amino acid substitution variants are characterized by at least one residue in the sequence being removed and another residue being inserted in its place. Preference is given to the modifications being in positions in the amino acid sequence which are not conserved between homologous proteins or peptides and/or to replacing amino acids with other

ones having similar properties. Preferably, amino acid changes in peptide and protein variants are conservative amino acid changes, i.e., substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring

5 amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In one

10 embodiment, conservative amino acid substitutions include substitutions within the following groups:

- glycine, alanine;
- valine, isoleucine, leucine;
- aspartic acid, glutamic acid;
- 15 asparagine, glutamine;
- serine, threonine;
- lysine, arginine; and
- phenylalanine, tyrosine.

Preferably the degree of similarity, preferably identity between a given amino

20 acid sequence and an amino acid sequence which is a variant of said given amino acid sequence will be at least about 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. The degree of similarity or identity is given preferably for an amino acid region which is at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least

25 about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90% or about 100% of the entire length of the reference amino acid sequence. For example, if the reference amino acid sequence consists of 200 amino acids, the degree of similarity or identity is given preferably for at least about 20, at least about 40, at least about 60, at least about 80, at least about 100, at least about 120, at least about 140, at

30 least about 160, at least about 180, or about 200 amino acids, in some embodiments continuous amino acids. In some embodiments, the degree of similarity or identity is given for the entire length of the reference amino acid sequence. The alignment for determining sequence similarity, preferably sequence identity can be done with art known tools, preferably using the best sequence alignment, for example, using Align,

35 using standard settings, preferably EMBOSS:needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5.

As used herein, "sequence similarity" indicates the percentage of amino acids that either are identical or that represent conservative amino acid substitutions. "Sequence identity" between two amino acid sequences indicates the percentage of amino acids that are identical between the sequences. "Sequence identity" between two nucleic acid sequences indicates the percentage of nucleotides that are identical between the sequences.

The terms "% identical", "% identity" or similar terms are intended to refer, in particular, to the percentage of nucleotides or amino acids which are identical in an optimal alignment between the sequences to be compared. Said percentage is purely statistical, and the differences between the two sequences may be but are not necessarily randomly distributed over the entire length of the sequences to be compared. Comparisons of two sequences are usually carried out by comparing the sequences, after optimal alignment, with respect to a segment or "window of comparison", in order to identify local regions of corresponding sequences. The optimal alignment for a comparison may be carried out manually or with the aid of the local homology algorithm by Smith and Waterman, 1981, *Adv. App. Math.* 2, 482, with the aid of the local homology algorithm by Needleman and Wunsch, 1970, *J. Mol. Biol.* 48, 443, with the aid of the similarity search algorithm by Pearson and Lipman, 1988, *Proc. Natl Acad. Sci. USA* 88, 2444, or with the aid of computer programs using said algorithms (GAP, BESTFIT, FASTA, BLAST P, BLAST N and TFASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.). In some embodiments, percent identity of two sequences is determined using the BLASTN or BLASTP algorithm, as available on the United States National Center for Biotechnology Information (NCBI) website (e.g., at blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LC=align2seq). In some embodiments, the algorithm parameters used for BLASTN algorithm on the NCBI website include: (i) Expect Threshold set to 10; (ii) Word Size set to 28; (iii) Max matches in a query range set to 0; (iv) Match/Mismatch Scores set to 1, -2; (v) Gap Costs set to Linear; and (vi) the filter for low complexity regions being used. In some embodiments, the algorithm parameters used for BLASTP algorithm on the NCBI website include: (i) Expect Threshold set to 10; (ii) Word Size set to 3; (iii) Max matches in a query range set to 0; (iv) Matrix set to BLOSUM62; (v) Gap Costs set to Existence: 11 Extension: 1; and (vi) conditional compositional score matrix adjustment.

Percentage identity is obtained by determining the number of identical positions at which the sequences to be compared correspond, dividing this number by the number of positions compared (e.g., the number of positions in the reference sequence) and multiplying this result by 100.

In some embodiments, the degree of similarity or identity is given for a region which is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about

90% or about 100% of the entire length of the reference sequence. For example, if the reference nucleic acid sequence consists of 200 nucleotides, the degree of identity is given for at least about 100, at least about 120, at least about 140, at least about 160, at least about 180, or about 200 nucleotides, in some embodiments continuous nucleotides. In some embodiments, the degree of similarity or identity is given for the entire length of the reference sequence.

Homologous amino acid sequences exhibit according to the disclosure at least 40%, in particular at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and preferably at least 95%, at least 98 or at least 99% identity of the amino acid residues. The amino acid sequence variants/mutants described herein may readily be prepared by the skilled person, for example, by recombinant DNA manipulation. The manipulation of DNA sequences for preparing peptides or proteins having substitutions, additions, insertions or deletions, is described in detail in Sambrook et al. (1989), for example. Furthermore, the peptides and amino acid variants described herein may be readily prepared with the aid of known peptide synthesis techniques such as, for example, by solid phase synthesis and similar methods.

In one aspect, a fragment or variant/mutant of an amino acid sequence (peptide or protein) is preferably a "functional fragment" or "functional variant". The term "functional fragment" or "functional variant/mutant" of an amino acid sequence relates to any fragment or variant/mutant exhibiting one or more functional properties identical or similar to those of the amino acid sequence from which it is derived, i.e., it is functionally equivalent. With respect to antigens or antigenic sequences, one particular function is one or more immunogenic activities displayed by the amino acid sequence from which the fragment or variant is derived. The term "functional fragment" or "functional variant/mutant", as used herein, in particular refers to a variant/mutant molecule or sequence that comprises an amino acid sequence that is altered by one or more amino acids compared to the amino acid sequence of the parent molecule or sequence and that is still capable of fulfilling one or more of the functions of the parent molecule or sequence, e.g., inducing an immune response. In one aspect, the modifications in the amino acid sequence of the parent molecule or sequence do not significantly affect or alter the characteristics of the molecule or sequence. In different embodiments, the function of the functional fragment or functional variant may be reduced but still significantly present, e.g., immunogenicity of the functional variant may be at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the parent molecule or sequence. However, in other embodiments, immunogenicity of the functional fragment or functional variant may be enhanced compared to the parent molecule or sequence.

As used herein, "isolated" means altered or removed from the natural state. For example, a nucleic acid or a peptide naturally present in a living animal is not "isolated", but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is "isolated". An isolated nucleic acid or protein
 5 can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

I. Polypeptides Derived from *E. coli* and Fragments Thereof

In one aspect, disclosed herein is a mammalian cell that includes a polynucleotide that
 10 encodes a polypeptide derived from *E. coli* or a fragment thereof. The term "derived from" as used herein refers to a polypeptide that comprises an amino acid sequence of a FimH polypeptide or FimCH polypeptide complex or a fragment thereof as described herein that has been altered by the introduction of an amino acid residue substitution, deletion or addition. Preferably, the polypeptide derived from *E. coli* or a fragment thereof includes a sequence
 15 having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has at least 85% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or
 20 fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has at least 90% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has at least 95% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has at least 98% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or
 25 fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has at least 99% identity to the sequence of the corresponding wild-type *E. coli* FimH polypeptide or fragment. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof has the identical total length of amino acids as the corresponding wild-type FimH polypeptide or
 30 FimCH polypeptide complex or a fragment thereof. In some embodiments, the polypeptide derived from *E. coli* has the identical total length of amino acids as the corresponding wild-type FimH polypeptide or FimCH polypeptide complex.

The fragments should include at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or
 35 more). Preferably the fragments include an epitope from the sequence. In some embodiments, the fragment includes an amino acid sequence of at least 50 consecutive amino acid residues,

at least 100 consecutive amino acid residues, at least 125 consecutive amino acid residues, at least 150 consecutive amino acid residues, at least 175 consecutive amino acid residues, at least 200 consecutive amino acid residues, or at least 250 consecutive amino acid residues of the amino acid sequence of a polypeptide derived from *E. coli*.

- 5 In some embodiments, the fragment includes an amino acid sequence of at least 50 consecutive amino acid residues of the amino acid sequence of a polypeptide derived from *E. coli*. In some embodiments, the fragment includes an amino acid sequence of at least 100 consecutive amino acid residues of the amino acid sequence of a polypeptide derived from *E. coli*. In some embodiments, the fragment includes an amino acid
- 10 sequence of at least 150 consecutive amino acid residues of the amino acid sequence of a polypeptide derived from *E. coli*. In some embodiments, the fragment includes an amino acid sequence of at least 200 consecutive amino acid residues of the amino acid sequence of a polypeptide derived from *E. coli*. In some embodiments, the fragment includes an amino acid sequence of at least 250 consecutive amino acid residues of the
- 15 amino acid sequence of a polypeptide derived from *E. coli*.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes one or more non-classical amino acids, as compared to a corresponding wild-type *E. coli* FimH polypeptide or fragment.

- As used herein, the term "FimH polypeptide" refers to any FimH polypeptide or
- 20 fragment thereof as described herein that has been altered by the introduction of an amino acid substitution, deletion or addition, any FimH domain of the full-length wild type *E. coli* FimH polypeptide, any combination of domains of the full-length wild type *E. coli* FimH polypeptide, or the full-length *E. coli* FimH polypeptide. For example, in one embodiment the present disclosure provides a FimH polypeptide that is a FimH_{LD}
- 25 polypeptide, or a FimH-DSG polypeptide.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof possess a similar or identical function as a corresponding wild-type FimH polypeptide or a fragment thereof.

- In a preferred embodiment, polypeptides or polypeptide complexes or fragments
- 30 thereof of the invention are isolated or purified.

In some embodiments, the polynucleotide encoding the polypeptide derived from *E. coli* or a fragment thereof is integrated into the genomic DNA of the mammalian cell, and, when cultured in a suitable condition, said polypeptide derived from *E. coli* or a fragment thereof is expressed by the mammalian cell.

- 35 In a preferred embodiment, the polypeptide derived from *E. coli* or a fragment thereof is soluble.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof is secreted from the mammalian host cell.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may include additional amino acid residues, such as N-terminal or C-terminal extensions. Such extensions may include one or more tags, which may facilitate detection (e.g. an epitope tag for detection by monoclonal antibodies) and/or purification (e.g. a polyhistidine-tag to allow purification on a nickel-chelating resin) of the polypeptide or fragment thereof. In some embodiments, the tag includes the amino acid sequence selected from any one of SEQ ID NO: 21 and SEQ ID NO: 25. Such affinity-purification tags are known in the art. Examples of affinity-purification tags include, e.g., His tag (hexahistidine, which may, for example, bind to metal ion), maltose-binding protein (MBP), which may, for example, bind to amylose), glutathione-S- transferase (GST), which may, for example, bind to glutathione, FLAG tag, which may, for example, bind to an anti-flag antibody), Strep tag, which may, for example, bind to streptavidin or a derivative thereof). In preferred embodiments, the polypeptide derived from *E. coli* or a fragment thereof does not include additional amino acid residues, such as N-terminal or C-terminal extensions. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof described herein does not include an exogenous tag sequence.

While specific strains of *E. coli* may be referenced herein, it should be understood that the polypeptide derived from *E. coli* or a fragment thereof are not limited to specific strains unless specified.

In some embodiments, the polypeptide derived from *E. coli* FimH or a fragment thereof includes a phenylalanine residue at the N-terminus of the polypeptide. In some embodiments, the polypeptide derived from FimH or fragment thereof includes a phenylalanine residue within the first 20 residue positions of the N-terminus. Preferably, the phenylalanine residue is located at position 1 of the polypeptide. For example, in some embodiments, the polypeptide derived from *E. coli* FimH or a fragment thereof does not include an additional glycine residue at the N-terminus of the polypeptide derived from *E. coli* FimH or a fragment thereof.

In some embodiments, the phenylalanine residue at position 1 of the wild-type mature *E. coli* FimH is replaced by an aliphatic hydrophobic amino acid, such as, for example, any one of Ile, Leu and Val residues.

In some embodiments, a signal peptide may be used for expressing the polypeptide derived from *E. coli* or a fragment thereof. Signal sequences and expression cassettes for producing proteins are known in the art. In general, leader peptides are 5-30 amino acids long, and are typically present at the N-terminus of a newly synthesized polypeptide. The signal peptide generally contains a long stretch of hydrophobic amino acids that has a tendency to form a single alpha-helix. In addition, many signal peptides begin with a short positively charged

stretch of amino acids, which may help to enforce proper topology of the polypeptide during translocation. At the end of the signal peptide there is typically a stretch of amino acids that is recognized and cleaved by signal peptidase. Signal peptidase may cleave either during or after completion of translocation to generate a free signal peptide and a mature protein. In some embodiments, the signal peptide includes the amino acid sequence having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.9%, or 100% identity to any one of SEQ ID NO: 9, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 22. In some embodiments, the signal peptide includes the amino acid sequence having at least 99% identity to any one of SEQ ID NO: 9, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 22. In some embodiments, the signal peptide has the amino acid sequence selected from any one of SEQ ID NO: 9, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 22. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof described herein may include a cleavable linker. Such linkers allow for the tag to be separated from the purified complex, for example by the addition of an agent capable of cleaving the linker. Cleavable linkers are known in the art. Such linkers may be cleaved for example, by irradiation of a photolabile bond or acid-catalyzed hydrolysis. Another example of a cleavable linker includes a polypeptide linker, which incorporates a protease recognition site and may be cleaved by the addition of a suitable protease enzyme.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes a modification as compared to the corresponding wild-type *E. coli* FimH polypeptide or fragment. The modification may include a covalent attachment of a molecule to the polypeptide. For example, such modifications may include glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may include a modification, such as by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical cleavage, acetylation, formylation, metabolic synthesis of tunicamycin, etc., as compared to a corresponding wild-type *E. coli* FimH polypeptide or fragment. In another embodiment, the modification may include a covalent attachment of a lipid molecule to the polypeptide. In some embodiments, the polypeptide does not include a covalent attachment of a molecule to the polypeptide as compared to the corresponding wild-type *E. coli* FimH polypeptide or fragment thereof.

For example, proteins and polypeptides produced in cell culture may be glycoproteins that contain covalently linked carbohydrate structures including oligosaccharide chains. These oligosaccharide chains are linked to the protein via either N-linkages or O-linkages. The oligosaccharide chains may comprise a significant portion of the mass of the glycoprotein.

- 5 Generally, N-linked oligosaccharide is added to the amino group on the side chain of an asparagine residue within the target consensus sequence of Asn-X-Ser/Thr, where X may be any amino acid except proline. In some embodiments, the glycosylation site includes an amino acid sequence selected from any one of the following: asparagine-glycine-threonine (NGT), asparagine-isoleucine-threonine (NIT), asparagine-glycine-serine (NGS), asparagine-serine-
10 threonine (NST), and asparagine-threonine-serine (NTS). The polypeptide derived from *E. coli* or a fragment thereof produced in mammalian cells may be glycosylated. The glycosylation may occur at the N-linked glycosylation signal Asn-Xaa-Ser/Thr in the sequence of the polypeptide derived from *E. coli* or a fragment thereof. "N-linked glycosylation" refers to the attachment of the carbohydrate moiety via GlcNAc to an asparagine residue in a polypeptide
15 chain. The N-linked carbohydrate contains a common Man 1-6(Man1-3)Man β 1-4GlcNAc β 1-4GlcNAc β -R core structure, where R represents an asparagine residue of the produced polypeptide derived from *E. coli* or a fragment thereof.

- In some embodiments, a glycosylation site in the polypeptide derived from *E. coli* or a fragment thereof is removed by a mutation within the sequence of the polypeptide derived from
20 *E. coli* or a fragment thereof. For example, in some embodiments, the Asn residue of a glycosylation motif (Asn-Xaa-Ser/Thr) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, Thr, and Gln.

- In some embodiments, the Ser residue of a glycosylation motif may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any
25 one of Asp, Thr, and Gln.

In some embodiments, the Thr residue of a glycosylation motif may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, and Gln.

- In some embodiments, a glycosylation site (such as Asn-Xaa-Ser/Thr) in the polypeptide
30 derived from *E. coli* or a fragment thereof is not removed or modified. In some embodiments, a compound to decrease or inhibit glycosylation may be added to the cell culture medium. In such embodiments, the polypeptide or protein includes at least one more unglycosylated (i.e., aglycosylated) site, that is, a completely unoccupied glycan site with no carbohydrate moiety attached thereto, or comprises at least one carbohydrate moiety less at the same potential
35 glycosylation site than an otherwise identical polypeptide or protein which is produced by a cell under otherwise identical conditions but in the absence of a glycosylation inhibiting compound.

Such compounds are known in the art and may include, but are not limited to, tunicamycin, tunicamycin homologs, streptovirudin, mycosporidin, amphomycin, tsushimycin, antibiotic 24010, antibiotic MM 19290, bacitracin, corynetoxin, showdomycin, duimycin, 1-deoxymannonojirimycin, deoxynojirimycin, N-methyl-1-dexoymannojirimycin, brefeldin A, glucose and mannose analogs, 2-deoxy- D-glucose, 2-deoxyglucose, D-(+)-mannose, D-(+) galactose, 2-deoxy-2-fluoro-D-glucose, 1,4-dideoxy- 1,4-imino-D-mannitol (DIM), fluoroglucose, fluoromannose, UDP-2-deoxyglucose, GDP-2-deoxyglucose, hydroxymethylglutaryl-CoA reductase inhibitors, 25-hydroxycholesterol, hydroxycholesterol, swainsonine, cycloheximide, puromycin, actinomycin D, monensin, m-Chlorocarbonyl-cyanide phenylhydrazine (CCCP), compactin, dolichyl-phosphoryl¹-deoxyglucose, N-Acetyl-D-Glucosamine, hygoxanthine, thymidine, cholesterol, glucosamine, mannosamine, castanospermine, glutamine, bromoconduritol, conduritol epoxide and conduritol derivatives, glycosylmethyl-p-nitrophenyltriazines, β -Hydroxynorvaline, threo- β -fluoroasparagine, D-(+)-Gluconic acid δ -lactone, di(2-ethyl hexyl)phosphate, tributyl phosphate, dodecyl phosphate, 2-dimethylamino ethyl ester of (diphenyl methyl)-phosphoric acid, [2-(diphenyl phosphinyloxy)ethyl]trimethyl ammonium iodide, iodoacetate, and/or fluoroacetate One of ordinary skill in the art will readily recognize or will be able to determine glycosylation-inhibiting substances that may be used in accordance with methods and compositions of the present invention without undue experimentation. In such embodiments, glycosylation of the polypeptide or fragment thereof may be controlled without the introduction of an amino acid mutation into the polypeptide or fragment thereof.

In some embodiments, the level of glycosylation (e.g., number of glycan sites that are occupied on the polypeptide or fragment thereof, the size and/or complexity of glycoform at the site, and the like) of the polypeptide or fragment thereof produced by the mammalian cell are lower than levels of glycosylation of the polypeptide or fragment thereof produced under otherwise identical conditions in an otherwise identical medium that lacks such a glycolysis-inhibiting compound and/or mutation.

In some embodiments, the sequence of a polypeptide derived from *E. coli* or a fragment thereof does not include a site of N-linked protein glycosylation. In some embodiments, the sequence of a polypeptide derived from *E. coli* or a fragment thereof does not include at least one site of N-linked protein glycosylation. In some embodiments, the sequence of a polypeptide derived from *E. coli* or a fragment thereof does not include any sites of N-linked protein glycosylation. In some embodiments, the sequence of a polypeptide derived from *E. coli* or a fragment thereof includes a site for N-linked protein glycosylation. In some embodiments, the sequence of a polypeptide

derived from *E. coli* or a fragment thereof includes at most 1 site of N-linked protein glycosylation. In some embodiments, the sequence of a polypeptide derived from *E. coli* or a fragment thereof includes at most 2 sites of N-linked protein glycosylation.

5 A polypeptide derived from *E. coli* or a fragment thereof expressed by different cell lines or in transgenic animals may have different glycan site occupancies, glycoforms and/or glycosylation patterns compared with each other. In some embodiments, the invention encompasses a polypeptide derived from *E. coli* or a fragment thereof regardless of the the glycosylation, glycan occupancy or glycoform pattern of the polypeptide derived from *E. coli* or a fragment thereof produced in a mammalian cell.

10 In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may be derived from an *E. coli* FimH polypeptide, wherein the amino acid residue at position 1 of the polypeptide is phenylalanine, not methionine, for example, a polypeptide having the amino acid sequence SEQ ID NO: 2. Preferably, the polypeptide derived from *E. coli* FimH comprises a phenylalanine at position 1 of the amino acid sequence of the polypeptide derived from *E. coli*.
 15 In another preferred embodiment, the polypeptide derived from *E. coli* FimH comprises the amino acid sequence SEQ ID NO: 3, preferably wherein the residue at position 1 of the amino acid sequence of the polypeptide derived from *E. coli* is phenylalanine. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may include the amino acid sequence SEQ ID NO: 4, which may be derived from an *E. coli* FimH polypeptide.

20 In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the amino acid sequence having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.9%, or 100% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the amino acid sequence having at least 95% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113. In some
 30 embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the amino acid sequence having at least 99% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113. In some
 35 embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the amino acid sequence having at least 99% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ

ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may be derived from an *E. coli* FimG polypeptide, for example, having the amino acid sequence SEQ ID NO: 9. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may be derived from an *E. coli* FimC polypeptide, for example, having the amino acid sequence SEQ ID NO: 10.

A. Polypeptides Derived from *E. coli* FimH and Fragments Thereof

In a preferred embodiment, the polypeptide or fragment thereof is derived from an *E. coli* FimH. In some embodiments, the polypeptide or fragment thereof includes full length *E. coli* FimH. Full length FimH includes two domains: an N-terminal lectin domain and a C-terminal pilin domain, which are connected by a short linker. In some embodiments, the full length of *E. coli* FimH includes 279 amino acids, which includes the full length of the mature protein of *E. coli* FimH. In some embodiments, the full length of *E. coli* FimH includes 300 amino acids, which includes the full length of the mature protein of *E. coli* FimH and a signal peptide sequence having 21 amino acids in length. The primary structure of the 300 amino acid-long wild type FimH is highly conserved across *E. coli* strains.

An exemplary sequence for a full-length *E. coli* FimH is SEQ ID NO: 1. The full length FimH sequence includes a sequence for a lectin domain and a sequence for a pilin domain. The lectin domain of FimH contains the carbohydrate recognition domain, which is responsible for binding to the mannosylated uroplakin 1a on the urothelial cell surface. The pilin domain is anchored to the core of the pilus via a donor strand of the subsequent FimG subunit, which is a process termed donor strand complementation.

Starting from the N-terminus, the names and in parenthesis the exemplary amino acid sequences of each domain of a full length FimH are as follows: FimH lectin (SEQ ID NO: 2) and FimH pilin (SEQ ID NO: 3).

Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include variants that have various degrees of identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20,

SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. Other suitable polypeptides have least 95% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. Other suitable polypeptides have least 99% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. In certain embodiments, the FimH variant proteins: (i) form part of the FimH-FimC; (ii) comprise at least one epitope from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113; and/or (iii) may elicit antibodies in vivo which immunologically cross react with an *E. coli* FimH.

In some embodiments, the composition includes a polypeptide having at least *n* consecutive amino acids from any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113, wherein *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments include an epitope from the sequence. In some embodiments, composition includes a polypeptide having at least 50 consecutive amino acid residues, at least 100 consecutive amino acid residues, at least 125 consecutive amino acid residues, at least 150 consecutive amino acid residues, at least 175 consecutive amino acid residues, at least 200 consecutive amino acid residues, or at least 250 consecutive amino acid residues of the amino acid sequence of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. In some embodiments, composition includes a polypeptide having at least 50 consecutive amino acid residues of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. In some embodiments, composition includes a polypeptide having at least 100 consecutive amino acid residues of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. In some embodiments, composition includes a polypeptide having at least 200 consecutive amino acid residues of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113.

NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113. In some embodiments, composition includes a polypeptide having at least 250 consecutive amino acid residues of the amino acid sequence of any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23,
5 SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113.

In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or
10 99.9% identity to SEQ ID NO: 1. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 1. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 1. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 1. In some embodiments, the composition includes a polypeptide as set
15 forth at SEQ ID NO: 1. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 2. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 2. In
20 some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 2. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 2. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 2. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%,
25 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%,
30 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 4. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 4. In some embodiments, the composition includes a

polypeptide having at least 95% identity to SEQ ID NO: 4. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 4. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 4. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 5. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 5. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 5. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 5. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 5. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 6. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 6. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 6. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 6. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 6. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 20. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 20. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 20. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 20. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 20. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 23. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 23. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 23. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 23. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 23. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%,

- 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide
- 5 having at least 99% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 26.
- 10 In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO:
- 15 26. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 27. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 27. In some embodiments, the
- 20 composition includes a polypeptide having at least 95% identity to SEQ ID NO: 27. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 27. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 27. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%,
- 25 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 28. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 28. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 28. In some embodiments, the composition includes a polypeptide
- 30 having at least 99% identity to SEQ ID NO: 28. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 28. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 29.
- 35 In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 29. In some embodiments, the composition includes a

polypeptide having at least 95% identity to SEQ ID NO: 29. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 29. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 29. In some embodiments, the composition includes a polypeptide having at least 70%,

5 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 110. In some embodiments, the

10 composition includes a polypeptide having at least 99% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 111.

15 In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 111. In some embodiments, the

20 composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 112. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 112. In some embodiments, the composition includes a polypeptide having at least 95%

25 identity to SEQ ID NO: 112. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 112. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 112. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,

30 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 113. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 113. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 113. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 113. In some embodiments, the composition includes a polypeptide as

35 set forth at SEQ ID NO: 113.

Another example of a suitable polypeptide and fragments thereof derived from *E. coli* FimH described herein is shown as SEQ ID NO: 2, which lacks the wild-type N-terminal signal sequence, and corresponds to amino acid residues 22-300 of SEQ ID NO: 1. Another example of a FimH fragment includes the entire N-terminal signal sequence and the mature protein, such as set forth in SEQ ID NO: 1.

In some embodiments, a glycosylation site in the polypeptide derived from *E. coli* or a fragment thereof is removed by a mutation within the sequence of the polypeptide derived from *E. coli* or a fragment thereof. For example, in some embodiments, the Asn residue at position 7 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the Asn residue at position 7 of a lectin domain of an *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 3) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, Thr, and Gln.

In some embodiments, the Thr residue at position 10 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the Thr residue at position 7 of a lectin domain of an *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 3) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, and Gln.

In some embodiments, the Asn residue at position N235 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the Asn residue at position N228 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, Thr, and Gln.

In some embodiments, the Asn residue at position 70 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the Asn residue at position 70 of a lectin domain of an *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 3) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Ser, Asp, Thr, and Gln.

In some embodiments, the Ser residue at position 72 of a mature *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ ID NO: 2) may be mutated, preferably by a substitution. In some embodiments, the Ser residue at position 72 of a lectin domain of an *E. coli* FimH polypeptide (e.g., according to the numbering of SEQ

ID NO: 3) may be mutated, preferably by a substitution. In some embodiments, the residue substitution is selected from any one of Asp, Thr, and Gln.

By the term "fragment" as used herein refers to a polypeptide and is defined as any discrete portion of a given polypeptide that is unique to or characteristic of that polypeptide.

- 5 The term as used herein also refers to any discrete portion of a given polypeptide that retains at least a fraction of the activity of the full-length polypeptide. In certain embodiments, the fraction of activity retained is at least 10% of the activity of the full-length polypeptide. In certain embodiments, the fraction of activity retained is at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the activity of the full-length polypeptide. In certain embodiments, the fraction of
- 10 activity retained is at least 95%, 96%, 97%, 98% or 99% of the activity of the full-length polypeptide. In certain embodiments, the fraction of activity retained is 100% or more of the activity of the full-length polypeptide. In some embodiments, a fragment includes at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more consecutive amino acids of the full-length polypeptide.

15 **B. Complex of FimH, FimC, and fragments thereof**

- In some embodiments, the polypeptide derived from *E. coli* FimH or fragment thereof is present in a complex with polypeptide derived from *E. coli* FimC or fragment thereof. In a preferred embodiment, the polypeptide derived from *E. coli* FimH or fragment thereof and the polypeptide derived from *E. coli* FimC or fragment thereof are present in a complex, preferably
- 20 in a 1:1 ratio in the complex. Without being bound by theory or mechanism, the full length FimH may be stabilized in an active conformation by the periplasmic chaperone FimC, thereby making it possible to purify full-length FimH protein. Accordingly, in some embodiments, the polypeptide or fragment thereof includes full length FimH and full length FimC.

- In some embodiments, the polypeptide or fragment thereof includes a fragment of FimH and a fragment of FimC. In some embodiments, the polypeptide or fragment thereof includes full length FimH and a fragment of FimC. An exemplary sequence for *E. coli* FimC is set forth in SEQ ID NO: 10. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes complex-forming fragments of FimH.

- A complex-forming fragment of FimH may be any part or portion of the FimH protein that
- 30 retain the ability to form a complex with FimC or a fragment thereof. A suitable complex-forming fragment of FimH may also be obtained or determined by standard assays known in the art, such as co-immunoprecipitation assay, cross-linking, or co-localization by fluorescent staining, etc. SDS-PAGE or western blot may also be used (e.g., by showing that the FimH fragment and FimC or fragment thereof are in a complex as evidenced by gel electrophoresis).
- 35 In certain embodiments, the complex-forming fragment of FimH (i) forms part of the FimH-FimC

- complex; (ii) comprises at least one epitope from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, and SEQ ID NO: 111;
- 5 and/or (iii) may elicit antibodies in vivo which immunologically cross react with an *E. coli* FimH.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes full length FimH, wherein the FimH is not complexed with FimC. In further embodiments, the polypeptide or fragment thereof includes a fragment of FimH, wherein

10 the fragment is not complexed with FimC. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof FimC includes SEQ ID NO: 10. In some embodiments, the the complex may be expressed from the same plasmid, preferably under the the control of separate promoters for each polypeptide or fragment thereof.

In some embodiments, the polypeptide derived from *E. coli* FimH or a fragment

15 thereof binds to a polypeptide derived from *E. coli* FimC or a fragment thereof, which may be engineered into the structure of the polypeptide derived from *E. coli* FimH or fragment thereof. The portion of the FimC molecule that binds to the FimH in the complex is called a "donor strand" and the mechanism of formation of the native FimH structure using the strand from FimC that binds to FimH in the FimCH complex is known

20 as "donor strand complementation."

In some embodiments, the polypeptide derived from *E. coli* FimH or a fragment thereof may be expressed by the appropriate donor strand complemented version of FimH, wherein the amino acid sequence of FimC that interacts with FimH in the FimCH complex is itself engineered at the C-terminal end of FimH to provide the native

25 conformation without the need for the remainder of the FimC molecule to be present. In some embodiments, the polypeptide derived from *E. coli* FimH or a fragment thereof may be expressed in the form of a complex that includes isolated domains thereof, such as the lectin binding domain and the piling domain, and such domains may be linked together covalently or non-covalently. For example, in some embodiments, the linking

30 segment may include amino acid sequences or other oligomeric structures, including simple polymer structures.

The methods and compositions of the invention may include complexes described herein, in which said polypeptides or fragments thereof derived from *E. coli* are co-expressed or formed in a combined state.

C. Lectin domain, Pilin domain, and Variants thereof

Conformation and ligand-binding properties of the lectin domain of FimH may be under the allosteric control of the pilin domain of FimH. Under static conditions, the interaction of the two domains of full length FimH stabilizes the lectin domain in a low-affinity to monomannose state (for example, $K_d \sim 300 \mu\text{M}$), which is characterized by a shallow binding pocket. Binding to a mannoside ligand may induce a conformational change leading to a medium affinity state, in which the lectin and pilin domains remain in close contact. However, upon shear stress, the lectin and pilin domains may separate and induce the high-affinity state (for example, $K_d < 1.2 \mu\text{M}$).

Because of the absence of negative allosteric regulation exerted by the pilin domain, isolated lectin domain of FimH is locked in the high-affinity state (for example, $K_d < 1.2 \mu\text{M}$). The isolated, recombinant lectin domain, which is locked in the high-affinity state. Locking the adhesin in a low-affinity conformation (for example, $K_d \sim 300 \mu\text{M}$), however, induces the production of adhesion-inhibiting antibodies. Accordingly, there is an interest in stabilizing the lectin domain in the low-affinity state.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the lectin domain of an *E. coli* FimH. Exemplary sequences for a lectin domain include any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111. In some embodiments, the lectin domain of an *E. coli* FimH includes cysteine substitutions. In a preferred embodiment, the lectin domain of an *E. coli* FimH includes cysteine substitutions within the first 50 amino acid residues of the lectin domain. In some embodiments, the lectin domain may include 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 cysteine substitutions. Preferably, the lectin domain includes 2 cysteine substitutions. See, for example, pSB02158 and pSB02198.

Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH lectin domain variants that have various degrees of identity to SEQ ID NO: 3, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 3. In some embodiments, the composition includes a polypeptide as set

forth at SEQ ID NO: 3. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the pilin domain of an *E. coli* FimH. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH pilin domain variants that have various degrees of identity to SEQ ID NO: 7, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 7. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 7. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 7. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 7. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 7. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 7. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH lectin domain variants that have various degrees of identity to SEQ ID NO: 8, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 8. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 8. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 8. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 8. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 8. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 8. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the pilin domain of an *E. coli* FimH. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH pilin domain variants that have various degrees of identity to SEQ ID NO: 24, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%,

88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 24. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 24. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH lectin domain variants that have various degrees of identity to SEQ ID NO: 26, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 26. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 26. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH lectin domain variants that have various degrees of identity to SEQ ID NO: 110, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 95% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 110. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 110. Other suitable polypeptides and fragments thereof derived from *E. coli* FimH include FimH lectin domain variants that have various degrees of identity to SEQ ID NO: 111, such as at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to the sequence recited in SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%,

77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.9% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 90% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide

5 having at least 95% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 99% identity to SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide as set forth at SEQ ID NO: 111.

10 In some embodiments, the composition includes a polypeptide having at least n consecutive amino acids from any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111, wherein n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments include an epitope from the sequence. In some embodiments, the composition includes a

15 polypeptide having at least 50 consecutive amino acid residues, at least 100 consecutive amino acid residues, at least 125 consecutive amino acid residues, at least 150 consecutive amino acid residues, at least 175 consecutive amino acid residues, at least 200 consecutive amino acid residues, or at least 250 consecutive amino acid residues of the amino acid sequence of any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ

20 ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 50 consecutive amino acid residues of any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 100

25 consecutive amino acid residues of any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111. In some embodiments, the composition includes a polypeptide having at least 150 consecutive amino acid residues of any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111. In

30 some embodiments, the composition includes a polypeptide having at least 250 consecutive amino acid residues of the amino acid sequence of any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 110, and SEQ ID NO: 111.

35 The location and length of a lectin domain of *E. coli* FimH or a homologue or a variant thereof may be predicted based on pairwise alignment of its sequence to any one of SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 24, SEQ ID NO: 26,

SEQ ID NO: 110, and SEQ ID NO: 111, for example by aligning the amino acid sequence of a FimH to SEQ ID NO: 1, and identifying the sequence that aligns to residues 22-179 of SEQ ID NO: 1.

D. Wild-type N-terminal signal sequence

- 5 In some embodiments, the N-terminal wild type signal sequence of full-length FimH is cleaved in a host cell to produce a mature FimH polypeptide. As such, the FimH expressed by the host cell may lack the N-terminal signal sequence. In preferred embodiments, the polypeptide derived from *E. coli* or a fragment thereof may be encoded by a nucleotide sequence that lacks the coding sequence for the wild type N-terminal signal sequence.
- 10 In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof includes the FimH-FimC complex forming fragments of FimH, the N-terminal signal sequence (such as, residues 1-21 of SEQ ID NO: 1), or a combination thereof. A complex-forming fragment of FimH may be any part or portion of the FimH protein that retains the ability to form a complex with FimC.
- 15 In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may lack between 1 and 21 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21 amino acid residues, or lack 1-21 residues, 1-20 residues, 1-15 residues, 1-10 residues, 2-20 residues, 2-15 residues, 2-10 residues, 5-20 residues, 5-15 residues, or 5-10 residues) at the N-terminus and/or C-terminus of the full-length FimH
- 20 polypeptide, which may include the signal sequence, lectin domain, and pilin domain. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 1-21 residues at the N-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 1-10 residues at the N-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a
- 25 fragment thereof lacks 5-15 residues at the N-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 5-10 residues at the N-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 1-21 residues at the C-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or
- 30 a fragment thereof lacks 1-10 residues at the C-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 5-15 residues at the C-terminus of the full-length FimH polypeptide. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof lacks 5-10 residues at the C-terminus of the full-length FimH polypeptide.

II. Nucleic acids

In one aspect, nucleic acids encoding the polypeptide derived from *E. coli* or a fragment thereof are disclosed. One or more nucleic acid constructs encoding the polypeptide derived from *E. coli* or a fragment thereof may be used for genomic integration and subsequent expression of the polypeptide derived from *E. coli* or a fragment thereof. For example, a single nucleic acid construct encoding the polypeptide derived from *E. coli* or fragment thereof may be introduced to a host cell. Alternatively, the coding sequences for the polypeptide derived from *E. coli* or a fragment thereof may be carried by two or more nucleic acid constructs, which are then introduced into host cell simultaneously or sequentially.

For example, in one exemplary embodiment, a single nucleic acid construct encodes the lectin domain and pilin domain of an *E. coli* FimH. In another exemplary embodiment, one nucleic acid construct encodes the lectin domain and a second nucleic acid construct encodes the pilin domain of an *E. coli* FimH. In some embodiments, genomic integration is achieved.

The nucleic acid construct may comprise genomic DNA that comprises one or more introns, or cDNA. Some genes are expressed more efficiently when introns are present. In some embodiments, the nucleic acid sequence is suitable for the expression of exogenous polypeptides in said mammalian cell.

In some embodiments, the nucleic acid encoding the polypeptide or fragment thereof is codon optimized to increase the level of expression in any particular cell.

In some embodiments, the nucleic acid construct includes a signal sequence that encodes a peptide that directs secretion of the polypeptide derived from *E. coli* or a fragment thereof. In some embodiments, the nucleic acid includes the native signal sequence of the polypeptide derived from *E. coli* FimH. In some embodiments where the polypeptide derived from *E. coli* or a fragment thereof includes an endogenous signal sequence, the nucleic acid sequence encoding the signal sequence may be codon optimized to increase the level of expression of the protein in a host cell.

In some embodiments, the signal sequence is any one of the following lengths: 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30 amino acids long. In some embodiments, the signal sequence is 20 amino acids long. In some embodiments, the signal sequence is 21 amino acids long.

In some embodiments, where the polypeptide or fragment thereof includes a signal sequence, the endogenous signal sequence naturally associated with the polypeptide may be replaced with a signal sequence not associated with the wild type polypeptide to improve the level of expression of the polypeptide or fragment thereof in

cultured cells. Accordingly, in some embodiments, the nucleic acid does not include the native signal sequence of the polypeptide derived from *E. coli* or a fragment thereof. In some embodiments, the nucleic acid does not include the native signal sequence of the polypeptide derived from *E. coli* FimH. In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may be expressed with a heterologous peptide, which is preferably a signal sequence or other peptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide derived from *E. coli* or a fragment thereof. For example, the polypeptide derived from *E. coli* FimH or a fragment thereof may be expressed with a heterologous peptide (e.g., IgK signal sequence), which is preferably a signal sequence or other peptide having a specific cleavage site at the N-terminus of the mature *E. coli* FimH protein. In preferred embodiments, the specific cleavage site at the N-terminus of the mature protein *E. coli* FimH occurs immediately before the initial phenylalanine residue of the mature *E. coli* FimH protein. The heterologous sequence selected is preferably one that is recognized and processed (i.e., cleaved by signal peptidase) by the host cell.

In preferred embodiments, the signal sequence is an IgK signal sequence. In some embodiments, the nucleic acid encodes the amino acid sequence SEQ ID NO: 18. In some embodiments, the nucleic acid encodes the amino acid sequence SEQ ID NO: 19. In some embodiments, the nucleic acid encodes the amino acid sequence SEQ ID NO: 22. In preferred embodiments, the signal sequence is a mouse IgK signal sequence.

Suitable mammalian expression vectors for producing the polypeptide derived from *E. coli* or fragments thereof are known in the art and may be commercially available, such as pSecTag2 expression vector from Invitrogen™. An exemplary mouse Ig Kappa signal peptide sequence includes the sequence ETDTLLLWLLLWPGSTG (SEQ ID NO: 54). In some embodiments, the vector includes pBudCE4.1 mammalian expression vector from Thermo Fisher. Additional exemplary and suitable vectors include the pcDNA™3.1 mammalian expression vector (Thermo Fisher).

In some embodiments, the signal sequence does not include a hemagglutinin signal sequence.

In some embodiments, the nucleic acid includes the native signal sequence of the polypeptide derived from *E. coli* or a fragment thereof. In some embodiments, the signal sequence is not an IgK signal sequence. In some embodiments, the signal sequence includes a hemagglutinin signal sequence.

In one aspect, disclosed herein are vectors that include the coding sequences for the polypeptide derived from *E. coli* or a fragment thereof. Exemplary vectors include plasmids that are able to replicate autonomously or to be replicated in a mammalian cell. Typical expression vectors contain suitable promoters, enhancers, and terminators that are useful for regulation of

the expression of the coding sequence(s) in the expression construct. The vectors may also include selection markers to provide a phenotypic trait for selection of transformed host cells (such as conferring resistance to antibiotics such as ampicillin or neomycin).

Suitable promoters are known in the art. Exemplary promoters include, e.g.,
 5 CMV promoter, adenovirus, EF1 a, GAPDH metallothionine promoter, SV-40 early promoter, SV-40 later promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, etc. Promoters may be constitutive or inducible. One or more vectors may be used (e.g., one vector encoding all subunits or domains or fragments thereof, or multiple vectors together encoding the subunits or domains or
 10 fragments thereof).

Internal ribosome entry site (IRES) and 2A peptide sequences may also be used. IRES and 2A peptide provides alternative approaches for co-expression of multiple sequences. IRES is a nucleotide sequence that allows for translation initiation in the middle of a messenger RNA (mRNA) sequence as part of the greater process of protein
 15 synthesis. Usually, in eukaryotes, translation may be initiated only at the 5' end of the mRNA molecule. IRES elements allow expression of multiple genes in one transcript. IRES-based polycistronic vectors, which express multiple proteins from one transcript, may reduce the escape of non-expressing clones from selection. The 2A peptide allows translation of multiple proteins in a single open reading frame into a polyprotein that is
 20 subsequently cleaved into individual proteins through a ribosome-skipping mechanism. 2A peptide may provide more balanced expression of multiple protein products. Exemplary IRES sequences include, e.g., EV71 IRES, EMCV IRES, HCV IRES. For genomic integration, the integration may be site-specific or random. Site-specific recombination may be achieved by introducing homologous sequence(s) into the nucleic
 25 acid constructs described herein. Such homologous sequence substantially matches the endogenous sequence at a specific target site in the host genome. Alternatively, random integration may be used. Sometimes, the expression level of a protein may vary depending upon the integration site. Therefore, it may be desirable to select a number of clones according to recombinant protein expression level to identify a clone that
 30 achieves the desired level of expression.

Exemplary nucleic acid constructs are further described in the figures, such as any one of **FIG. 2A-2T**.

In one aspect, the nucleic acid sequence encodes the amino acid sequence having at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%,
 35 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.9% or 100% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2,

SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113.

5 III. Host Cells

In one aspect, the invention relates to cells in which the sequences encoding the polypeptide derived from *E. coli* or a fragment thereof are expressed in a mammalian host cell. In one embodiment, the polypeptide derived from *E. coli* or a fragment thereof is transiently expressed in the host cell. In another embodiment, the polypeptide derived from *E. coli* or a
 10 fragment thereof is stably integrated into the genome of the host cells, and, when cultured under a suitable condition, express the polypeptide derived from *E. coli* or a fragment thereof. In a preferred embodiment, the polynucleotide sequence is expressed with high efficiency and genomic stability.

Suitable mammalian host cells are known in the art. Preferably, the host cell is suitable
 15 for producing protein at industrial manufacturing scale. Exemplary mammalian host cells include any one of the following and derivatives thereof: Chinese Hamster Ovary (CHO) cells, COS cells (a cell line derived from monkey kidney (African green monkey), Vero cells, Hela cells, baby hamster kidney (BHK) cells, Human Embryonic Kidney (HEK) cells, NSO cells (Murine myeloma cell line), and C127 cells (nontumorigenic mouse cell line). Further
 20 exemplary mammalian host cells include mouse Sertoli (TM4), buffalo rat liver (BRL 3A), mouse mammary tumor (MMT), rat hepatoma (HTC), mouse myeloma (NSO), murine hybridoma (Sp2/0), mouse thymoma (EL4), Chinese Hamster Ovary (CHO) and CHO cell derivatives, murine embryonic (NIH/3T3, 3T3 Li), rat myocardial (H9c2), mouse myoblast (C2C12), and mouse kidney (miMCD-3). Further examples of mammalian cell lines include NSO/1, Sp2/0,
 25 Hep G2, PER.C6, COS-7, TM4, CV1, VERO-76, MDCK, BRL 3A, W138, MMT 060562, TR1, MRC5, and FS4.

Any cell susceptible to cell culture may be utilized in accordance with the present invention. In some embodiments, the cell is a mammalian cell. Non-limiting examples of
 30 mammalian cells that may be used in accordance with the present invention include BALB/c mouse myeloma line (NSO/I, ECACC No: 85110503); human retinoblasts (PER.C6, CruCell, Leiden, The Netherlands); monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59,1977); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells +/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci.
 35 USA, 77:4216, 1980); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251, 1980);

monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1 587); human cervical carcinoma cells (HeLa, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., *Annals N.Y. Acad. Sci.*, 383:44-68, 1982); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2). In some preferred embodiment, the cells are CHO cells. In some preferred embodiments, the cells are GS-cells.

Additionally, any number of commercially and non-commercially available hybridoma cell lines may be utilized in accordance with the present invention. The term “hybridoma” as used herein refers to a cell or progeny of a cell resulting from fusion of an immortalized cell and an antibody-producing cell. Such a resulting hybridoma is an immortalized cell that produces antibodies. Individual cells used to create the hybridoma can be from any mammalian source, including, but not limited to, rat, pig, rabbit, sheep, pig, goat, and human. In some embodiments, a hybridoma is a trioma cell line, which results when progeny of heterohybrid myeloma fusions, which are the product of a fusion between human cells and a murine myeloma cell line, are subsequently fused with a plasma cell. In some embodiments, a hybridoma is any immortalized hybrid cell line that produces antibodies such as, for example, quadromas (See, e.g., Milstein et al., *Nature*, 537:3053, 1983). One skilled in the art will appreciate that hybridoma cell lines might have different nutrition requirements and/or might require different culture conditions for optimal growth, and will be able to modify conditions as needed.

In some embodiments, the cell comprises a first gene of interest, wherein the first gene of interest is chromosomally-integrated. In some embodiments, the first gene of interest comprises a reporter gene, a selection gene, a gene of interest (e.g., encoding a polypeptide derived from *E. coli* or a fragment thereof), an ancillary gene, or a combination thereof. In some embodiments, the gene of therapeutic interest comprises a gene encoding a difficult to express (DtE) protein.

In some embodiments, the first gene of interest is located between two of the distinct recombination target sites (RTS) in a site-specific integration (SSI) mammalian cell, wherein two RTS are chromosomally-integrated within the NL1 locus or the NL2 locus. See, for example, United States Patent Application Publication No. 20200002727, for a description of the NL1 locus, the NL2 locus, the NL3 locus, the NL4 locus, the NL5 locus, and the NL6 locus. In some embodiments, the first gene of interest is located within the NL1 locus. In some embodiments, the cell comprises a second gene of

interest, wherein the second gene of interest is chromosomally-integrated. In some embodiments, the second gene of interest comprises a reporter gene, a selection gene, a gene of therapeutic interest (such as a polypeptide derived from *E. coli* or a fragment thereof), an ancillary gene, or a combination thereof. In some embodiments, the gene of therapeutic interest comprises a gene encoding a DtE protein. In some embodiments, the second gene of interest is located between two of the RTS. In some embodiments, the second gene of interest is located within the NL1 locus or the NL2 locus. In some embodiments, the first gene of interest is located within the NL1 locus, and the second gene of interest is located within the NL2 locus. In some embodiments, the cell comprises a third gene of interest, wherein the third gene of interest is chromosomally-integrated. In some embodiments, the third gene of interest comprises a reporter gene, a selection gene, a gene of therapeutic interest (such as a polypeptide derived from *E. coli* or a fragment thereof), an ancillary gene, or a combination thereof. In some embodiments, the gene of therapeutic interest comprises a gene encoding a DtE protein. In some embodiments, the third gene of interest is located between two of the RTS. In some embodiments, the third gene of interest is located within the NL1 locus or the NL2 locus. In some embodiments, the third gene of interest is located within a locus distinct from the NL1 locus and the NL2 locus. In some embodiments, the first gene of interest, the second gene of interest, and the third gene of interest are within three separate loci. In some embodiments, at least one of the first genes of interest, the second gene of interest, and the third gene of interest is within the NL1 locus, and at least one of the first gene of interest, the second gene of interest, and the third gene of interest is within the NL2 locus. In some embodiments, the cell comprises a site-specific recombinase gene. In some embodiments, the site-specific recombinase gene is chromosomally-integrated.

In some embodiments, the present disclosure provides a mammalian cell comprising at least four distinct RTS, wherein the cell comprises (a) at least two distinct RTS are chromosomally-integrated within the NL1 locus or NL2 locus; (b) a first gene of interest is integrated between the at least two RTS of (a), wherein the first gene of interest comprises a reporter gene, a gene encoding a DtE protein, an ancillary gene or a combination thereof; (c) and a second gene of interest is integrated within a second chromosomal locus distinct from the locus of (a), wherein the second gene of interest comprises a reporter gene, a gene encoding a DtE protein (such as a polypeptide derived from *E. coli* or a fragment thereof), an ancillary gene or a combination thereof. In some embodiments, the present disclosure provides a mammalian cell comprising at least four distinct RTS, wherein the cell comprises (a) at least two distinct RTS are chromosomally-integrated within the Fer1L4 locus; (b) at least two distinct RTS are chromosomally-integrated within the NL1 locus or the NL2 locus; (c) a first gene of interest is chromosomally-integrated within the Fer1L4 locus, wherein the first gene of interest comprises

a reporter gene, a gene encoding a DtE protein, an ancillary gene or a combination thereof; and (d) a second gene of interest is chromosomally-integrated within the within the NL1 locus or NL2 locus of (b), wherein the second gene of interest comprises a reporter gene, a gene encoding a DtE protein (such as a polypeptide derived from *E. coli* or a fragment thereof), an ancillary gene or a combination thereof.

In some embodiments, the present disclosure provides a mammalian cell comprising at least six distinct RTS, wherein the cell comprises (a) at least two distinct RTS and a first gene of interest are chromosomally-integrated within the *Fer1L4* locus; (b) at least two distinct RTS and a second gene of interest are chromosomally-integrated within the NL1 locus; and (c) at least two distinct RTS and a third gene of interest are chromosomally-integrated within the NL2 locus.

As referred to herein, the terms “in operable combination,” “in operable order,” and “operably linked” refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced. In some embodiments, a gene of interest is operably linked to a promoter, wherein the gene of interest is chromosomally-integrated into the host cell. In some embodiments, the gene of interest is operably linked to a heterologous promoter; where in the gene of interest is chromosomally-integrated into the host cell. In some embodiments, an ancillary gene is operably linked to a promoter, wherein the ancillary gene is chromosomally-integrated into the host cell genome. In some embodiments, the ancillary gene is operably linked to a heterologous promoter; where in the ancillary gene is chromosomally-integrated into the host cell genome. In some embodiments, a gene encoding a DtE protein is operably linked to a promoter, wherein the gene encoding a DtE protein is chromosomally-integrated into the host cell genome. In some embodiments, the gene encoding a DtE protein is operably linked to a heterologous promoter, where in the gene encoding a DtE protein is chromosomally-integrated into the host cell genome. In some embodiments, a recombinase gene is operably linked to a promoter, wherein the recombinase gene is chromosomally-integrated into the host cell. In some embodiments, the recombinase gene is operably linked to a promoter, where in the recombinase gene is not integrated into the host cell genome. In some embodiments, a recombinase gene is operably linked to a heterologous promoter, wherein the recombinase gene is not chromosomally-integrated into the host cell genome. In some embodiments, the recombinase gene is operably linked to a

heterologous promoter, wherein the recombinase gene is not chromosomally-integrated into the host cell genome.

As referred to herein, the term "chromosomally-integrated" or "chromosomal integration" refers to the stable incorporation of a nucleic acid sequence into the chromosome of a host cell, e.g. a mammalian cell. i.e., a nucleic acid sequence that is chromosomally-integrated into the genomic DNA (gDNA) of a host cell, e.g. a mammalian cell. In some embodiments, a nucleic acid sequence that is chromosomally-integrated is stable. In some embodiments, a nucleic acid sequence that is chromosomally-integrated is not located on a plasmid or a vector. In some embodiments, a nucleic acid sequence that is chromosomally-integrated is not excised. In some embodiments, chromosomal integration is mediated by the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein (Cas) gene editing system (CRISPR/CAS).

In some embodiments, the host cells are suitable for growth in suspension cultures. Suspension competent host cells are generally monodisperse or grow in loose aggregates without substantial aggregation. Suspension competent host cells include cells that are suitable for suspension culture without adaptation or manipulation (e.g., hematopoietic cells, lymphoid cells) and cells that have been made suspension competent by modification or adaptation of attachment-dependent cells (e.g., epithelial cells, fibroblasts).

In some embodiments, the expression level or activity of the polypeptide derived from *E. coli* or fragment thereof is increased by at least 2-fold, at least 3 fold, at least 5 fold, at least 10 fold, at least 20 fold, at least 30 fold, at least 40 fold, at least 50 fold, at least 60 fold, at least 70 fold, at least 75 fold, at least 80 fold, at least 90 fold, at least 100 fold, as compared to expression of the polypeptide derived from *E. coli* or a fragment thereof in a bacterial cell, such as, for example, an *E. coli* host cell. In some embodiments, the expression level of the polypeptide derived from *E. coli* or fragment thereof is increased by at least 2-fold as compared to expression of the polypeptide derived from *E. coli* or a fragment thereof in an *E. coli* host cell. In some embodiments, the expression level of the polypeptide derived from *E. coli* or fragment thereof is increased by at least 50-fold as compared to expression of the polypeptide derived from *E. coli* or a fragment thereof in an *E. coli* host cell. In some embodiments, the expression level of the polypeptide derived from *E. coli* or fragment thereof is increased by at least 100-fold as compared to expression of the polypeptide derived from *E. coli* or a fragment thereof in an *E. coli* host cell.

The host cells described herein are suitable for large scale culture. For example, the cell cultures may be 10 L, 30 L, 50 L, 100 L, 150 L, 200 L, 300 L, 500 L, 1000 L, 2000 L, 3000 L, 4000 L, 5000 L, 10,000 L or larger. In some embodiments, the cell culture size may range from 10 L to 5000 L, from 10 L to 10,000 L, from 10 L, to 20,000 L, from 10 L, to 50,000 L, from

40 l, to 50,000 L, from 100 L to 50,000 L, from 500 L to 50,000 L, from 1000 L to 50,000 L, from 2000 L to 50,000 L, from 3000 L to 50,000 L, from 4000 L to 50,000 L, from 4500 L to 50,000 L, from 1000 L to 10,000 L, from 1000 L to 20,000 L, from 1000 L to 25,000 L, from 1000 L to 30,000 L, from 15 L to 2000 L, from 40 L to 1000 L, from 100 L to 500 L, from 200 L to 400 L, or any integer there between. Media components for cell culture are known in the art, and may include, e.g., buffer, amino acid content, vitamin content, salt content, mineral content, serum content, carbon source content, lipid content, nucleic acid content, hormone content, trace element content, ammonia content, co-factor content, indicator content, small molecule content, hydrolysate content and enzyme modulator content.

The terms "medium", "cell culture medium" and "culture medium" as used herein refer to a solution containing nutrients which nourish growing mammalian cells. Typically, such solutions provide essential and non-essential amino acids, vitamins, energy sources, lipids, and trace elements required by the cell for minimal growth and/or survival. Such a solution may also contain supplementary components that enhance growth and/or survival above the minimal rate, including, but not limited to, hormones and/or other growth factors, particular ions (such as sodium, chloride, calcium, magnesium, and phosphate), buffers, vitamins, nucleosides or nucleotides, trace elements (inorganic compounds usually present at very low final concentrations), inorganic compounds present at high final concentrations (e.g., iron), amino acids, lipids, and/or glucose or other energy source. In some embodiments, a medium is advantageously formulated to a pH and salt concentration optimal for cell survival and proliferation. In some embodiments, a medium is a feed medium that is added after the beginning of the cell culture.

In some embodiments, cells may be grown in one of a variety of chemically defined media, wherein the components of the media are both known and controlled. In some embodiments, cells may be grown in a complex medium, in which not all components of the medium are known and/or controlled. Chemically defined growth media for mammalian cell culture have been extensively developed and published over the last several decades. All components of defined media are well characterized, and so defined media do not contain complex additives such as serum or hydrolysates. Early media formulations were developed to permit cell growth and maintenance of viability with little or no concern for protein production. More recently, media formulations have been developed with the express purpose of supporting highly productive recombinant protein producing cell cultures. Such media are preferred for use in the method of the invention. Such media generally comprises high amounts of

nutrients and in particular of amino acids to support the growth and/or the maintenance of cells at high density. If necessary, these media can be modified by the skilled person for use in the method of the invention. For example, the skilled person may decrease the amount of phenylalanine, tyrosine, tryptophan and/or methionine in these media for their use as base media or feed media in a method as disclosed herein.

Not all components of complex media are well characterized, and so complex media may contain additives such as simple and/or complex carbon sources, simple and/or complex nitrogen sources, and serum, among other things. In some embodiments, complex media suitable for the present invention contains additives such as hydrolysates in addition to other components of defined medium as described herein. In some embodiments, defined media typically includes roughly fifty chemical entities at known concentrations in water. Most of them also contain one or more well-characterized proteins such as insulin, IGF-1, transferrin or BSA, but others require no protein components and so are referred to as protein-free defined media. Typical chemical components of the media fall into five broad categories: amino acids, vitamins, inorganic salts, trace elements, and a miscellaneous category that defies neat categorization.

Cell culture medium may be optionally supplemented with supplementary components. The term "supplementary components" as used herein refers to components that enhance growth and/or survival above the minimal rate, including, but not limited to, hormones and/or other growth factors, particular ions (such as sodium, chloride, calcium, magnesium, and phosphate), buffers, vitamins, nucleosides or nucleotides, trace elements (inorganic compounds usually present at very low final concentrations), amino acids, lipids, and/or glucose or other energy source. In some embodiments, supplementary components may be added to the initial cell culture. In some embodiments, supplementary components may be added after the beginning of the cell culture. Typically, trace elements refer to a variety of inorganic salts included at micromolar or lower levels. For example, commonly included trace elements are zinc, selenium, copper, and others. In some embodiments, iron (ferrous or ferric salts) can be included as a trace element in the initial cell culture medium at micromolar concentrations. Manganese is also frequently included among the trace elements as a divalent cation (MnCl_2 or MnSO_4) in a range of nanomolar to micromolar concentrations. Numerous less common trace elements are usually added at nanomolar concentrations.

In some embodiments, the medium used in the method of the invention is a medium suitable for supporting high cell density, such as for example 1×10^6 cells/mL, 5×10^6 cells/mL, 1×10^7 cells/mL, 5×10^7 cells/mL, 1×10^8 cells/mL or 5×10^8 cells/mL, in a cell culture. In some embodiments, the cell culture is a mammalian cell fed-batch culture, preferably a CHO cells fed-batch culture.

In some embodiments, the cell culture medium comprises phenylalanine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises tyrosine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises tryptophan at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises leucine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises serine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises threonine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises two of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises phenylalanine and tyrosine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises phenylalanine and tryptophan at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises phenylalanine and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises tyrosine and tryptophan at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises tyrosine and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments,

the cell culture medium comprises tryptophan and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises three of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration

5 below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises phenylalanine, tyrosine and tryptophan at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

10 embodiments, the cell culture medium comprises phenylalanine, tyrosine and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises phenylalanine, tryptophan and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

15 embodiments, the cell culture medium comprises tyrosine, tryptophan and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises four of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

20 embodiments, the cell culture medium comprises phenylalanine, tyrosine, tryptophan and methionine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises five of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

25 embodiments, the cell culture medium comprises six of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some embodiments, the cell culture medium comprises seven of phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

30 embodiments, the cell culture medium comprises phenylalanine, tyrosine, tryptophan, methionine, leucine, serine, threonine and glycine at a concentration below 2mM, below 1mM, between 0.1 and 2mM, between 0.1 to 1mM, between 0.5 and 1.5mM or between 0.5 to 1mM. In some

35 In some embodiments, the cell culture medium further comprises at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 of glycine, valine, leucine, isoleucine, proline, serine, threonine, lysine,

arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium further comprises at least 5 of glycine, valine, leucine, isoleucine, proline, serine, threonine, lysine, arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium further comprises glycine, valine, leucine, isoleucine, proline, serine, threonine, lysine, arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium further comprises at least 1, 2, 3, 4, 5, 6, 7, 8, or 9 of valine, isoleucine, proline, lysine, arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium further comprises at least 5 of valine, isoleucine, proline, lysine, arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium further comprises valine, isoleucine, proline, lysine, arginine, histidine, aspartate, glutamate and asparagine at a concentration above 2mM, 3mM, 4mM, 5mM, 10mM, 15mM, preferably 2mM. In some embodiments, the cell culture medium comprises serine at a concentration above 3mM, 5mM, 7mM, 10mM, 15mM or 20mM, preferably 10mM. In some embodiments, the cell culture medium comprises valine at a concentration above 3mM, 5mM, 7mM, 10mM, 15mM or 20mM, preferably 10mM. In some embodiments, the cell culture medium comprises cysteine at a concentration above 3mM, 5mM, 7mM, 10mM, 15mM or 20mM, preferably 10mM. In some embodiments, the cell culture medium comprises isoleucine at a concentration above 3mM, 5mM, 7mM, 10mM, 15mM or 20mM, preferably 10mM. In some embodiments, the cell culture medium comprises leucine at a concentration above 3mM, 5mM, 7mM, 10mM, 15mM or 20mM, preferably 10mM. In some embodiments, the above cell culture medium is for use in a method as disclosed herein. In some embodiments, the above cell culture medium is used in a method as disclosed herein as a base media. In some embodiments, the above cell culture medium is used a method as disclosed herein as a feed media.

IV. Method of Producing

In one aspect, the invention includes a method of producing a polypeptide derived from *E. coli* or a fragment thereof. The method includes culturing a mammalian cell under a suitable condition, thereby expressing the polypeptide derived from *E. coli*

or a fragment thereof. The method may further include harvesting the polypeptide derived from *E. coli* or a fragment thereof from the culture. The process may further include purifying the polypeptide derived from *E. coli* or a fragment thereof.

5 In some embodiments, the method produces the polypeptide or fragment thereof at a yield as 0.1 g/L to 0.5 g/L.

In some embodiments, the cells may be grown in batch or fed-batch cultures, where the culture is terminated after sufficient expression of the polypeptide, after which the expressed polypeptide is harvested and optionally purified. In some embodiments, the cells may be grown in perfusion cultures, where the culture is not terminated and new nutrients and other
10 components are periodically or continuously added to the culture, during which the expressed polypeptide is periodically or continuously harvested.

In some embodiments, the cells may be grown in small scale reaction vessels ranging in volume from a few milliliters to several liters. In some embodiments, the cells may be grown in large scale commercial bioreactors ranging in volume from approximately least 1 liter to 10,
15 100, 250, 500, 1,000, 2,500, 5,000, 8,000, 10,000, 12,000 liters or more, or any volume in between.

The temperature of the cell culture will be selected based primarily on the range of temperatures at which the cell culture remains viable, at which a high level of polypeptide is produced, the temperature at which production or accumulation of metabolic waste products is
20 minimized, and/or any combination of these or other factors deemed important by the practitioner. As one non-limiting example, CHO cells grow well and produce high levels of protein or polypeptide at approximately 37°C. In general, most mammalian cells grow well and/or can produce high levels of protein or polypeptide within a range of about 25°C to 42°C, although methods taught by the present disclosure are not limited to these temperatures.
25 Certain mammalian cells grow well and/or can produce high levels of protein or polypeptide within the range of about 35°C to 40°C. In certain embodiments, the cell culture is grown at a temperature of 20°C, 21°C, 22°C, 23°C, 24°C, 25°C, 26°C, 27°C, 28°C, 29°C, 30°C, 31°C, 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, 38°C, 39°C, 40°C, 41°C, 42°C, 43°C, 44°C, or 45°C at one or more times during the cell culture process.

30 The terms "culture" and "cell culture" as used herein refer to a cell population that is suspended in a medium under conditions suitable to survival and/or growth of the cell population. As will be clear to those of ordinary skill in the art, in some embodiments, these terms as used herein refer to the combination comprising the cell population and the medium in which the population is suspended. In some embodiments, the cells of the cell culture comprise
35 mammalian cells.

The present invention may be used with any cell culture method that is amenable to the desired process (e.g., production of a recombinant protein (e.g., antibody)). As a non-limiting example, cells may be grown in batch or fed-batch cultures, where the culture is terminated after sufficient expression of the recombinant protein (e.g., antibody), after which the expressed protein (e.g., antibody) is harvested. Alternatively, as another non-limiting example, cells may be grown in batch-refeed, where the culture is not terminated and new nutrients and other components are periodically or continuously added to the culture, during which the expressed recombinant protein (e.g., antibody) is harvested periodically or continuously. Other suitable methods (e.g., spin-tube cultures) are known in the art and can be used to practice the present invention.

In some embodiments, a cell culture suitable for the present invention is a fed-batch culture. The term "fed-batch culture" as used herein refers to a method of culturing cells in which additional components are provided to the culture at a time or times subsequent to the beginning of the culture process. Such provided components typically comprise nutritional components for the cells which have been depleted during the culturing process. A fed-batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified. In some embodiments, the fed-batch culture comprises a base medium supplemented with feed media.

Cells may be grown in any convenient volume chosen by the practitioner. For example, cells may be grown in small scale reaction vessels ranging in volume from a few milliliters to several liters. Alternatively, cells may be grown in large scale commercial Bioreactors ranging in volume from approximately at least 1 liter to 10, 50, 100, 250, 500, 1000, 2500, 5000, 8000, 10,000, 12,000, 15000, 20000 or 25000 liters or more, or any volume in between.

The temperature of a cell culture will be selected based primarily on the range of temperatures at which the cell culture remains viable and the range in which a high level of desired product (e.g., a recombinant protein) is produced. In general, most mammalian cells grow well and can produce desired products (e.g., recombinant proteins) within a range of about 25°C to 42°C, although methods taught by the present disclosure are not limited to these temperatures. Certain mammalian cells grow well and can produce desired products (e.g., recombinant proteins or antibodies) within the range of about 35°C to 40°C. In certain embodiments, a cell culture is grown at a temperature of 20°C, 21°C, 22°C, 23°C, 24°C, 25°C, 26°C, 27°C, 28°C, 29°C, 30°C, 31°C, 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, 38°C, 39°C, 40°C, 41°C, 42°C, 43°C, 44°C, or 45°C at one or more times during the cell culture process. Those of ordinary skill in the art will

be able to select appropriate temperature or temperatures in which to grow cells, depending on the particular needs of the cells and the particular production requirements of the practitioner. The cells may be grown for any amount of time, depending on the needs of the practitioner and the requirement of the cells themselves. In some embodiment, the cells are grown at 37°C. In some embodiments, the cells are grown at 36.5°C.

In some embodiments, the cells may be grown during the initial growth phase (or growth phase) for a greater or lesser amount of time, depending on the needs of the practitioner and the requirement of the cells themselves. In some embodiments, the cells are grown for a period of time sufficient to achieve a predefined cell density. In some embodiments, the cells are grown for a period of time sufficient to achieve a cell density that is a given percentage of the maximal cell density that the cells would eventually reach if allowed to grow undisturbed. For example, the cells may be grown for a period of time sufficient to achieve a desired viable cell density of 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 99 percent of maximal cell density. In some embodiments, the cells are grown until the cell density does not increase by more than 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% per day of culture. In some embodiments, the cells are grown until the cell density does not increase by more than 5% per day of culture.

In some embodiment the cells are allowed to grow for a defined period of time. For example, depending on the starting concentration of the cell culture, the temperature at which the cells are grown, and the intrinsic growth rate of the cells, the cells may be grown for 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more days, preferably for 4 to 10 days. In some cases, the cells may be allowed to grow for a month or more. The practitioner of the present invention will be able to choose the duration of the initial growth phase depending on protein production requirements and the needs of the cells themselves.

The cell culture may be agitated or shaken during the initial culture phase in order to increase oxygenation and dispersion of nutrients to the cells. In accordance with the present invention, one of ordinary skill in the art will understand that it can be beneficial to control or regulate certain internal conditions of the bioreactor during the initial growth phase, including but not limited to pH, temperature, oxygenation, etc.

At the end of the initial growth phase, at least one of the culture conditions may be shifted so that a second set of culture conditions is applied and a metabolic shift occurs in the culture. A metabolic shift can be accomplished by, e.g., a change in the temperature, pH, osmolality or chemical inductant level of the cell culture. In one non-limiting embodiment, the culture conditions are shifted by shifting the temperature of the culture. However, as is known in the art, shifting temperature is not the only mechanism through which an appropriate metabolic shift can be achieved. For example, such a metabolic shift can also be achieved by

shifting other culture conditions including, but not limited to, pH, osmolality, and sodium butyrate levels. The timing of the culture shift will be determined by the practitioner of the present invention, based on protein production requirements or the needs of the cells themselves.

5 When shifting the temperature of the culture, the temperature shift may be relatively gradual. For example, it may take several hours or days to complete the temperature change. Alternatively, the temperature shift may be relatively abrupt. For example, the temperature change may be complete in less than several hours. Given the appropriate production and control equipment, such as is standard in the commercial
10 large-scale production of polypeptides or proteins, the temperature change may even be complete within less than an hour.

 In some embodiments, once the conditions of the cell culture have been shifted as discussed above, the cell culture is maintained for a subsequent production phase under a second set of culture conditions conducive to the survival and viability of the cell
15 culture and appropriate for expression of the desired polypeptide or protein at commercially adequate levels.

 As discussed above, the culture may be shifted by shifting one or more of a number of culture conditions including, but not limited to, temperature, pH, osmolality, and sodium butyrate levels. In some embodiments, the temperature of the culture is
20 shifted. According to this embodiment, during the subsequent production phase, the culture is maintained at a temperature or temperature range that is lower than the temperature or temperature range of the initial growth phase. As discussed above, multiple discrete temperature shifts may be employed to increase cell density or viability or to increase expression of the recombinant protein.

25 In some embodiments, the cells may be maintained in the subsequent production phase until a desired cell density or production titer is reached. In another embodiment of the present invention, the cells are allowed to grow for a defined period of time during the subsequent production phase. For example, depending on the concentration of the cell culture at the start of the subsequent growth phase, the
30 temperature at which the cells are grown, and the intrinsic growth rate of the cells, the cells may be grown for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more days. In some cases, the cells may be allowed to grow for a month or more. The practitioner of the present invention will be able to choose the duration of the subsequent production phase depending on polypeptide or protein production
35 requirements and the needs of the cells themselves.

The cell culture may be agitated or shaken during the subsequent production phase in order to increase oxygenation and dispersion of nutrients to the cells. In accordance with the present invention, one of ordinary skill in the art will understand that it can be beneficial to control or regulate certain internal conditions of the bioreactor during the subsequent growth phase, including but not limited to pH, temperature, oxygenation, etc.

In some embodiments, the cells express a recombinant protein and the cell culture method of the invention comprises a growth phase and a production phase.

In some embodiments, step (ii) of any of the methods disclosed herein is applied during the totality of the cell culture method. In some embodiments, step (ii) of any of the methods disclosed herein is applied during a part of the cell culture method. In some embodiments, step (ii) is applied until a predetermined viable cell density is obtained.

In some embodiments, the cell culture method of the invention comprises a growth phase and a production phase and step (ii) is applied during the growth phase. In some embodiments, the cell culture method of the invention comprises a growth phase and a production phase and step (ii) is applied during a part of the growth phase. In some embodiments, the cell culture method of the invention comprises a growth phase and a production phase and step (ii) is applied during the growth phase and the production phase.

In step (ii) of any of the methods disclosed herein, the term "maintaining" can refer to maintaining the concentration of amino acid or metabolite below C1 or C2 for the entire culture process (until harvesting) or for a part of the culture process such as for example the growth phase, a part of the growth phase or until a predetermined cell density is obtained.

In some embodiments of any of the above mentioned methods, cell growth and/or productivity are increased as compared to a control culture, said control culture being identical except that it does not comprise step (ii).

In some embodiments of any of the above mentioned methods, the method of the invention is a method for improving cell growth. In some embodiment, the method of the invention is a method for improving cell growth in high density cell culture at high cell density.

High cell density as used herein refers to cell density above 1×10^6 cells/mL, 5×10^6 cells/mL, 1×10^7 cells/mL, 5×10^7 cells/mL, 1×10^8 cells/mL or 5×10^8 cells/mL, preferably above 1×10^7 cells/mL, more preferably above 5×10^7 cells/mL.

In some embodiments, the method of the invention is a method for improving cell growth in a cell culture where cell density is above 1×10^6 cells/mL, 5×10^6 cells/mL, 1×10^7 cells/mL, 5×10^7 cells/mL, 1×10^8 cells/mL or 5×10^8 cells/mL. In some embodiments, the method of the invention is a method for improving cell growth in a cell culture where maximum cell density is above 1×10^6 cells/mL, 5×10^6 cells/mL, 1×10^7 cells/mL, 5×10^7 cells/mL, 1×10^8 cells/mL or 5×10^8 cells/mL.

In some embodiments, cell growth is determined by viable cell density (VCD), maximum viable cell density, or Integrated viable cell count (IVCC). In some embodiments, cell growth is determined by maximum viable cell density.

5 The term “viable cell density” as used herein refers to the number of cells present in a given volume of medium. Viable cell density can be measured by any method known to the skilled person. Preferably, Viable cell density is measured using an automated cell counter such as Bioprofile Flex®. The term maximum cell density as used herein refers to the maximum cell density achieved during the cell culture. The term “cell viability” as used herein refers to the ability of cells in culture to survive under
10 a given set of culture conditions or experimental variations. Those of ordinary skill in the art will appreciate that one of many methods for determining cell viability are encompassed in this invention. For example, one may use a dye (e.g., trypan blue) that does not pass through the membrane of a living cell, but can pass through the disrupted membrane of a dead or dying cell in order to determine cell viability.

15 The term “Integrated viable cell count (IVCC)” as used herein refers to as the area under the viable cell density (VCD) curve. IVCC can be calculated using the following formula: $IVCC_{t+1} = IVCC_t + (VCD_t + VCD_{t+1}) * (\Delta t) / 2$, where Δt is the time difference between t and t+1 time points. $IVCC_{t=0}$ can be assumed negligible. VCD_t and VCD_{t+1} are viable cell densities at t and t+1 time points.

20 The term “titer” as used herein refers, for example, to the total amount of recombinantly expressed protein produced by a cell culture in a given amount of medium volume. Titer is typically expressed in units of grams of protein per liter of medium.

In some embodiments, cell growth is increased by at least 5%, 10%, 15%, 20%
25 or 25% as compared to the control culture. In some embodiments, cell growth is increased by at least 10% as compared to the control culture. In some embodiments, cell growth is increased by at least 20% as compared to the control culture.

In some embodiments, the productivity is determined by titer and/or volumetric productivity.

30 The term “titer” as used herein refers, for example, to the total amount of recombinantly expressed protein produced by a cell culture in a given amount of medium volume. Titer is typically expressed in units of grams of protein per liter of medium.

In some embodiments, the productivity is determined by titer. In some
35 embodiments, the productivity is increased by at least 5%, 10%, 15%, 20% or 25% as compared to the control culture. In some embodiments, the productivity is increased by

at least 10% as compared to a control culture. In some embodiments, the productivity is increased by at least 20% as compared to a control culture.

In some embodiments, the maximum cell density of the cell culture is greater than 1×10^6 cells/mL, 5×10^6 cells/mL, 1×10^7 cells/mL, 5×10^7 cells/mL, 1×10^8 cells/mL or 5×10^8 cells/mL. In some embodiments, the maximum cell density of the cell culture is greater than 5×10^6 cells/mL. In some embodiments, the maximum cell density of the cell culture is greater than 1×10^8 cells/mL.

V. Purification

In some embodiments, the method for producing a polypeptide derived from *E. coli* or a fragment thereof includes isolating and/or purifying the polypeptide derived from *E. coli* or a fragment thereof. In some embodiments, the expressed polypeptide derived from *E. coli* or a fragment thereof is secreted into the medium and thus cells and other solids may be removed by centrifugation and/or filtration.

The polypeptide derived from *E. coli* or a fragment thereof produced in accordance with the methods described herein may be harvested from host cells and purified using any suitable method known to the skilled person. Suitable methods for purifying the polypeptide or fragment thereof include precipitation and various types of chromatography, such as hydrophobic interaction, ion exchange, affinity, chelation, and size exclusion, all of which are known in the art. Suitable purification schemes may include two or more of these or other suitable methods. In some embodiments, one or more of the polypeptide or fragments thereof derived from *E. coli* may include a "tag" that facilitates purification, such as an epitope tag or a HIS tag, Strep tag. Such tagged polypeptides may conveniently be purified, for example from conditioned media, by chelating chromatography or affinity chromatography. Optionally, the tag sequence may be cleaved post-purification.

In some embodiments, the polypeptide derived from *E. coli* or a fragment thereof may include a tag for affinity purification. Affinity purification tags are known in the art. Examples include, e.g., His tag (binds to metal ion), an antibody, maltose-binding protein (MBP) (binds to amylose), glutathione-S- transferase (GST) (binds to glutathione), FLAG tag, Strep tag (binds to streptavidin or a derivative thereof).

In a preferred embodiment, the polypeptide derived from *E. coli* or a fragment thereof does not include a purification tag.

In some embodiments, the yield of the polypeptide derived from *E. coli* or a fragment thereof is at least about 1 mg/L, at least about 2 mg/L, at least about 3 mg/L, at least about 4 mg/L, at least about 5 mg/L, at least about 6 mg/L, at least about 7 mg/L, at least about 8 mg/L, at least about 9 mg/L, at least about 10 mg/L, at least about 11 mg/L, at least about 12

mg/L, at least about 13 mg/L, at least about 14 mg/L, at least about 15 mg/L, at least about 16 mg/L, at least about 17 mg/L, at least about 18 mg/L, at least about 19 mg/L, at least about 20 mg/L, at least about 25 mg/L, at least about 30 mg/L, at least about 35 mg/L, at least about 40 mg/L, at least about 45 mg/L, at least about 50 mg/L, at least about 55 mg/L, at least about 60 mg/L, at least about 65 mg/L, at least about 70 mg/L, at least about 75 mg/L, at least about 80 mg/L, at least about 85 mg/L, at least about 90 mg/L, at least about 95 mg/L, or at least about 100 mg/L.

In some embodiments, the culture is at least about 10 liters in size, e.g., a volume of at least about 10L, at least about 20L, at least about 30L, at least about 40L, at least about 50L, at least about 60 L, at least about 70L, at least about 80L, at least about 90L, at least about 100L, at least about 150L, at least about 200L, at least about 250L, at least about 300L, at least about 400L, at least about 500L, at least about 600L, at least about 700L, at least about 800L, at least about 900L, at least about 1000 L, at least about 2000 L, at least about 3000 L, at least about 4000 L, at least about 5000 L, at least about 6000 L, at least about 10,000 L, at least about 15,000 L, at least about 20,000 L, at least about 25,000 L, at least about 30,000 L, at least about 35,000 L, at least about 40,000 L, at least about 45,000 L, at least about 50,000 L, at least about 55,000 L, at least about 60,000 L, at least about 65,000 L, at least about 70,000 L, at least about 75,000 L, at least about 80,000 L, at least about 85,000 L, at least about 90,000 L, at least about 95,000 L, at least about 100,000 L, etc.

VI. Compositions and Formulations

In one aspect, the invention includes a composition that includes a polypeptide derived from *E. coli* or a fragment thereof. In some embodiments, the composition elicits an immune response, including antibodies, that may confer immunity to pathogenic species of *E. coli*.

In some embodiments, the composition includes the polypeptide derived from *E. coli* or fragment thereof as the only antigen. In some embodiments, the composition does not include a conjugate.

In some embodiments, the composition includes the polypeptide derived from *E. coli* or fragment thereof and an additional antigen. In some embodiments, the composition includes the polypeptide derived from *E. coli* or fragment thereof and an additional *E. coli* antigen. In some embodiments, the composition includes the polypeptide derived from *E. coli* or fragment thereof and a glycoconjugate from *E. coli*.

In some embodiments, the polypeptide or a fragment thereof is derived from *E. coli* FimH.

In some embodiments, the composition includes a polypeptide derived from *E. coli* FimC or a fragment thereof.

In some embodiments, the composition includes a polypeptide derived from *E. coli* FimH or a fragment thereof; and a polypeptide derived from *E. coli* FimC or a fragment thereof.

In one aspect, the invention includes a composition including a polypeptide derived from *E. coli* FimH or a fragment thereof; and a saccharide comprising a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula

O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 1 to 100.

In some embodiments, the composition includes any one of the saccharides disclosed herein. In preferred embodiments, the composition includes any one of the conjugates disclosed herein.

In some embodiments, the composition includes at least one glycoconjugate from *E. coli* serotype O25, preferably serotype O25b. In one embodiment, the composition includes at least one glycoconjugate from *E. coli* serotype O1, preferably serotype O1a. In one embodiment, the composition includes at least one glycoconjugate from *E. coli* serotype O2. In one embodiment, the composition includes at least one glycoconjugate from *E. coli* serotype O6.

In one embodiment, the composition comprises at least one glycoconjugate selected from any one of the following *E. coli* serotypes O25, O1, O2, and O6, preferably O25b, O1a, O2, and O6. In one embodiment, the composition comprises at least two glycoconjugates selected from any one of the following *E. coli* serotypes O25, O1, O2, and O6, preferably O25b, O1a, O2, and O6. In one embodiment, the composition comprises at least three glycoconjugates selected from any one of the following *E. coli* serotypes O25, O1, O2, and O6, preferably O25b, O1a, O2, and O6. In one embodiment, the composition comprises a glycoconjugate from each of the following *E. coli* serotypes O25, O1, O2, and O6, preferably O25b, O1a, O2, and O6.

In a preferred embodiment, the glycoconjugate of any of the above compositions is individually conjugated to CRM₁₉₇. In another preferred embodiment, the glycoconjugate of any of the above compositions is individually conjugated to SCP.

Accordingly, in some embodiments, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from at least one *E. coli* serotype. In a preferred embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from more than 1 *E. coli* serotype. For example, the composition may include an O-antigen from two different *E. coli* serotypes (or "v", valences) to 12 different serotypes (12v). In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 3 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 4 different

E. coli serotypes. In one embodiment, the composition includes an O- antigen from 5 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 6 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 7 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 8 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 9 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 10 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 11 different *E. coli* serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 12 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 13 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 14 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 15 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 16 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 17 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 18 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 19 different serotypes. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 20 different serotypes.

Preferably, the number of *E. coli* saccharides can range from 1 serotype (or "v", valences) to 26 different serotypes (26v). In one embodiment there is one serotype. In one embodiment there are 2 different serotypes. In one embodiment there are 3 different serotypes. In one embodiment there are 4 different serotypes. In one embodiment there are 5 different serotypes. In one embodiment there are 6 different serotypes. In one embodiment there are 7 different serotypes. In one embodiment there are 8 different serotypes. In one embodiment there are 9 different serotypes. In one embodiment there are 10 different serotypes. In one embodiment there are 11 different serotypes. In one embodiment there are 12 different serotypes. In one embodiment there are 13 different serotypes. In one embodiment there are 14 different serotypes. In one embodiment there are 15 different serotypes. In one embodiment

there are 16 different serotypes. In one embodiment there are 17 different serotypes. In one embodiment there are 18 different serotypes. In one embodiment there are 19 different serotypes. In one embodiment there are 20 different serotypes. In one embodiment there are 21 different serotypes. In one embodiment there are 22 different serotypes. In one embodiment there are 23 different serotypes. In one embodiment there are 24 different serotypes. In an embodiment there are 25 different serotypes. In one embodiment there are 26 different serotypes. The saccharides are conjugated to a carrier protein to form glycoconjugates as described herein.

In one aspect, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and a glycoconjugate that includes an O-antigen from at least one *E. coli* serogroup, wherein the O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from more than 1 *E. coli* serotype, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 2 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 3 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 4 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 5 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 6 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 7 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 8 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 9 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes an O-antigen from a polypeptide derived from *E. coli* or a fragment thereof; and 10 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one

embodiment, the composition includes an O-antigen from a polypeptide derived from *E. coli* or a fragment thereof; and 11 different *E. coli* serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 12 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 13 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 14 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 15 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 16 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 17 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 18 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 19 different serotypes, wherein each O-antigen is conjugated to a carrier protein. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-antigen from 20 different serotypes, wherein each O-antigen is conjugated to a carrier protein.

In another aspect, the composition includes an O-polysaccharide from at least one *E. coli* serotype. In a preferred embodiment, the composition includes an O-polysaccharide from more than 1 *E. coli* serotype. For example, the composition may include an O-polysaccharide from two different *E. coli* serotypes to 12 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 3 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 4 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 5 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 6 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 7 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 8 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 9 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 10 different *E. coli* serotypes. In one embodiment, the composition

includes an O-polysaccharide from 11 different *E. coli* serotypes. In one embodiment, the composition includes an O-polysaccharide from 12 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 13 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 14 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 15 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 16 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 17 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 18 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 19 different serotypes. In one embodiment, the composition includes an O-polysaccharide from 20 different serotypes.

In a preferred embodiment, the composition includes an O-polysaccharide from at least one *E. coli* serotype, wherein the O-polysaccharide is conjugated to a carrier protein. In a preferred embodiment, the composition includes an O-polysaccharide from more than 1 *E. coli* serotype, wherein each O-polysaccharide is conjugated to a carrier protein. For example, the composition may include an O-polysaccharide from two different *E. coli* serotypes to 12 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 3 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 4 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 5 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 6 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 7 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 8 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 9 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 10 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 11 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition

includes an O-polysaccharide from 12 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 13 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 14 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 15 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 16 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 17 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 18 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 19 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein. In one embodiment, the composition includes an O-polysaccharide from 20 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein.

In a most preferred embodiment, the composition includes an O-polysaccharide from at least one *E. coli* serotype, wherein the O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In a preferred embodiment, the composition includes an O-polysaccharide from more than 1 *E. coli* serotype, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. For example, the composition may include an O-polysaccharide from two different *E. coli* serotypes to 12 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 3 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 4 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 5 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 6 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the

composition includes an O-polysaccharide from 7 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 8 different *E. coli* serotypes, wherein
5 each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 9 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the
10 composition includes an O-polysaccharide from 10 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 11 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-
15 polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 12 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 13 different serotypes, wherein each O-
20 polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 14 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition
25 includes an O-polysaccharide from 15 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 16 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide
30 includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 17 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 18 different serotypes, wherein each O-
35 polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition

includes an O-polysaccharide from 19 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes an O-polysaccharide from 20 different serotypes, wherein each O-polysaccharide is conjugated to a carrier protein, and
 5 wherein the O-polysaccharide includes the O-antigen and core saccharide. In a preferred embodiment, the carrier protein is CRM₁₉₇.

In another preferred embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O25a, wherein n is at least 40, and the core saccharide. In a
 10 preferred embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O25b, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O1a, wherein n is at least 40, and the core saccharide. In another embodiment, the composition
 15 further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O2, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O6, wherein n is at least 40, and the core saccharide.

In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O17, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O15, wherein n is at least 40, and the core saccharide. In another embodiment, the composition
 25 further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O18A, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O75, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide
 30 conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O4, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O16, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide
 35 includes Formula O13, wherein n is at least 40, and the core saccharide. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇,

wherein the O-polysaccharide includes Formula O7, wherein n is at least 40, and the core saccharide.

In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O8, wherein n is at least 40, and the core saccharide. In another embodiment, the O-polysaccharide includes Formula O8, wherein n is 1-20, preferably 2-5, more preferably 3. Formula O8 is shown, e.g., in **FIG. 10B**. In another embodiment, the composition further includes an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O9, wherein n is at least 40, and the core saccharide. In another embodiment, the O-polysaccharide includes Formula O9, wherein n is 1-20, preferably 4-8, more preferably 5. Formula O9 is shown, e.g., in **FIG. 10B**. In another embodiment, the O-polysaccharide includes Formula O9a, wherein n is 1-20, preferably 4-8, more preferably 5. Formula O9a is shown, e.g., in **FIG. 10B**.

In some embodiments, the O-polysaccharide includes selected from any one of Formula O20ab, Formula O20ac, Formula O52, Formula O97, and Formula O101, wherein n is 1-20, preferably 4-8, more preferably 5. See, e.g., **FIG. 10B**.

As described above, the composition may include a polypeptide derived from *E. coli* or a fragment thereof; and any combination of conjugated O-polysaccharides (antigens). In one exemplary embodiment, the composition includes a polysaccharide that includes Formula O25b, a polysaccharide that includes Formula O1A, a polysaccharide that includes Formula O2, and a polysaccharide that includes Formula O6. More specifically, such as a composition that includes: (i) an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O25b, wherein n is at least 40, and the core saccharide; (ii) an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O1a, wherein n is at least 40, and the core saccharide; (iii) an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O2, wherein n is at least 40, and the core saccharide; and (iv) an O-polysaccharide conjugated to CRM₁₉₇, wherein the O-polysaccharide includes Formula O6, wherein n is at least 40, and the core saccharide.

In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and at least one O-polysaccharide derived from any *E. coli* serotype, wherein the serotype is not O25a. For example, in one embodiment, the composition does not include a saccharide that includes the Formula O25a. Such a composition may include, for example, an O-polysaccharide that includes Formula O25b, an O-polysaccharide that includes Formula O1A, an O-polysaccharide that includes Formula O2, and an O-polysaccharide that includes Formula O6.

In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 2 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide
5 derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 3 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 4 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated
10 to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 5 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide
15 derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 6 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 7 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated
20 to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 8 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide
25 derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 9 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 10 different *E. coli* serotypes, wherein each O-polysaccharide is
30 conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 11 different *E. coli* serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide
35 derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 12 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide

includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 13 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇ and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 14 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 15 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 16 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 17 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 18 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 19 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide. In one embodiment, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and an O-polysaccharide from 20 different serotypes, wherein each O-polysaccharide is conjugated to CRM₁₉₇, and wherein the O-polysaccharide includes the O-antigen and core saccharide.

30 In one aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound to a carrier protein, wherein the saccharide includes Formula O25b, wherein n is 15 ± 2 . In one aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound to a carrier protein, wherein the saccharide includes Formula O25b, wherein n is 17 ± 2 . In one aspect, the invention relates to a composition

that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O25b, wherein n is 55 ± 2 . In another aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a
5 saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O25b, wherein n is 51 ± 2 . In one embodiment, the saccharide further includes the *E. coli* R1 core saccharide moiety. In another embodiment, the saccharide further includes the *E. coli* K12 core saccharide moiety. In another embodiment, the saccharide further includes the KDO moiety. Preferably, the carrier protein is CRM₁₉₇. In one embodiment, the conjugate is prepared by
10 single end linked conjugation. In one embodiment, the conjugate is prepared by reductive amination chemistry, preferably in DMSO buffer. In one embodiment, the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl) carbamate (eTEC) spacer. Preferably, the composition further includes a pharmaceutically acceptable diluent.

In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said
15 antibodies being capable of binding an *E. coli* serotype O25B polysaccharide at a concentration of at least 0.2 pg/ml, 0.3 pg/ml, 0.35 pg/ml, 0.4 pg/ml or 0.5 pg/ml as determined by ELISA assay. Therefore, comparison of OPA activity of pre- and post-immunization serum with the immunogenic composition of the invention can be conducted and compared for their response to serotype O25B to assess the potential increase of responders. In one embodiment, the
20 immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of killing *E. coli* serotype O25B as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition elicits functional antibodies in humans, said antibodies being capable of killing *E. coli* serotype O25B as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition of the invention
25 increases the proportion of responders against *E. coli* serotype O25B (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one embodiment, the immunogenic composition elicits a titer of at least 1:8 against *E. coli* serotype O25B in at least 50% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the
30 invention elicits a titer of at least 1:8 against *E. coli* serotype O25B in at least 60%, 70%, 80%, or at least 90% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention significantly increases the proportion of responders against *E. coli* serotypes O25B (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized
35 population. In one embodiment, the immunogenic composition of the invention significantly

increases the OPA titers of human subjects against *E. coli* serotype O25B as compared to the pre-immunized population.

In one aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a
5 carrier protein, wherein the saccharide includes Formula O1a, wherein n is 39 ± 2 . In another aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O1a, wherein n is 13 ± 2 . In one embodiment, the saccharide further includes the *E. coli* R1 core saccharide moiety.
10 In one embodiment, the saccharide further includes the KDO moiety. Preferably, the carrier protein is CRM₁₉₇. In one embodiment, the conjugate is prepared by single end linked conjugation. In one embodiment, the conjugate is prepared by reductive amination chemistry, preferably in DMSO buffer. In one embodiment, the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl) carbamate (eTEC)
15 spacer. Preferably, the composition further includes a pharmaceutically acceptable diluent.

In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of binding an *E. coli* serotype O1A polysaccharide at a concentration of at least 0.2 pg/ml, 0.3 pg/ml, 0.35 pg/ml, 0.4 pg/ml
20 or 0.5 pg/ml as determined by ELISA assay. Therefore, comparison of OPA activity of pre- and post-immunization serum with the immunogenic composition of the invention can be conducted and compared for their response to serotype O1A to assess the potential increase of responders. In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of killing *E. coli* serotype
25 O1A as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition elicits functional antibodies in humans, said antibodies being capable of killing *E. coli* serotype O1A as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition of the invention increases the proportion of responders against *E. coli* serotype O1A (i.e., individual with a serum
30 having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one embodiment, the immunogenic composition elicits a titer of at least 1:8 against *E. coli* serotype O1A in at least 50% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention elicits a titer of at least 1:8 against *E. coli* serotype O1A in
35 at least 60%, 70%, 80%, or at least 90% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of

the invention significantly increases the proportion of responders against *E. coli* serotypes O1A (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one embodiment, the immunogenic composition of the invention significantly increases the OPA titers of human subjects against *E. coli* serotype O1A as compared to the pre-immunized population.

In one aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O2, wherein n is 43 ± 2 . In another aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O2, wherein n is 47 ± 2 . In another aspect, the invention relates to a composition that includes a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O2, wherein n is 17 ± 2 . In another aspect, the invention relates to a composition that includes a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O2, wherein n is 18 ± 2 . In one embodiment, the saccharide further includes the *E. coli* R1 core saccharide moiety. In another embodiment, the saccharide further includes the *E. coli* R4 core saccharide moiety. In another embodiment, the saccharide further includes the KDO moiety. Preferably, the carrier protein is CRM₁₉₇. In one embodiment, the conjugate is prepared by single end linked conjugation. In one embodiment, the conjugate is prepared by reductive amination chemistry, preferably in DMSO buffer. In one embodiment, the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer. Preferably, the composition further includes a pharmaceutically acceptable diluent.

In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of binding an *E. coli* serotype O2 polysaccharide at a concentration of at least 0.2 pg/ml, 0.3 pg/ml, 0.35 pg/ml, 0.4 pg/ml or 0.5 pg/ml as determined by ELISA assay. Therefore, comparison of OPA activity of pre- and post-immunization serum with the immunogenic composition of the invention can be conducted and compared for their response to serotype O2 to assess the potential increase of responders. In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of killing *E. coli* serotype O2 as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition elicits functional antibodies in humans, said antibodies being capable of killing *E. coli* serotype O2 as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition of the invention increases the proportion of responders against *E. coli* serotype O2 (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized

population. In one embodiment, the immunogenic composition elicits a titer of at least 1:8 against *E. coli* serotype O2 in at least 50% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention elicits a titer of at least 1:8 against *E. coli* serotype O2 in at least 60%, 70%, 80%, or at least 90% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention significantly increases the proportion of responders against *E. coli* serotypes O2 (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one embodiment, the immunogenic composition of the invention significantly increases the OPA titers of human subjects against *E. coli* serotype O2 as compared to the pre-immunized population.

In one aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O6, wherein n is 42 ± 2 . In another aspect, the invention relates to a composition that includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O6, wherein n is 50 ± 2 . In another aspect, the invention relates to a composition that includes a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O6, wherein n is 17 ± 2 . In another aspect, the invention relates to a composition that includes a conjugate including a saccharide covalently bound a carrier protein, wherein the saccharide includes Formula O6, wherein n is 18 ± 2 . In one embodiment, the saccharide further includes the *E. coli* R1 core saccharide moiety. In one embodiment, the saccharide further includes the KDO moiety. Preferably, the carrier protein is CRM₁₉₇. In one embodiment, the conjugate is prepared by single end linked conjugation. In one embodiment, the conjugate is prepared by reductive amination chemistry, preferably in DMSO buffer. In one embodiment, the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl) carbamate (eTEC) spacer. Preferably, the composition further includes a pharmaceutically acceptable diluent.

In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of binding an *E. coli* serotype O6 polysaccharide at a concentration of at least 0.2 pg/ml, 0.3 pg/ml, 0.35 pg/ml, 0.4 pg/ml or 0.5 pg/ml as determined by ELISA assay. Therefore, comparison of OPA activity of pre- and post-immunization serum with the immunogenic composition of the invention can be conducted and compared for their response to serotype O6 to assess the potential

increase of responders. In one embodiment, the immunogenic composition elicits IgG antibodies in humans, said antibodies being capable of killing *E. coli* serotype O6 as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition elicits functional antibodies in humans, said antibodies being capable of

5 killing *E. coli* serotype O6 as determined by *in vitro* opsonophagocytic assay. In one embodiment, the immunogenic composition of the invention increases the proportion of responders against *E. coli* serotype O6 (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one

10 embodiment, the immunogenic composition elicits a titer of at least 1:8 against *E. coli* serotype O6 in at least 50% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention elicits a titer of at least 1:8 against *E. coli* serotype O6 in at least 60%, 70%, 80%, or at least 90% of the subjects as determined by *in vitro* opsonophagocytic killing assay. In one embodiment, the immunogenic composition of the invention significantly increases the proportion of responders against *E. coli*

15 serotypes O6 (i.e., individual with a serum having a titer of at least 1:8 as determined by *in vitro* OPA) as compared to the pre-immunized population. In one embodiment, the immunogenic composition of the invention significantly increases the OPA titers of human subjects against *E. coli* serotype O6 as compared to the pre-immunized population.

In one aspect, the composition includes a polypeptide derived from *E. coli* or a fragment thereof; and a conjugate including a saccharide covalently bound to a carrier protein, wherein the saccharide includes a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7,

25 Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula

30 O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60,

35 Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula

O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 1 to 100, preferably from 31 to 90. In one embodiment, the saccharide further includes the *E. coli* R1 core saccharide moiety. In one embodiment, the saccharide further includes the *E. coli* R2 core saccharide moiety. In one embodiment, the saccharide further includes the *E. coli* R3 core saccharide moiety. In another embodiment, the saccharide further includes the *E. coli* R4 core saccharide moiety. In one embodiment, the saccharide further includes the *E. coli* K12 core saccharide moiety. In another embodiment, the saccharide further includes the KDO moiety. Preferably, the carrier protein is CRM₁₉₇. In one embodiment, the conjugate is prepared by single end linked conjugation. In one embodiment, the conjugate is prepared by reductive amination chemistry, preferably in DMSO buffer. In one embodiment, the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer. Preferably, the composition further includes a pharmaceutically acceptable diluent. In one embodiment, the composition further includes at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 additional conjugates to at most 30 additional

conjugates, each conjugate including a saccharide covalently bound to a carrier protein, wherein the saccharide includes a structure selected from any one of said Formulas.

A. Saccharide

In one embodiment, the saccharide is produced by expression (not necessarily overexpression) of different Wzz proteins (e.g., WzzB) to control of the size of the saccharide.

As used herein, the term "saccharide" refers to a single sugar moiety or monosaccharide unit as well as combinations of two or more single sugar moieties or monosaccharide units covalently linked to form disaccharides, oligosaccharides, and polysaccharides. The saccharide may be linear or branched.

In one embodiment, the saccharide is produced in a recombinant Gram-negative bacterium. In one embodiment, the saccharide is produced in a recombinant *E. coli* cell. In one embodiment, the saccharide is produced in a recombinant *Salmonella* cell. Exemplary bacteria include *E. coli* O25K5H1, *E. coli* BD559, *E. coli* GAR2831, *E. coli* GAR865, *E. coli* GAR868, *E. coli* GAR869, *E. coli* GAR872, *E. coli* GAR878, *E. coli* GAR896, *E. coli* GAR1902, *E. coli* O25a ETC NR-5, *E. coli* O157:H7:K-, *Salmonella enterica* serovar Typhimurium strain LT2, *E. coli* GAR2401, *Salmonella enterica* serotype Enteritidis CVD 1943, *Salmonella enterica* serotype Typhimurium CVD 1925, *Salmonella enterica* serotype Paratyphi A CVD 1902, and *Shigella flexneri* CVD 1208S. In one embodiment, the bacterium is not *E. coli* GAR2401. This genetic approach towards saccharide production allows for efficient production of O-polysaccharides and O-antigen molecules as vaccine components.

The term "wzz protein," as used herein, refers to a chain length determinant polypeptide, such as, for example, wzzB, wzz, wzz_{SF}, wzz_{ST}, fepE, wzz_{fepE}, wzz1 and wzz2. The GenBank accession numbers for the exemplary wzz gene sequences are AF011910 for E4991/76, AF011911 for F186, AF011912 for M70/1-1, AF011913 for 79/311, AF011914 for Bi7509-41, AF011915 for C664-1992, AF011916 for C258-94, AF011917 for C722-89, and AF011919 for EDL933. The GenBank accession numbers for the G7 and Bi316-41 wzz genes sequences are U39305 and U39306, respectively. Further GenBank accession numbers for exemplary wzz gene sequences are NP_459581 for *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2 FepE; AIG66859 for *E. coli* O157:H7 Strain EDL933 FepE; NP_461024 for *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2 WzzB. NP_416531 for *E. coli* K-12 substr. MG1655 WzzB, NP_415119 for *E. coli* K-12 substr. MG1655 FepE. In preferred embodiments, the wzz family protein is any one of wzzB, wzz, wzz_{SF}, wzz_{ST}, fepE, wzz_{fepE}, wzz1 and wzz2, most preferably wzzB, more preferably fepE.

Exemplary wzzB sequences include sequences set forth in SEQ ID Nos: 30-34. Exemplary FepE sequences include sequences set forth in SEQ ID Nos: 35-39.

In some embodiments, a modified saccharide (modified as compared to the corresponding wild-type saccharide) may be produced by expressing (not necessarily overexpressing) a wzz family protein (e.g., *fepE*) from a Gram-negative bacterium in a Gram-negative bacterium and/or by switching off (i.e., repressing, deleting, removing) a second wzz gene (e.g., *wzzB*) to generate high molecular weight saccharides, such as lipopolysaccharides, containing intermediate or long O-antigen chains. For example, the modified saccharides may be produced by expressing (not necessarily overexpressing) *wzz2* and switching off *wzzI*. Or, in the alternative, the modified saccharides may be produced by expressing (not necessarily overexpressing) *wzzfepE* and switching off *wzzB*. In another embodiment, the modified saccharides may be produced by expressing (not necessarily overexpressing) *wzzB* but switching off *wzzfepE*. In another embodiment, the modified saccharides may be produced by expressing *fepE*. Preferably, the wzz family protein is derived from a strain that is heterologous to the host cell.

In some embodiments, the saccharide is produced by expressing a wzz family protein having an amino acid sequence that is at least 30%, 50%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, and SEQ ID NO: 39. In one embodiment, the wzz family protein includes a sequence selected from any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, and SEQ ID NO: 39. Preferably, the wzz family protein has at least 30%, 50%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34. In some embodiments, the saccharide is produced by expressing a protein having an amino acid sequence that is at least 30%, 50%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an *fepE* protein.

In one aspect, the invention relates to saccharides produced by expressing a wzz family protein, preferably *fepE*, in a Gram-negative bacterium to generate high molecular weight saccharides containing intermediate or long O-antigen chains, which have an increase of at least 1, 2, 3, 4, or 5 repeating units, as compared to the corresponding wild-type O-polysaccharide. In one aspect, the invention relates to saccharides produced by a Gram-negative bacterium in culture that expresses (not necessarily overexpresses) a wzz family protein (e.g., *wzzB*) from a Gram-negative bacterium to generate high molecular weight saccharides containing intermediate or long O-antigen chains, which have an increase of at least 1, 2, 3, 4, or 5 repeating units,

as compared to the corresponding wild-type O-antigen. See description of O-polysaccharides and O-antigens below for additional exemplary saccharides having increased number of repeat units, as compared to the corresponding wild-type saccharides. A desired chain length is the one which produces improved or maximal immunogenicity in the context of a given vaccine construct.

In another embodiment, the saccharide includes any one Formula selected from **Table 1**, wherein the number of repeat units n in the saccharide is greater than the number of repeat units in the corresponding wild-type O-polysaccharide by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more repeat units. Preferably, the saccharide includes an increase of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 repeat units, as compared to the corresponding wild-type O-polysaccharide. See, for example, **Table 24**. Methods of determining the length of saccharides are known in the art. Such methods include nuclear magnetic resonance, mass spectroscopy, and size exclusion chromatography, as described in **Example 13**.

In a preferred embodiment, the invention relates to a saccharide produced in a recombinant *E. coli* host cell, wherein the gene for an endogenous wzz O-antigen length regulator (e.g., wzzB) is deleted and is replaced by a (second) wzz gene from a Gram-negative bacterium heterologous to the recombinant *E. coli* host cell (e.g., *Salmonella* fepE) to generate high molecular weight saccharides, such as lipopolysaccharides, containing intermediate or long O-antigen chains. In some embodiments, the recombinant *E. coli* host cell includes a wzz gene from *Salmonella*, preferably from *Salmonella enterica*.

In one embodiment, the host cell includes the heterologous gene for a wzz family protein as a stably maintained plasmid vector. In another embodiment, the host cell includes the heterologous gene for a wzz family protein as an integrated gene in the chromosomal DNA of the host cell. Methods of stably expressing a plasmid vector in an *E. coli* host cell and methods of integrating a heterologous gene into the chromosome of an *E. coli* host cell are known in the art. In one embodiment, the host cell includes the heterologous genes for an O-antigen as a stably maintained plasmid vector. In another embodiment, the host cell includes the heterologous genes for an O-antigen as an integrated gene in the chromosomal DNA of the host cell. Methods of stably expressing a plasmid vector in an *E. coli* host cell and a *Salmonella* host cell are known in the art. Methods of integrating a heterologous gene into the chromosome of an *E. coli* host cell and a *Salmonella* host cell are known in the art.

In one aspect, the recombinant host cell is cultured in a medium that comprises a carbon source. Carbon sources for culturing *E. coli* are known in the art. Exemplary carbon sources include sugar alcohols, polyols, aldol sugars or keto sugars including but not limited to arabinose, cellobiose, fructose, glucose, glycerol, inositol, lactose, maltose, mannitol, mannose, 5 rhamnose, raffinose, sorbitol, sorbose, sucrose, trehalose, pyruvate, succinate and methylamine. In a preferred embodiment, the medium includes glucose. In some embodiments, the medium includes a polyol or aldol sugar, for example, mannitol, inositol, sorbose, glycerol, sorbitol, lactose and arabinose as the carbon source. All of the carbon sources may be added to the medium before the start of culturing, or it may 10 be added step by step or continuously during culturing.

An exemplary culture medium for the recombinant host cell includes an element selected from any one of KH_2PO_4 , K_2HPO_4 , $(\text{NH}_4)_2\text{SO}_4$, sodium citrate, Na_2SO_4 , aspartic acid, glucose, MgSO_4 , $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, H_3BO_3 , $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, ZnCl_2 and $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$. Preferably, the medium includes KH_2PO_4 , 15 K_2HPO_4 , $(\text{NH}_4)_2\text{SO}_4$, sodium citrate, Na_2SO_4 , aspartic acid, glucose, MgSO_4 , $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, H_3BO_3 , $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, ZnCl_2 and $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$.

The medium used herein may be solid or liquid, synthetic (i.e. man-made) or natural, and may include sufficient nutrients for the cultivation of the recombinant host 20 cell. Preferably, the medium is a liquid medium.

In some embodiments, the medium may further include suitable inorganic salts. In some embodiments, the medium may further include trace nutrients. In some embodiments, the medium may further include growth factors. In some embodiments, the medium may further include an additional carbon source. In some embodiments, 25 the medium may further include suitable inorganic salts, trace nutrients, growth factors, and a supplementary carbon source. Inorganic salts, trace nutrients, growth factors, and supplementary carbon sources suitable for culturing *E. coli* are known in the art.

In some embodiments, the medium may include additional components as appropriate, such as peptone, N-Z Amine, enzymatic soy hydrolysate, additional yeast 30 extract, malt extract, supplemental carbon sources and various vitamins. In some embodiments, the medium does not include such additional components, such as peptone, N-Z Amine, enzymatic soy hydrolysate, additional yeast extract, malt extract, supplemental carbon sources and various vitamins.

Illustrative examples of suitable supplemental carbon sources include, but are 35 not limited to other carbohydrates, such as glucose, fructose, mannitol, starch or starch hydrolysate, cellulose hydrolysate and molasses; organic acids, such as acetic acid,

propionic acid, lactic acid, formic acid, malic acid, citric acid, and fumaric acid; and alcohols, such as glycerol, inositol, mannitol and sorbitol.

In some embodiments, the medium further includes a nitrogen source. Nitrogen sources suitable for culturing *E. coli* are known in the art. Illustrative examples of suitable nitrogen sources include, but are not limited to ammonia, including ammonia gas and aqueous ammonia; ammonium salts of inorganic or organic acids, such as ammonium chloride, ammonium nitrate, ammonium phosphate, ammonium sulfate and ammonium acetate; urea; nitrate or nitrite salts, and other nitrogen-containing materials, including amino acids as either pure or crude preparations, meat extract, peptone, fish meal, fish hydrolysate, corn steep liquor, casein hydrolysate, soybean cake hydrolysate, yeast extract, dried yeast, ethanol-yeast distillate, soybean flour, cottonseed meal, and the like.

In some embodiments, the medium includes an inorganic salt. Illustrative examples of suitable inorganic salts include, but are not limited to salts of potassium, calcium, sodium, magnesium, manganese, iron, cobalt, zinc, copper, molybdenum, tungsten and other trace elements, and phosphoric acid.

In some embodiments, the medium includes appropriate growth factors. Illustrative examples of appropriate trace nutrients, growth factors, and the like include, but are not limited to coenzyme A, pantothenic acid, pyridoxine-HCl, biotin, thiamine, riboflavin, flavine mononucleotide, flavine adenine dinucleotide, DL-6,8-thioctic acid, folic acid, Vitamin B₁₂, other vitamins, amino acids such as cysteine and hydroxyproline, bases such as adenine, uracil, guanine, thymine and cytosine, sodium thiosulfate, p- or r-aminobenzoic acid, niacinamide, nitriloacetate, and the like, either as pure or partially purified chemical compounds or as present in natural materials. The amounts may be determined empirically by one skilled in the art according to methods and techniques known in the art.

In another embodiment, the modified saccharide (as compared to the corresponding wild-type saccharide) described herein is synthetically produced, for example, *in vitro*. Synthetic production or synthesis of the saccharides may facilitate the avoidance of cost- and time-intensive production processes. In one embodiment, the saccharide is synthetically synthesized, such as, for example, by using sequential glycosylation strategy or a combination of sequential glycosylations and [3+2] block synthetic strategy from suitably protected monosaccharide intermediates. For example, thioglycosides and glycosyl trichloroacetimidate derivatives may be used as glycosyl donors in the glycosylations. In one embodiment, a saccharide that is synthetically synthesized *in vitro* has the identical structure to a saccharide produced by recombinant means, such as by manipulation of a wzz family protein described above.

The saccharide produced (by recombinant or synthetic means) includes a structure derived from any *E. coli* serotype including, for example, any one of the following *E. coli* serotypes: O1 (e.g., O1A, O1B, and O1C), O2, O3, O4 (e.g., O4:K52 and O4:K6), O5 (e.g., O5ab and O5ac (strain 180/C3)), O6 (e.g., O6:K2; K13; K15 and O6:K54), O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18 (e.g., O18A, O18ac, O18A1, O18B, and O18B1), O19, O20, O21, O22, O23 (e.g., O23A), O24, O25 (e.g., O25a and O25b), O26, O27, O28, O29, O30, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45 (e.g., O45 and O45rel), O46, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, 62D₁, O63, O64, O65, O66, O68, O69, O70, O71, O73 (e.g., O73 (strain 73-1)), O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O95, O96, O97, O98, O99, O100, O101, O102, O103, O104, O105, O106, O107, O108, O109, O110, O111, O112, O113, O114, O115, O116, O117, O118, O119, O120, O121, O123, O124, O125, O126, O127, O128, O129, O130, O131, O132, O133, O134, O135, O136, O137, O138, O139, O140, O141, O142, O143, O144, O145, O146, O147, O148, O149, O150, O151, O152, O153, O154, O155, O156, O157, O158, O159, O160, O161, O162, O163, O164, O165, O166, O167, O168, O169, O170, O171, O172, O173, O174, O175, O176, O177, O178, O179, O180, O181, O182, O183, O184, O185, O186, and O187.

The individual polysaccharides are typically purified (enriched with respect to the amount of polysaccharide-protein conjugate) through methods known in the art, such as, for example, dialysis, concentration operations, diafiltration operations, tangential flow filtration, precipitation, elution, centrifugation, precipitation, ultra-filtration, depth filtration, and/or column chromatography (ion exchange chromatography, multimodal ion exchange chromatography, DEAE, and hydrophobic interaction chromatography). Preferably, the polysaccharides are purified through a method that includes tangential flow filtration.

Purified polysaccharides may be activated (e.g., chemically activated) to make them capable of reacting (e.g., either directly to the carrier protein or via a linker such as an eTEC spacer) and then incorporated into glycoconjugates of the invention, as further described herein.

In one preferred embodiment, the saccharide of the invention is derived from an *E. coli* serotype, wherein the serotype is O25a. In another preferred embodiment, the serotype is O25b. In another preferred embodiment, the serotype is O1A. In another preferred embodiment, the serotype is O2. In another preferred embodiment, the

serotype is O6. In another preferred embodiment, the serotype is O17. In another preferred embodiment, the serotype is O15. In another preferred embodiment, the serotype is O18A. In another preferred embodiment, the serotype is O75. In another preferred embodiment, the serotype is O4. In another preferred embodiment, the serotype is O16. In another preferred embodiment, the serotype is O13. In another preferred embodiment, the serotype is O7. In another preferred embodiment, the serotype is O8. In another preferred embodiment, the serotype is O9.

As used herein, reference to any of the serotypes listed above, refers to a serotype that encompasses a repeating unit structure (O-unit, as described below) known in the art and is unique to the corresponding serotype. For example, the term "O25a" serotype (also known in the art as serotype "O25") refers to a serotype that encompasses Formula O25 shown in **Table 1**. As another example, the term "O25b" serotype refers to a serotype that encompasses Formula O25b shown in **Table 1**.

As used herein, the serotypes are referred generically herein unless specified otherwise such that, for example, the term Formula "O18" refers generically to encompass Formula O18A, Formula O18ac, Formula 18A1, Formula O18B, and Formula O18B1.

As used herein, the term "O1" refers generically to encompass the species of Formula that include the generic term "O1" in the Formula name according to **Table 1**, such as any one of Formula O1A, Formula O1A1, Formula O1B, and Formula O1C, each of which is shown in **Table 1**. Accordingly, an "O1 serotype" refers generically to a serotype that encompasses any one of Formula O1A, Formula O1A1, Formula O1B, and Formula O1C.

As used herein, the term "O6" refers generically to species of Formula that include the generic term "O6" in the Formula name according to **Table 1**, such as any one of Formula O6:K2; K13; K15; and O6:K54, each of which is shown in **Table 1**. Accordingly, an "O6 serotype" refers generically to a serotype that encompasses any one of Formula O6:K2; K13; K15; and O6:K54.

Other examples of terms that refer generically to species of a Formula that include the generic term in the Formula name according to **Table 1** include: "O4", "O5", "O18", and "O45".

As used herein, the term "O2" refers to Formula O2 shown in **Table 1**. The term "O2 O-antigen" refers to a saccharide that encompasses Formula O2 shown in **Table 1**.

As used herein, reference to an O-antigen from a serotype listed above refers to a saccharide that encompasses the formula labeled with the corresponding serotype name. For example, the term "O25B O-antigen" refers to a saccharide that encompasses Formula O25B shown in **Table 1**.

As another example, the term "O1 O-antigen" generically refers to a saccharide that encompasses a Formula including the term "O1," such as the Formula O1A, Formula O1A1, Formula O1B, and Formula O1C, each of which are shown in **Table 1**.

5 As another example, the term "O6 O-antigen" generically refers to a saccharide that encompasses a Formula including the term "O6," such as Formula O6:K2; Formula O6:K13; Formula O6:K15 and Formula O6:K54, each of which are shown in **Table 1**.

B. O-Polysaccharide

As used herein, the term "O-polysaccharide" refers to any structure that includes an O-antigen, provided that the structure does not include a whole cell or Lipid A. For
10 example, in one embodiment, the O-polysaccharide includes a lipopolysaccharide wherein the Lipid A is not bound. The step of removing Lipid A is known in the art and includes, as an example, heat treatment with addition of an acid. An exemplary process includes treatment with 1% acetic acid at 100°C for 90 minutes. This process is combined with a process of isolating Lipid A as removed. An exemplary process for
15 isolating Lipid A includes ultracentrifugation.

In one embodiment, the O-polysaccharide refers to a structure that consists of the O-antigen, in which case, the O-polysaccharide is synonymous with the term O-antigen. In one preferred embodiment, the O-polysaccharide refers to a structure that includes repeating units of the O-antigen, without the core saccharide. Accordingly, in
20 one embodiment, the O-polysaccharide does not include an *E. coli* R1 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R2 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R3 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R4 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli*
25 K12 core moiety. In another preferred embodiment, the O-polysaccharide refers to a structure that includes an O-antigen and a core saccharide. In another embodiment, the O-polysaccharide refers to a structure that includes an O-antigen, a core saccharide, and a KDO moiety.

Methods of purifying an O-polysaccharide, which includes the core
30 oligosaccharide, from LPS are known in the art. For example, after purification of LPS, purified LPS may be hydrolyzed by heating in 1% (v/v) acetic acid for 90 minutes at 100 degrees Celsius, followed by ultracentrifugation at 142,000 x g for 5 hours at 4 degrees Celsius. The supernatant containing the O-polysaccharide is freeze-dried and stored at 4 degrees Celsius. In certain embodiments, deletion of capsule synthesis genes to
35 enable simple purification of O-polysaccharide is described.

The O-polysaccharide can be isolated by methods including, but not limited to mild acid hydrolysis to remove lipid A from LPS. Other embodiments may include use of hydrazine as an agent for O-polysaccharide preparation. Preparation of LPS can be accomplished by known methods in the art.

5 In certain embodiments, the O-polysaccharides purified from wild-type, modified, or attenuated Gram-negative bacterial strains that express (not necessarily overexpress) a Wzz protein (e.g., wzzB) are provided for use in conjugate vaccines. In preferred embodiments, the O-polysaccharide chain is purified from the Gram-negative bacterial strain expressing (not necessarily overexpressing) wzz protein for use as a vaccine antigen either as a conjugate or
10 complexed vaccine.

In one embodiment, the O-polysaccharide has a molecular weight that is increased by about 1-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-
15 fold, 25-fold, 26-fold, 27-fold, 28-fold, 29-fold, 30-fold, 31-fold, 32-fold, 33-fold, 34-fold, 35-fold, 36-fold, 37-fold, 38-fold, 39-fold, 40-fold, 41-fold, 42-fold, 43-fold, 44-fold, 45-fold, 46-fold, 47-fold, 48-fold, 49-fold, 50-fold, 51-fold, 52-fold, 53-fold, 54-fold, 55-fold, 56-fold, 57-fold, 58-fold, 59-fold, 60-fold, 61-fold, 62-fold, 63-fold, 64-fold, 65-fold, 66-fold, 67-fold, 68-fold, 69-fold, 70-fold, 71-fold, 72-fold, 73-fold, 74-fold, 75-fold, 76-fold, 77-fold, 78-fold, 79-fold, 80-fold, 81-fold, 82-fold, 83-fold, 84-fold, 85-fold, 86-fold, 87-fold, 88-fold, 89-fold, 90-fold, 91-fold, 92-fold, 93-
20 fold, 94-fold, 95-fold, 96-fold, 97-fold, 98-fold, 99-fold, 100-fold or more, as compared to the corresponding wild-type O-polysaccharide. In a preferred embodiment, the O-polysaccharide has a molecular weight that is increased by at least 1-fold and at most 5-fold, as compared to the corresponding wild-type O-polysaccharide. In another embodiment, the O-polysaccharide has a molecular weight that is increased by at least 2-fold and at most 4-fold, as compared to
25 the corresponding wild-type O-polysaccharide. An increase in molecular weight of the O-polysaccharide, as compared to the corresponding wild-type O-polysaccharide, is preferably associated with an increase in number of O-antigen repeat units. In one embodiment, the increase in molecular weight of the O-polysaccharide is due to the wzz family protein.

In one embodiment, the O-polysaccharide has a molecular weight that is increased by
30 about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 kDa or more, as compared to the corresponding wild-type O-polysaccharide. In one embodiment,
35 the O-polysaccharide of the invention has a molecular weight that is increased by at least 1 and at most 200 kDa, as compared to the corresponding wild-type O-polysaccharide. In one

[illegible]

is increased by at least 20 and at most 43kDa. In one embodiment, the molecular weight is increased by at least 20 and at most 42kDa. In one embodiment, the molecular weight is increased by at least 20 and at most 41kDa. Such an increase in molecular weight of the O-polysaccharide, as compared to the corresponding wild-type O-polysaccharide, is preferably associated with an increase in number of O-antigen repeat units. In one embodiment, the increase in molecular weight of the O-polysaccharide is due to the wzz family protein. See, for example, **Table 21**.

In another embodiment, the O-polysaccharide includes any one Formula selected from **Table 1**, wherein the number of repeat units n in the O-polysaccharide is greater than the number of repeat units in the corresponding wild-type O-polysaccharide by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more repeat units. Preferably, the saccharide includes an increase of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 repeat units, as compared to the corresponding wild-type O-polysaccharide. See, for example, **Table 21**.

C. O-Antigen

The O-antigen is part of the lipopolysaccharide (LPS) in the outer membrane of Gram-negative bacteria. The O-antigen is on the cell surface and is a variable cell constituent. The variability of the O-antigen provides a basis for serotyping of Gram-negative bacteria. The current *E. coli* serotyping scheme includes O-polysaccharides 1 to 181.

The O-antigen includes oligosaccharide repeating units (O-units), the wild type structure of which usually contains two to eight residues from a broad range of sugars. The O-units of exemplary *E. coli* O-antigens are shown in **Table 1**, see also **FIG. 9A-9C** and **FIG. 10A-10B**.

In one embodiment, the saccharide of the invention may be one oligosaccharide unit. In one embodiment, the saccharide of the invention is one repeating oligosaccharide unit of the relevant serotype. In such embodiments, the saccharide may include a structure selected from any one of Formula O1a, Formula O2, Formula O6, Formula O8, Formula O9a, Formula O9, Formula O20ab, Formula O20ac, Formula O25b, Formula O52, Formula O97, and Formula O101. In a further embodiment, the saccharide may include a structure selected from any one of Formula O1a, Formula O2, Formula O6, and Formula O25b.

In one embodiment, the saccharide of the invention may be oligosaccharides. Oligosaccharides have a low number of repeat units (typically 5-15 repeat units) and are typically derived synthetically or by hydrolysis of polysaccharides. In such embodiments, the

saccharide may include a structure selected from any one of Formula O1a, Formula O2, Formula O6, Formula O8, Formula O9a, Formula O9, Formula O20ab, Formula O20ac, Formula O25b, Formula O52, Formula O97, and Formula O101. In a further embodiment, the saccharide may include a structure selected from any one of Formula O1a, Formula O2,
5 Formula O6, and Formula O25b.

Preferably, all of the saccharides of the present invention and in the immunogenic compositions of the present invention are polysaccharides. High molecular weight polysaccharides may induce certain antibody immune responses due to the epitopes present on the antigenic surface. The isolation and purification of high
10 molecular weight polysaccharides are preferably contemplated for use in the conjugates, compositions and methods of the present invention.

In some embodiments, the number of repeat O units in each individual O-antigen polymer (and therefore the length and molecular weight of the polymer chain) depends on the wzz chain length regulator, an inner membrane protein. Different wzz proteins
15 confer different ranges of modal lengths (4 to >100 repeat units). The term "modal length" refers to the number of repeating O-units. Gram-negative bacteria often have two different Wzz proteins that confer two distinct OAg modal chain lengths, one longer and one shorter. The expression (not necessarily the overexpression) of wzz family proteins (e.g., wzzB) in Gram-negative bacteria may allow for the manipulation of O-
20 antigen length, to shift or to bias bacterial production of O-antigens of certain length ranges, and to enhance production of high-yield large molecular weight lipopolysaccharides. In one embodiment, a "short" modal length as used herein refers to a low number of repeat O-units, e.g., 1-20. In one embodiment, a "long" modal length as used herein refers to a number of repeat O-units greater than 20 and up to a
25 maximum of 40. In one embodiment, a "very long" modal length as used herein refers to greater than 40 repeat O-units.

In one embodiment, the saccharide produced has an increase of at least 10 repeating units, 15 repeating units, 20 repeating units, 25 repeating units, 30 repeating units, 35 repeating units, 40 repeating units, 45 repeating units, 50 repeating units, 55
30 repeating units, 60 repeating units, 65 repeating units, 70 repeating units, 75 repeating units, 80 repeating units, 85 repeating units, 90 repeating units, 95 repeating units, or 100 repeating units, as compared to the corresponding wild-type O-polysaccharide.

In another embodiment, the saccharide of the invention has an increase of 1, 2,
3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,
35 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73,

74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more repeat units, as compared to the corresponding wild-type O-polysaccharide. Preferably, the saccharide includes an increase of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 repeat units, as compared to the corresponding wild-type O-polysaccharide. See, for example, **Table 21**. Methods of determining the length of saccharides are known in the art. Such methods include nuclear magnetic resonance, mass spectroscopy, and size exclusion chromatography, as described in **Example 13**.

Methods of determining the number of repeat units in the saccharide are also known in the art. For example, the number of repeat units (or " n " in the Formula) may be calculated by dividing the molecular weight of the polysaccharide (without the molecular weight of the core saccharide or KDO residue) by the molecular weight of the repeat unit (i.e., molecular weight of the structure in the corresponding Formula, shown for example in **Table 1**, which may be theoretically calculated as the sum of the molecular weight of each monosaccharide within the Formula). The molecular weight of each monosaccharide within the Formula is known in the art. The molecular weight of a repeat unit of Formula O25b, for example, is about 862 Da. The molecular weight of a repeat unit of Formula O1a, for example, is about 845 Da. The molecular weight of a repeat unit of Formula O2, for example, is about 829 Da. The molecular weight of a repeat unit of Formula O6, for example, is about 893 Da. When determining the number of repeat units in a conjugate, the carrier protein molecular weight and the protein:polysaccharide ratio is factored into the calculation. As defined herein, " n " refers to the number of repeating units (represented in brackets in **Table 1**) in a polysaccharide molecule. As is known in the art, in biological macromolecules, repeating structures may be interspersed with regions of imperfect repeats, such as, for example, missing branches. In addition, it is known in the art that polysaccharides isolated and purified from natural sources such as bacteria may be heterogenous in size and in branching. In such a case, n may represent an average or median value for n for the molecules in a population.

In one embodiment, the O-polysaccharide has an increase of at least one repeat unit of an O-antigen, as compared to the corresponding wild-type O-polysaccharide. The repeat units of O-antigens are shown in **Table 1**. In one embodiment, the O-polysaccharide includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more total repeat units. Preferably, the saccharide has a total of at least 3 to at most 80 repeat units. In another embodiment, the O-polysaccharide has an increase of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,

15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37,
 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60,
 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83,
 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more repeat units,
 5 as compared to the corresponding wild-type O-polysaccharide.

In one embodiment, the saccharide includes an O-antigen wherein n in any of the O-antigen formulas (such as, for example, the Formulas shown in **Table 1** (see also **FIG. 9A-9C** and **FIG. 10A-10B**)) is an integer of at least 1, 2, 3, 4, 5, 10, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, and at most 200, 100,
 10 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 68, 67, 66, 65, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51, or 50. Any minimum value and any maximum value may be combined to define a range. Exemplary ranges include, for example, at least 1 to at most 1000; at least 10 to at most 500; and at least 20 to at most 80, preferably at most 90. In one preferred embodiment, n is at
 15 least 31 to at most 90. In a preferred embodiment, n is 40 to 90, more preferably 60 to 85.

In one embodiment, the saccharide includes an O-antigen wherein n in any one of the O-antigen Formulas is at least 1 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 5 and at most 200. In one embodiment, n in
 20 any one of the O-antigen Formulas is at least 10 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 25 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 50 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 75 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 100 and at most
 25 200. In one embodiment, n in any one of the O-antigen Formulas is at least 125 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 150 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 175 and at most 200. In one embodiment, n in any one of the O-antigen Formulas is at least 1 and at most 100. In one embodiment, n in any one of the O-antigen Formulas is
 30 at least 5 and at most 100. In one embodiment, n in any one of the O-antigen Formulas is at least 10 and at most 100. In one embodiment, n in any one of the O-antigen Formulas is at least 25 and at most 100. In one embodiment, n in any one of the O-antigen Formulas is at least 50 and at most 100. In one embodiment, n in any one of the O-antigen Formulas is at least 75 and at most 100. In one embodiment, n in any one of
 35 the O-antigen Formulas is at least 1 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 5 and at most 75. In one embodiment, n in any one

of the O-antigen Formulas is at least 10 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 20 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 25 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 30 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 40 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 50 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 30 and at most 90. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 85. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 75. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 70. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 60. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 50. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 49. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 48. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 47. In one embodiment, n in any one of the O-antigen Formulas is at least 35 and at most 46. In one embodiment, n in any one of the O-antigen Formulas is at least 36 and at most 45. In one embodiment, n in any one of the O-antigen Formulas is at least 37 and at most 44. In one embodiment, n in any one of the O-antigen Formulas is at least 38 and at most 43. In one embodiment, n in any one of the O-antigen Formulas is at least 39 and at most 42. In one embodiment, n in any one of the O-antigen Formulas is at least 39 and at most 41.

For example, in one embodiment, n in the saccharide is 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90, most preferably 40. In another embodiment, n is at least 35 to at most 60. For example, in one embodiment, n is any one of 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, and 60, preferably 50. In another preferred embodiment, n is at least 55 to at most 75. For example, in one embodiment, n is 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, or 69, most preferably 60.

The saccharide structure may be determined by methods and tools known art, such as, for example, NMR, including 1D, 1H, and/or 13C, 2D TOCSY, DQF-COSY, NOESY, and/or HMQC.

In some embodiments, the purified polysaccharide before conjugation has a molecular weight of between 5 kDa and 400 kDa. In other such embodiments, the saccharide has a molecular weight of between 10 kDa and 400 kDa; between 5 kDa and 400 kDa; between 5 kDa and 300 kDa; between 5 kDa and 200 kDa; between 5 kDa and 150 kDa; between 10 kDa

and 100 kDa; between 10 kDa and 75 kDa; between 10 kDa and 60 kDa; between 10 kDa and 40 kDa; between 10 kDa and 100 kDa; 10 kDa and 200 kDa; between 15 kDa and 150 kDa; between 12 kDa and 120 kDa; between 12 kDa and 75 kDa; between 12 kDa and 50 kDa; between 12 and 60 kDa; between 35 kDa and 75 kDa; between 40 kDa and 60 kDa; between 35 kDa and 60 kDa; between 20 kDa and 60 kDa; between 12 kDa and 20 kDa; or between 20 kDa and 50 kDa. In further embodiments, the polysaccharide has a molecular weight of between 7 kDa to 15 kDa; 8 kDa to 16 kDa; 9 kDa to 25 kDa; 10 kDa to 100; 10 kDa to 60 kDa; 10 kDa to 70 kDa; 10 kDa to 160 kDa; 15 kDa to 600 kDa; 20 kDa to 1000 kDa; 20 kDa to 600 kDa; 20 kDa to 400 kDa; 30 kDa to 1,000 kDa; 30 kDa to 60 kDa; 30 kDa to 50 kDa or 5 kDa to 60 kDa. Any whole number integer within any of the above ranges is contemplated as an embodiment of the disclosure.

As used herein, the term "molecular weight" of polysaccharide or of carrier protein-polysaccharide conjugate refers to molecular weight calculated by size exclusion chromatography (SEC) combined with multiangle laser light scattering detector (MALLS).

A polysaccharide can become slightly reduced in size during normal purification procedures. Additionally, as described herein, polysaccharide can be subjected to sizing techniques before conjugation. Mechanical or chemical sizing maybe employed. Chemical hydrolysis may be conducted using acetic acid. Mechanical sizing may be conducted using High Pressure Homogenization Shearing. The molecular weight ranges mentioned above refer to purified polysaccharides before conjugation (e.g., before activation).

Table 1: *E. coli* serogroups/serotypes and O-unit moieties

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O1A, O1A1	[$\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 3)- β -L-Rha-(1 $\rightarrow$ 4)- β -D-GlcNAc-(1 $\rightarrow$ β -D-ManNAc-(1 $\rightarrow$ 2)] _n	Formula O1A
O1B	[$\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 2)- α -D-Gal-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ β -D-ManNAc-(1 $\rightarrow$ 2)] _n	Formula O1B

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O1C	$[\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \beta\text{-D-ManNAc-(1}\rightarrow 2)\text{]}_n$	Formula O1C
O2	$[\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\beta\text{-L-Rha-(1}\rightarrow 4)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \alpha\text{-D-Fuc3NAc-(1}\rightarrow 2)\text{]}_n$	Formula O2
O3	$[\beta\text{-L-RhaNAc(1}\rightarrow 4)\alpha\text{-D-Glc-(1}\rightarrow 4) \rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow 3)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \text{]}_n$	Formula O3
O4:K52	$[\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-FucNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc(1}\rightarrow \text{]}_n$	Formula O4:K52
O4:K6	$[\alpha\text{-D-Glc-(1}\rightarrow 3)\text{ } \rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-FucNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc(1}\rightarrow \text{]}_n$	Formula O4:K6
O5ab	$[\rightarrow 4)\text{-}\beta\text{-D-Qui3NAc-(1}\rightarrow 3)\text{-}\beta\text{-D-Ribf-(1}\rightarrow 4)\text{-}\beta\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GalNAc(1}\rightarrow \text{]}_n$	Formula O5ab
O5ac (strain 180/C3)	$[\rightarrow 2)\text{-}\beta\text{-D-Qui3NAc-(1}\rightarrow 3)\text{-}\beta\text{-D-Ribf-(1}\rightarrow 4)\text{-}\beta\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GalNAc(1}\rightarrow \text{]}_n$	Formula O5ac (strain 180/C3)
O6:K2; K13; K15	$[\rightarrow 4)\text{-}\alpha\text{-D-GalNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-Man-(1}\rightarrow 4)\text{-}\beta\text{-D-Man-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \beta\text{-D-Glc-(1}\rightarrow 2)\text{]}_n$	Formula O6:K2; K13; K15
O6:K54	$[\rightarrow 4)\text{-}\alpha\text{-D-GalNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-Man-(1}\rightarrow 4)\text{-}\beta\text{-D-Man-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \beta\text{-D-GlcNAc-(1}\rightarrow 2)\text{]}_n$	Formula O6:K54
O7	$[\alpha\text{-L-Rha-(1}\rightarrow 3)\text{ } \rightarrow 3)\text{-}\beta\text{-D-Qui4NAc-(1}\rightarrow 2)\text{-}\alpha\text{-D-Man-(1}\rightarrow 4)\text{-}\beta\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \text{]}_n$	Formula O7
O10	$[\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \alpha\text{-D-Fuc4NAcyl-(1}\rightarrow 2)\text{ Acyl=acetyl (60%) or (R)-3-hydroxybutyryl (40%)]}_n$	Formula O10

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O16	$[\rightarrow 2)\text{-}\beta\text{-D-Galf-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rha2Ac-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow)_n$	Formula O16
O17	$[\alpha\text{-D-Glc-(1}\rightarrow 6) \mid \rightarrow 6)\text{-}\alpha\text{-D-Man-(1}\rightarrow 2)\text{-}\alpha\text{-D-Man-(1}\rightarrow 2)\text{-}\beta\text{-D-Man-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc(1}\rightarrow)_n$	Formula O17
O18A, O18ac	$[\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 4)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \mid \beta\text{-D-GlcNAc-(1}\rightarrow 3)]_n$	Formula O18A, Formula O18ac
O18A1	$[\alpha\text{-D-Glc-(1}\rightarrow 6) \mid \rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 4)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \mid \beta\text{-D-GlcNAc-(1}\rightarrow 3)]_n$	Formula O18A1
O18B	$[\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 4)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \mid \beta\text{-D-Glc-(1}\rightarrow 3)]_n$	Formula O18B
O18B1	$[\alpha\text{-D-Glc-(1}\rightarrow 4) \mid \rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 4)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow \mid \beta\text{-D-Glc-(1}\rightarrow 3)]_n$	Formula O18B1
O21	$[\beta\text{-D-Gal-(1}\rightarrow 4) \mid \rightarrow 3)\text{-}\beta\text{-D-Gal-(1}\rightarrow 4)\text{-}\beta\text{-D-Glc-(1}\rightarrow 3)\text{-}\beta\text{-D-GalNAc-(1}\rightarrow \mid \beta\text{-D-GlcNAc-(1}\rightarrow 2)]_n$	Formula O21
O23A	$[\alpha\text{-D-Glc-(1}\rightarrow 6) \mid \rightarrow 6)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 4)\text{-}\beta\text{-D-Gal-(1}\rightarrow 3)\text{-}\alpha\text{-D-GalNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \mid \beta\text{-D-GlcNAc(1}\rightarrow 3)]_n$	Formula O23A
O24	$[\rightarrow 7)\text{-}\alpha\text{-Neu5Ac-(2}\rightarrow 3)\text{-}\beta\text{-D-Glc-(1}\rightarrow 3)\text{-}\beta\text{-D-GalNAc-(1}\rightarrow \mid \alpha\text{-D-Glc-(1}\rightarrow 2)]_n$	Formula O24
O25/O25a	$[\beta\text{-D-Glc-(1}\rightarrow 6) \mid \rightarrow 4)\text{-}\alpha\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-FucNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \mid \alpha\text{-L-Rha-(1}\rightarrow 3)]_n$	Formula O25a
O25b	$\begin{array}{c} \beta\text{-Glc}p\text{-} \\ 1 \\ \downarrow \\ 6 \end{array}$ $[\alpha\text{-Rhap-(1}\rightarrow 3)\text{-}\alpha\text{-Glc}p\text{-(1}\rightarrow 3)\text{-}\alpha\text{-Rhap2OAc-(1}\rightarrow 3)\text{-}\beta\text{-Glc}p\text{NAc-}]_n$	Formula O25b

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O26	$[\rightarrow 3)\text{-}\alpha\text{-L-Rha-(1}\rightarrow 4)\text{-}\alpha\text{-L-FucNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow]_n$	Formula O26
O28	$[\rightarrow 2)\text{-(R)-Gro-1-P}\rightarrow 4)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow 3)\text{-}\beta\text{-D-Galf2Ac-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc-(1}\rightarrow]_n$	Formula O28
O36	$\begin{array}{c} [\alpha\text{-L-Rhap-(1}\rightarrow 2)\text{-}\alpha\text{-L-Fucp} \\ 1 \\ \downarrow \\ 3 \\ \rightarrow 4)\text{-}\alpha\text{-D-Manp-(1}\rightarrow 3)\text{-}\alpha\text{-L-Fucp-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcpNAc-(1}\rightarrow]_n \end{array}$	Formula O36
O44	$[\alpha\text{-D-Glc-(1}\rightarrow 4) \mid \rightarrow 6)\text{-}\alpha\text{-D-Man-(1}\rightarrow 2)\text{-}\alpha\text{-D-Man-(1}\rightarrow 2)\text{-}\beta\text{-D-Man-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcNAc(1}\rightarrow]_n$	Formula O44
O45	$[\rightarrow 2)\text{-}\beta\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-6dTal2Ac-(1}\rightarrow 3)\text{-}\alpha\text{-D-FucNAc-(1}\rightarrow]_n$	Formula O45
O45rel	$[\rightarrow 2)\text{-}\beta\text{-D-Glc-(1}\rightarrow 3)\text{-}\alpha\text{-L-6dTal2Ac-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow]_n$	Formula O45rel
O54	$[\rightarrow 4)\text{-}\alpha\text{-d-GalpA-(1}\rightarrow 2)\text{-}\alpha\text{-l-Rhap-(1}\rightarrow 2)\text{-}\beta\text{-d-Ribf-(1}\rightarrow 4)\text{-}\beta\text{-d-Galp-(1}\rightarrow 3)\text{-}\beta\text{-d-GlcpNAc-(1}\rightarrow]_n$	Formula O54
O55	$[\rightarrow 6)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow 3)\text{-}\alpha\text{-D-Gal-(1}\rightarrow 3)\text{-}\beta\text{-D-GalNAc-(1}\rightarrow \mid \alpha\text{-Col-(1}\rightarrow 2)\text{-}\beta\text{-D-Gal-(1}\rightarrow 3)]_n$	Formula O55
O56	$[\rightarrow 7)\text{-}\alpha\text{-Neu5Ac-(2}\rightarrow 3)\text{-}\beta\text{-D-Glc-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow \mid \alpha\text{-D-Gal-(1}\rightarrow 2)]_n$	Formula O56
O57	$\begin{array}{ccc} [\rightarrow 3)\text{-}\alpha\text{-D-Galp-(1}\rightarrow 3)\text{-}\alpha\text{-L-FucpNAc-(1}\rightarrow 3)\text{-}\alpha\text{-D-GlcpNAc-(1}\rightarrow]_n \\ 2 & & 4 \\ \uparrow & & \uparrow \\ 1 & & 1 \\ \alpha\text{-D-GalpA2/3Ac} & & \beta\text{-D-Glcp} \end{array}$	Formula O57
O58	$[3\text{-O-}[(R)\text{-1-carboxyethyl}]\text{-}\alpha\text{-L-Rha-(1}\rightarrow 3) \mid \rightarrow 4)\text{-}\alpha\text{-D-Man-(1}\rightarrow 4)\text{-}\alpha\text{-D-Man2Ac-(1}\rightarrow 3)\text{-}\beta\text{-D-GlcNAc-(1}\rightarrow]_n$	Formula O58

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O64	$[\beta\text{-D-Gal-(1}\rightarrow\text{6)} \mid \rightarrow\text{3)-}\alpha\text{-D-ManNAc-(1}\rightarrow\text{3)-}\beta\text{-D-GlcA-(1}\rightarrow\text{3)-}\beta\text{-D-Gal-(1}\rightarrow\text{3)-}\beta\text{-D-GlcNAc(1}\rightarrow\text{)}_n$	Formula O64
O68	$\begin{array}{ccc} \alpha\text{-L-Rhap} & & \alpha\text{-D-Glcp} \\ 1 & & 1 \\ \downarrow & & \downarrow \\ 3 & & 3 \end{array}$ $\rightarrow\text{6)-}\alpha\text{-D-Manp-(1}\rightarrow\text{2)-}\alpha\text{-D-Manp-(1}\rightarrow\text{2)-}\alpha\text{-D-Manp-(1}\rightarrow\text{2)-}\beta\text{-D-Manp-(1}\rightarrow\text{3)-}\alpha\text{-D-GlcpNAc-(1}\rightarrow\text{)}_n$	Formula O68
O69	$[\rightarrow\text{2)-}\alpha\text{-L-Rha-(1}\rightarrow\text{2)-}\alpha\text{-L-Rha-(1}\rightarrow\text{2)-}\alpha\text{-D-Gal-(1}\rightarrow\text{3)-}\beta\text{-D-GlcNAc-(1}\rightarrow\text{)}_n$	Formula O69
O73 (Strain 73-1)	$[\alpha\text{-D-Glc-(1}\rightarrow\text{3)} \mid \rightarrow\text{4)-}\alpha\text{-D-Man-(1}\rightarrow\text{2)-}\alpha\text{-D-Man-(1}\rightarrow\text{2)-}\beta\text{-D-Man-(1}\rightarrow\text{3)-}\alpha\text{-D-GalNAc(1}\rightarrow\text{)}_n$	Formula O73 (Strain 73-1)
O74	$\begin{array}{c} \rightarrow\text{6)-}\alpha\text{-D-GlcpNAc-(1}\rightarrow\text{4)-}\beta\text{-D-GalpA-(1}\rightarrow\text{3)-}\beta\text{-D-GlcpNAc-(1}\rightarrow\text{)}_n \\ \\ [\beta\text{-D-Fucp3NAc-(1}\rightarrow\text{3)} \end{array}$	Formula O74
O75	$[\beta\text{-D-Man-(1}\rightarrow\text{4)} \mid \rightarrow\text{3)-}\alpha\text{-D-Gal-(1}\rightarrow\text{4)-}\alpha\text{-L-Rha-(1}\rightarrow\text{3)-}\beta\text{-D-GlcNAc-(1}\rightarrow\text{)}_n$	Formula O75
O76	$[\rightarrow\text{4)-}\beta\text{-D-GlcpA-(1}\rightarrow\text{4)-}\beta\text{-D-GalpNAc3Ac-(1}\rightarrow\text{4)-}\alpha\text{-D-GalpNAc-(1}\rightarrow\text{3)-}\beta\text{-D-GalpNAc-(1}\rightarrow\text{)}_n$	Formula O76
O77	$[\rightarrow\text{6)-}\alpha\text{-D-Man-(1}\rightarrow\text{2)-}\alpha\text{-D-Man-(1}\rightarrow\text{2)-}\beta\text{-D-Man-(1}\rightarrow\text{3)-}\alpha\text{-D-GlcNAc(1}\rightarrow\text{)}_n$	Formula O77
O78	$[\rightarrow\text{4)-}\beta\text{-D-GlcNAc-(1}\rightarrow\text{4)-}\beta\text{-D-Man-(1}\rightarrow\text{4)-}\alpha\text{-D-Man-(1}\rightarrow\text{3)-}\beta\text{-D-GlcNAc-(1}\rightarrow\text{)}_n$	Formula O78
O86	$[\alpha\text{-D-Gal-(1}\rightarrow\text{3)} \mid \rightarrow\text{4)-}\alpha\text{-L-Fuc-(1}\rightarrow\text{2)-}\beta\text{-D-Gal-(1}\rightarrow\text{3)-}\alpha\text{-D-GalNAc-(1}\rightarrow\text{3)-}\beta\text{-D-GalNAc-(1}\rightarrow\text{)}_n$	Formula O86

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O88	[α -L-6dTal-(1 $\rightarrow$ 3) $\rightarrow$ 4)- α -D-Man-(1 $\rightarrow$ 3)- α -D-Man-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$) _n	Formula O88
O90	[$\rightarrow$ 4)- α -L-Fuc2/3Ac-(1 $\rightarrow$ 2)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$) _n	Formula O90
O98	[$\rightarrow$ 3)- α -L-QuiNAc-(1 $\rightarrow$ 4)- α -D-GalNAcA-(1 $\rightarrow$ 3)- α -L-QuiNAc-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$) _n	Formula O98
O104	[$\rightarrow$ 4)- α -D-Gal-(1 $\rightarrow$ 4)- α -Neu5,7,9Ac ₃ -(2 $\rightarrow$ 3)- β -D-Gal-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$) _n	Formula O104
O111	[α -Col-(1 $\rightarrow$ 6) $\rightarrow$ 4)- α -D-Glc-(1 $\rightarrow$ 4)- α -D-Gal-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ α -Col-(1 $\rightarrow$ 3)) _n	Formula O111
O113	[$\rightarrow$ 4)- α -D-GalNAc-(1 $\rightarrow$ 4)- α -D-GalA-(1 $\rightarrow$ 3)- α -D-Gal-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ β -D-Gal-(1 $\rightarrow$ 3)) _n	Formula O113
O114	[$\rightarrow$ 4)- β -D-Qui3N(<i>N</i> -acetyl-L-seryl)-(1 $\rightarrow$ 3)- β -D-Ribf-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GlcNAc(1 $\rightarrow$) _n	Formula O114
O119	[β -D-RhaNAc3Nfo-(1 $\rightarrow$ 3) $\rightarrow$ 2)- β -D-Man-(1 $\rightarrow$ 3)- α -D-Gal-(1 $\rightarrow$ 4)- α -L-Rha-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$) _n	Formula O119
O121	[$\rightarrow$ 3)- β -D-Qui4N(<i>N</i> -acetyl-glycyl)-(1 $\rightarrow$ 4)- α -D-GalNAc3AcA6N-(1 $\rightarrow$ 4)- α -D-GalNAcA-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$) _n	Formula O121
O124	[4- <i>O</i> -[(<i>R</i>)-1-carboxyethyl]- β -D-Glc-(1 $\rightarrow$ 6)- α -D-Glc(1 $\rightarrow$ 4) $\rightarrow$ 3)- α -D-Gal-(1 $\rightarrow$ 6)- β -D-Galf-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$) _n	Formula O124
O125	[α -D-Glc-(1 $\rightarrow$ 3) $\rightarrow$ 4)- β -D-GalNAc-(1 $\rightarrow$ 2)- α -D-Man-(1 $\rightarrow$ 3)- α -L-Fuc-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ β -D-Gal-(1 $\rightarrow$ 3)) _n	Formula O125
O126	[$\rightarrow$ 2)- β -D-Man-(1 $\rightarrow$ 3)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ α -L-Fuc-(1 $\rightarrow$ 2)) _n	Formula O126

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O127	[$\rightarrow 2$)- α -L-Fuc-(1 $\rightarrow$ 2)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$) _n	Formula O127
O128	[α -L-Fuc-(1 $\rightarrow$ 2) $\rightarrow 6$)- β -D-Gal-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$ 4)- α -D-Gal-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$) _n	Formula O128
O136	[$\rightarrow 4$)- β -Pse5Ac7Ac-(2 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 4)- β -D-GlcNAc-(1 $\rightarrow$ β -Pse5Ac7Ac=5,7-diacetamido-3,5,7,9-tetradexy-L- <i>glycero</i> - β -L- <i>manno</i> -nonulosonic acid) _n	Formula O136
O138	[$\rightarrow 2$)- α -L-Rha-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 4)- α -D-GalNAcA-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$) _n	Formula O138
O140	$\begin{array}{c} [\alpha\text{-D-Galf-(1}\rightarrow\text{2)-}\alpha\text{-L-Rhap} \\ 1 \\ \downarrow \\ 4 \\ \rightarrow\text{3)-}\beta\text{-D-Galp-(1}\rightarrow\text{4)-}\alpha\text{-D-Glcp-(1}\rightarrow\text{4)-}\beta\text{-D-GlcpA-(1}\rightarrow\text{3)-} \\ \beta\text{-D-GalpNAc-(1}\rightarrow\text{)}_n \end{array}$	Formula O140
O141	[α -L-Rha-(1 $\rightarrow$ 3) $\rightarrow 4$)- α -D-Man-(1 $\rightarrow$ 3)- α -D-Man6Ac-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ β -D-GlcA-(1 $\rightarrow$ 2)) _n	Formula O141
O142	[$\rightarrow 2$)- α -L-Rha-(1 $\rightarrow$ 6)- α -D-GalNAc-(1 $\rightarrow$ 4)- α -D-GalNAc-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ β -D-GlcNAc-(1 $\rightarrow$ 3)) _n	Formula O142
O143	[$\rightarrow 2$)- β -D-GalA6R3,4Ac-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ 4)- β -D-GlcA-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ R=1,3-dihydroxy-2-propylamino) _n	Formula O143
O147	[$\rightarrow 2$)- α -L-Rha-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 4)- β -D-GalA-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$) _n	Formula O147
O149	[$\rightarrow 3$)- β -D-GlcNAc-(S)-4,6Py-(1 $\rightarrow$ 3)- β -L-Rha-(1 $\rightarrow$ 4)- β -D-GlcNAc-(1 $\rightarrow$ (S)-4,6Py=4,6-O-[(S)-1-carboxyethylidene]-) _n	Formula O149

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O152	[β -L-Rha-(1 $\rightarrow$ 4) $\rightarrow$ 3)- α -D-GlcNAc-(1- $P\rightarrow$ 6)- α -D-Glc-(1 $\rightarrow$ 2)- β -D-Glc-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O152
O157	[$\rightarrow$ 2)- α -D-Rha4NAc-(1 $\rightarrow$ 3)- α -L-Fuc-(1 $\rightarrow$ 4)- β -D-Glc-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$)] _n	Formula O157
O158	[α -D-Glc-(1 $\rightarrow$ 6) $\rightarrow$ 4)- α -D-Glc-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$ α -L-Rha-(1 $\rightarrow$ 3)] _n	Formula O158
O159	[α -L-Fuc-(1 $\rightarrow$ 4) $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ 4)- α -D-GalA-(1 $\rightarrow$ 3)- α -L-Fuc-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O159
O164	[β -D-Glc-(1 $\rightarrow$ 6)- α -D-Glc(1 $\rightarrow$ 4) $\rightarrow$ 3)- β -D-Gal-(1 $\rightarrow$ 6)- β -D-Galf-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$)] _n	Formula O164
O173	[α -L-Fuc-(1 $\rightarrow$ 4) $\rightarrow$ 3)- α -D-Glc-(1- $P\rightarrow$ 6)- α -D-Glc-(1 $\rightarrow$ 2)- β -D-Glc-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O173
62D ₁ Suggested as <i>Erwinia</i> <i>herbicola</i>	[α -D-Gal(1 $\rightarrow$ 6) $\rightarrow$ 2)- β -D-Qui3NAc-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 3)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-FucNAc-(1 $\rightarrow$)] _n	Formula 62D ₁
O22	[$\rightarrow$ 6)- α -D-Glc-(1 $\rightarrow$ 4)- β -D-GlcA-(1 $\rightarrow$ 4)- β -D-GalNAc3Ac-(1 $\rightarrow$ 3)- α -D-Gal-(1 $\rightarrow$ 3)- β -D-GalNAc-(1 $\rightarrow$)] _n	Formula O22
O35	[$\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ α -D-GalNAcA6N-(1 $\rightarrow$ 2)] _n	Formula O35
O65	[$\rightarrow$ 2)- β -D-Qui3NAc-(1 $\rightarrow$ 4)- α -D-GalA6N-(1 $\rightarrow$ 4)- α -D-GalNAc-(1 $\rightarrow$ 4)- β -D-GalA-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O65
O66	[$\rightarrow$ 2)- β -D-Man-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$ 2)- β -D-Glc3Ac-(1 $\rightarrow$ 3)- α -L-6dTal-(1 $\rightarrow$ 3)- α -D-GlcNAc(1 $\rightarrow$)] _n	Formula O66

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O83	[$\rightarrow 6$)- α -D-Glc-(1 $\rightarrow$ 4)- β -D-GlcA-(1 $\rightarrow$ 6)- β -D-Gal-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 4)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O83
O91	[$\rightarrow 4$)- α -D-Qui3NAcyl-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 4)- β -D-GlcNAc-(1 $\rightarrow$ 4)- β -D-GlcA6NGly-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$ Acyl=(R)-3-hydroxybutyryl)] _n	Formula O91
O105	[β -D-Ribf-(1 $\rightarrow$ 3) $\rightarrow 4$)- α -D-GlcA2Ac3Ac-(1 $\rightarrow$ 2)- α -L-Rha4Ac-(1 $\rightarrow$ 3)- β -L-Rha-(1 $\rightarrow$ 4)- β -L-Rha-(1 $\rightarrow$ 3)- β -D-GlcNAc6Ac-(1 $\rightarrow$)] _n	Formula O105
O116	[$\rightarrow 2$)- β -D-Qui4NAc-(1 $\rightarrow$ 6)- α -D-GlcNAc-(1 $\rightarrow$ 4)- α -D-GalNAc-(1 $\rightarrow$ 4)- α -D-GalA-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O116
O117	[$\rightarrow 4$)- β -D-GalNAc-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 4)- α -D-Glc-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GalNAc-(1 $\rightarrow$)] _n	Formula O117
O139	[β -D-Glc-(1 $\rightarrow$ 3) $\rightarrow 3$)- α -L-Rha-(1 $\rightarrow$ 4)- α -D-GalA-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 2)- α -L-Rha-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O139
O153	[$\rightarrow 2$)- β -D-Ribf-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 4)- α -D-GlcNAc-(1 $\rightarrow$ 4)- β -D-Gal-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O153
O167	[α -D-Galf-(1 $\rightarrow$ 4) $\rightarrow 2$)- β -D-GalA6N(L)Ala-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$ 2)- β -D-Galf-(1 $\rightarrow$ 5)- β -D-Galf-(1 $\rightarrow$ 3)- β -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O167
O172	[$\rightarrow 3$)- α -L-FucNAc-(1 $\rightarrow$ 4)- α -D-Glc6Ac-(1 $\rightarrow$ P $\rightarrow$ 4)- α -D-Glc-(1 $\rightarrow$ 3)- α -L-FucNAc-(1 $\rightarrow$ 3)- α -D-GlcNAc-(1 $\rightarrow$)] _n	Formula O172
O8	[$\rightarrow 2$)- α -D-Man-(1 $\rightarrow$ 2)- α -D-Man-(1 $\rightarrow$ 3)- β -D-Man-(1 $\rightarrow$)] _n	Formula O8
O9a	[$\rightarrow 2$)- α -D-Man-(1 $\rightarrow$ 2)- α -D-Man-(1 $\rightarrow$ 3)- α -D-Man-(1 $\rightarrow$ 3)- α -D-Man-(1 $\rightarrow$)] _n	Formula O9a
O9	[$\rightarrow 2$)-[α -D-Man-(1 $\rightarrow$ 2)] ₂ - α -D-Man-(1 $\rightarrow$ 3)- α -D-Man-(1 $\rightarrow$ 3)- α -D-Man-(1 $\rightarrow$)] _n	Formula O9

Serogroup/ Serotype	Moiety Structure (O-unit)	Moiety structure referred to herein as:
O20ab	[$\rightarrow$ 2)- β -D-Ribf-(1 $\rightarrow$ 4)- α -D-Gal-(1 $\rightarrow$) _n	Formula O20ab
O20ac	[α -D-Gal-(1 $\rightarrow$ 3) $\rightarrow$ 2)- β -D-Ribf-(1 $\rightarrow$ 4)- α -D-Gal-(1 $\rightarrow$) _n	Formula O20ac
O52	[$\rightarrow$ 3)- β -D-Fucf-(1 $\rightarrow$ 3)- β -D-6dmanHep2Ac-(1 $\rightarrow$) _n	Formula O52
O97	[$\rightarrow$ 3)- α -L-Rha-(1 $\rightarrow$ 3)- β -L-Rha-(1 $\rightarrow$ β -D-Xulf-(2 $\rightarrow$ 2) β -D-Xulf-(2 $\rightarrow$ 2)] _n	Formula O97

† β -D-6dmanHep2Ac is 2-O-acetyl-6-deoxy- β -D-manno-heptopyranosyl.

‡ β -D-Xulf is β -D-threo-pentofuranosyl.

5 D. Core Oligosaccharide

The core oligosaccharide is positioned between Lipid A and the O-antigen outer region in wild-type *E. coli* LPS. More specifically, the core oligosaccharide is the part of the polysaccharide that includes the bond between the O-antigen and the lipid A in wild type *E. coli*. This bond includes a ketosidic bond between the hemiketal function of the innermost 3-deoxy-d-manno-oct-2-ulosonic acid (KDO)) residue and a hydroxyl-group of a GlcNAc-residue of the lipid A. The core oligosaccharide region shows a high degree of similarity among wild-type *E. coli* strains. It usually includes a limited number of sugars. The core oligosaccharide includes an inner core region and an outer core region.

More specifically, the inner core is composed primarily of L-glycero-D-manno-heptose (heptose) and KDO residues. The inner core is highly conserved. A KDO residue includes the following Formula KDO:



The outer region of the core oligosaccharide displays more variation than the inner core region, and differences in this region distinguish the five chemotypes in *E. coli*: R1, R2, R3, R4, and K-12. See FIG. 24, which illustrates generalized structures of the carbohydrate backbone of the outer core oligosaccharides of the five known chemotypes. HepII is the last residue of the inner core oligosaccharide. While all of the outer core oligosaccharides share a structural

theme, with a (hexose)₃ carbohydrate backbone and two side chain residues, the order of hexoses in the backbone and the nature, position, and linkage of the side chain residues can all vary. The structures for the R1 and R4 outer core oligosaccharides are highly similar, differing in only a single β -linked residue.

5 The core oligosaccharides of wild-type *E. coli* are categorized in the art based on the structures of the distal oligosaccharide, into five different chemotypes: *E. coli* R1, *E. coli* R2, *E. coli* R3, *E. coli* R4, and *E. coli* K12.

In a preferred embodiment, the compositions described herein include glycoconjugates in which the O-polysaccharide includes a core oligosaccharide bound to the O-antigen. In one embodiment, the composition induces an immune response
10 against at least any one of the core *E. coli* chemotypes *E. coli* R1, *E. coli* R2, *E. coli* R3, *E. coli* R4, and *E. coli* K12. In another embodiment, the composition induces an immune response against at least two core *E. coli* chemotypes. In another embodiment, the composition induces an immune response against at least three core *E. coli*
15 chemotypes. In another embodiment, the composition induces an immune response against at least four core *E. coli* chemotypes. In another embodiment, the composition induces an immune response against all five core *E. coli* chemotypes.

In another preferred embodiment, the compositions described herein include glycoconjugates in which the O-polysaccharide does not include a core oligosaccharide bound to the O-antigen. In one embodiment, such a composition induces an immune
20 response against at least any one of the core *E. coli* chemotypes *E. coli* R1, *E. coli* R2, *E. coli* R3, *E. coli* R4, and *E. coli* K12, despite the glycoconjugate having an O-polysaccharide that does not include a core oligosaccharide.

E. coli serotypes may be characterized according to one of the five chemotypes.
25 **Table 2** lists exemplary serotypes characterized according to chemotype. The serotypes in bold represent the serotypes that are most commonly associated with the indicated core chemotype. Accordingly, in a preferred embodiment, the composition induces an immune response against at least any one of the core *E. coli* chemotypes *E. coli* R1, *E. coli* R2, *E. coli* R3, *E. coli* R4, and *E. coli* K12, which includes an immune
30 response against any one of the respective corresponding *E. coli* serotypes.

Table 2: Core Chemotype and associated *E. coli* Serotype

Core chemotype	Serotype
R1	O25a, O6, O2 , O1 , O75, O4, O16 , O8, O18, O9, O13, O20, O21, O91, and O163.

R2	O21, O44, O11, O89, O162, O9
R3	O25b, O15, O153, O21, O17, O11, O159, O22 O86, O93
R4	O2, O1, O86, O7, O102, O160, O166
K-12	O25b , O16

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an R1 chemotype, e.g., selected from a saccharide having Formula O25a, Formula O6, Formula O2, Formula O1, Formula O75, Formula O4, Formula O16, Formula O8, Formula O18, Formula O9, Formula O13, Formula O20, Formula O21, Formula O91, and Formula O163, wherein n is 1 to 100. In some embodiments, the saccharide in said composition further includes an *E. coli* R1 core moiety, e.g., shown in **FIG. 24**.

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an R1 chemotype, e.g., selected from a saccharide having Formula O25a, Formula O6, Formula O2, Formula O1, Formula O75, Formula O4, Formula O16, Formula O18, Formula O13, Formula O20, Formula O21, Formula O91, and Formula O163, wherein n is 1 to 100, preferably 31 to 100, preferably from 31 to 90, more preferably 35 to 90, most preferably 35 to 65. In some embodiments, the saccharide in said composition further includes an *E. coli* R1 core moiety in the saccharide.

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an R2 chemotype, e.g., selected from a saccharide having Formula O21, Formula O44, Formula O11, Formula O89, Formula O162, and Formula O9, wherein n is 1 to 100, preferably 31 to 100, preferably from 31 to 90, more preferably 35 to 90, most preferably 35 to 65. In some embodiments, the saccharide in said composition further includes an *E. coli* R2 core moiety, e.g., shown in **FIG. 24**.

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an R3 chemotype, e.g., selected from a saccharide having Formula O25b, Formula O15, Formula O153, Formula O21, Formula O17, Formula O11, Formula O159, Formula O22, Formula O86, and Formula O93, wherein n is 1 to 100, preferably 31 to 100, preferably from 31 to 90, more preferably 35 to 90, most preferably 35 to 65. In some embodiments, the saccharide in said composition further includes an *E. coli* R3 core moiety, e.g., shown in **FIG. 24**.

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an R4 chemotype, e.g., selected from a saccharide having Formula O2, Formula O1, Formula O86, Formula O7, Formula O102, Formula O160, and Formula O166, wherein n is 1 to 100, preferably 31 to 100, preferably from 31 to 90, more

preferably 35 to 90, most preferably 35 to 65. In some embodiments, the saccharide in said composition further includes an *E. coli* R4 core moiety, e.g., shown in FIG. 24.

In some embodiments, the composition includes a saccharide that includes a structure derived from a serotype having an K-12 chemotype (e.g., selected from a saccharide having Formula O25b and a saccharide having Formula O16), wherein n is 1 to 1000, preferably 31 to 100, preferably from 31 to 90, more preferably 35 to 90, most preferably 35 to 65. In some embodiments, the saccharide in said composition further includes an *E. coli* K-12 core moiety, e.g., shown in FIG. 24.

In some embodiments, the saccharide includes the core saccharide. Accordingly, in one embodiment, the O-polysaccharide further includes an *E. coli* R1 core moiety. In another embodiment, the O-polysaccharide further includes an *E. coli* R2 core moiety. In another embodiment, the O-polysaccharide further includes an *E. coli* R3 core moiety. In another embodiment, the O-polysaccharide further includes an *E. coli* R4 core moiety. In another embodiment, the O-polysaccharide further includes an *E. coli* K12 core moiety.

In some embodiments, the saccharide does not include the core saccharide. Accordingly, in one embodiment, the O-polysaccharide does not include an *E. coli* R1 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R2 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R3 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* R4 core moiety. In another embodiment, the O-polysaccharide does not include an *E. coli* K12 core moiety.

E. Conjugated O-Antigens

Chemical linkage of O-antigens or preferably O-polysaccharides to protein carriers may improve the immunogenicity of the O-antigens or O-polysaccharides. However, variability in polymer size represents a practical challenge for production. In commercial use, the size of the saccharide can influence the compatibility with different conjugation synthesis strategies, product uniformity, and conjugate immunogenicity. Controlling the expression of a Wzz family protein chain length regulator through manipulation of the O- antigen synthesis pathway allows for production of a desired length of O-antigen chains in a variety of Gram-negative bacterial strains, including *E. coli*.

In one embodiment, the purified saccharides are chemically activated to produce activated saccharides capable of reacting with the carrier protein. Once activated, each saccharide is separately conjugated to a carrier protein to form a conjugate, namely a

glycoconjugate. As used herein, the term "glycoconjugate" refers to a saccharide covalently linked to a carrier protein. In one embodiment a saccharide is linked directly to a carrier protein. In another embodiment, a saccharide is linked to a protein through a spacer/linker.

Conjugates may be prepared by schemes that bind the carrier to the O-antigen at one or at
5 multiple sites along the O-antigen, or by schemes that activate at least one residue of the core oligosaccharide.

In one embodiment, each saccharide is conjugated to the same carrier protein.
If the protein carrier is the same for 2 or more saccharides in the composition, the saccharides
may be conjugated to the same molecule of the carrier protein (e.g., carrier molecules having 2
10 or more different saccharides conjugated to it).

In a preferred embodiment, the saccharides are each individually conjugated to different
molecules of the protein carrier (each molecule of protein carrier only having one type of
saccharide conjugated to it). In said embodiment, the saccharides are said to be individually
conjugated to the carrier protein.

15 The chemical activation of the saccharides and subsequent conjugation to the carrier
protein can be achieved by the activation and conjugation methods disclosed herein. After
conjugation of the polysaccharide to the carrier protein, the glycoconjugates are purified
(enriched with respect to the amount of polysaccharide- protein conjugate) by a variety of
techniques. These techniques include concentration/diafiltration operations,
20 precipitation/elution, column chromatography, and depth filtration. After the individual
glycoconjugates are purified, they are compounded to formulate the immunogenic composition
of the present invention.

Activation. The present invention further relates to activated polysaccharides produced
from any of the embodiments described herein wherein the polysaccharide is activated with a
25 chemical reagent to produce reactive groups for conjugation to a linker or carrier protein. In
some embodiments, the saccharide of the invention is activated prior to conjugation to the
carrier protein. In some embodiments, the degree of activation does not significantly reduce the
molecular weight of the polysaccharide. For example, in some embodiments, the degree of
activation does not cleave the polysaccharide backbone. In some embodiments, the degree of
30 activation does not significantly impact the degree of conjugation, as measured by the number
of lysine residues modified in the carrier protein, such as, CRM₁₉₇ (as determined by amino acid
analysis). For example, in some embodiments, the degree of activation does not significantly
increase the number of lysine residues modified (as determined by amino acid analysis) in the
carrier protein by 3-fold, as compared to the number of lysine residues modified in the carrier
35 protein of a conjugate with a reference polysaccharide at the same degree of activation. In
some embodiments, the degree of activation does not increase the level of unconjugated free

saccharide. In some embodiments, the degree of activation does not decrease the optimal saccharide/protein ratio.

In some embodiments, the activated saccharide has a percentage of activation wherein moles of thiol per saccharide repeat unit of the activated saccharide is between 1-100%,
 5 such as, for example, between 2-80%, between 2-50%, between 3-30%, and between 4-25%. The degree of activation is at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, $\geq 20\%$, $\geq 30\%$, $\geq 40\%$, $\geq 50\%$, $\geq 60\%$, $\geq 70\%$, $\geq 80\%$, or $\geq 90\%$, or about 100%. Preferably, the degree of activation is at most 50%, more preferably at most 25%. In one embodiment, the degree of
 10 activation is at most 20%. Any minimum value and any maximum value may be combined to define a range.

In one embodiment, the polysaccharide is activated with 1 -cyano-4-dimethylamino pyridinium tetrafluoroborate (CDAP) to form a cyanate ester. The activated polysaccharide is then coupled directly or via a spacer (linker) group to an
 15 amino group on the carrier protein (preferably CRM₁₉₇ or tetanus toxoid).

For example, the spacer may be cystamine or cysteamine to give a thiolated polysaccharide which could be coupled to the carrier via a thioether linkage obtained after reaction with a maleimide-activated carrier protein (for example using N-[Y-maleimidobutyroxy]succinimide ester (GMBS)) or a haloacetylated carrier protein (for
 20 example using iodoacetamide, N-succinimidyl bromoacetate (SBA; SIB), N-succinimidyl(4-iodoacetyl)aminobenzoate (SIAB), sulfosuccinimidyl(4-iodoacetyl)aminobenzoate (sulfo-SIAB), N-succinimidyl iodoacetate (SIA), or succinimidyl 3-[bromoacetamido]propionate (SBAP)). In one embodiment, the cyanate ester (optionally made by CDAP chemistry) is coupled with hexane diamine or adipic
 25 acid dihydrazide (ADH) and the amino-derivatised saccharide is conjugated to the carrier protein (e.g., CRM₁₉₇) using carbodiimide (e.g., EDAC or EDC) chemistry via a carboxyl group on the protein carrier.

Other suitable techniques for conjugation use carbodiimides, hydrazides, active esters, norborane, p-nitrobenzoic acid, N-hydroxysuccinimide, S-NHS, EDC, TSTU.
 30 Conjugation may involve a carbonyl linker which may be formed by reaction of a free hydroxyl group of the saccharide with CDI followed by reaction with a protein to form a carbamate linkage. This may involve reduction of the anomeric terminus to a primary hydroxyl group, optional protection/deprotection of the primary hydroxyl group, reaction of the primary hydroxyl group with CDI to form a CDI carbamate intermediate and
 35 coupling the CDI carbamate intermediate with an amino group on a protein (CDI chemistry).

Molecular weight. In some embodiments, the glycoconjugate comprises a saccharide having a molecular weight of between 10 kDa and 2,000 kDa. In other embodiments, the saccharide has a molecular weight of between 50 kDa and 1,000 kDa. In other embodiments, the saccharide has a molecular weight of between 70 kDa and 900 kDa. In other embodiments, the saccharide has a molecular weight of between 100 kDa and 800 kDa. In other embodiments, the saccharide has a molecular weight of between 200 kDa and 600 kDa. In further embodiments, the saccharide has a molecular weight of 100 kDa to 1000 kDa; 100 kDa to 900 kDa; 100 kDa to 800 kDa; 100 kDa to 700 kDa; 100 kDa to 600 kDa; 100 kDa to 500 kDa; 100 kDa to 400 kDa; 100 kDa to 300 kDa; 150 kDa to 1,000 kDa; 150 kDa to 900 kDa; 150 kDa to 800 kDa; 150 kDa to 700 kDa; 150 kDa to 600 kDa; 150 kDa to 500 kDa; 150 kDa to 400 kDa; 150 kDa to 300 kDa; 200 kDa to 1,000 kDa; 200 kDa to 900 kDa; 200 kDa to 800 kDa; 200 kDa to 700 kDa; 200 kDa to 600 kDa; 200 kDa to 500 kDa; 200 kDa to 400 kDa; 200 kDa to 300; 250 kDa to 1,000 kDa; 250 kDa to 900 kDa; 250 kDa to 800 kDa; 250 kDa to 700 kDa; 250 kDa to 600 kDa; 250 kDa to 500 kDa; 250 kDa to 400 kDa; 250 kDa to 350 kDa; 300 kDa to 1,000 kDa; 300 kDa to 900 kDa; 300 kDa to 800 kDa; 300 kDa to 700 kDa; 300 kDa to 600 kDa; 300 kDa to 500 kDa; 300 kDa to 400 kDa; 400 kDa to 1,000 kDa; 400 kDa to 900 kDa; 400 kDa to 800 kDa; 400 kDa to 700 kDa; 400 kDa to 600 kDa; 500 kDa to 600 kDa. In one embodiment, the glycoconjugate having such a molecular weight is produced by single-end conjugation. In another embodiment, the glycoconjugate having such a molecular weight is produced by reductive amination chemistry (RAC) prepared in aqueous buffer. Any whole number integer within any of the above ranges is contemplated as an embodiment of the disclosure.

In some embodiments, the glycoconjugate of the invention has a molecular weight of between 400 kDa and 15,000 kDa; between 500 kDa and 10,000 kDa; between 2,000 kDa and 10,000 kDa; between 3,000 kDa and 8,000 kDa; or between 3,000 kDa and 5,000 kDa. In other embodiments, the glycoconjugate has a molecular weight of between 500 kDa and 10,000 kDa. In other embodiments, glycoconjugate has a molecular weight of between 1,000 kDa and 8,000 kDa. In still other embodiments, the glycoconjugate has a molecular weight of between 2,000 kDa and 8,000 kDa or between 3,000 kDa and 7,000 kDa. In further embodiments, the glycoconjugate of the invention has a molecular weight of between 200 kDa and 20,000 kDa; between 200 kDa and 15,000 kDa; between 200 kDa and 10,000 kDa; between 200 kDa and 7,500 kDa; between 200 kDa and 5,000 kDa; between 200 kDa and 3,000 kDa; between 200 kDa and 1,000 kDa; between 500 kDa and 20,000 kDa; between 500 kDa and 15,000 kDa; between 500 kDa and 12,500 kDa; between 500 kDa and 10,000 kDa; between 500 kDa and 7,500 kDa; between 500 kDa and 6,000 kDa; between 500 kDa and 5,000 kDa; between 500 kDa and 4,000 kDa; between 500 kDa and 3,000 kDa; between 500 kDa and 2,000 kDa;

between 500 kDa and 1,500 kDa; between 500 kDa and 1,000 kDa; between 750 kDa and 20,000 kDa; between 750 kDa and 15,000 kDa; between 750 kDa and 12,500 kDa; between 750 kDa and 10,000 kDa; between 750 kDa and 7,500 kDa; between 750 kDa and 6,000 kDa; between 750 kDa and 5,000 kDa; between 750 kDa and 4,000 kDa;

5 between 750 kDa and 3,000 kDa; between 750 kDa and 2,000 kDa; between 750 kDa and 1,500 kDa; between 1,000 kDa and 15,000 kDa; between 1,000 kDa and 12,500 kDa; between 1,000 kDa and 10,000 kDa; between 1,000 kDa and 7,500 kDa; between 1,000 kDa and 6,000 kDa; between 1,000 kDa and 5,000 kDa; between 1,000 kDa and 4,000 kDa; between 1,000 kDa and 2,500 kDa; between 2,000 kDa and 15,000 kDa;

10 between 2,000 kDa and 12,500 kDa; between 2,000 kDa and 10,000 kDa; between 2,000 kDa and 7,500 kDa; between 2,000 kDa and 6,000 kDa; between 2,000 kDa and 5,000 kDa; between 2,000 kDa and 4,000 kDa; or between 2,000 kDa and 3,000 kDa.

In one embodiment, the glycoconjugate having such a molecular weight is produced by eTEC conjugation described herein. In another embodiment, the glycoconjugate having

15 such a molecular weight is produced by reductive amination chemistry (RAC). In another embodiment, the glycoconjugate having such a molecular weight is produced by reductive amination chemistry (RAC) prepared in DMSO.

In further embodiments, the glycoconjugate of the invention has a molecular weight of between 1,000 kDa and 20,000 kDa; between 1,000 kDa and 15,000 kDa;

20 between 2,000 kDa and 10,000 kDa; between 2,000 kDa and 7,500 kDa; between 2,000 kDa and 5,000 kDa; between 3,000 kDa and 20,000 kDa; between 3,000 kDa and 15,000 kDa; between 3,000 kDa and 12,500 kDa; between 4,000 kDa and 10,000 kDa; between 4,000 kDa and 7,500 kDa; between 4,000 kDa and 6,000 kDa; or between 5,000 kDa and 7,000 kDa. In one embodiment, the glycoconjugate having such a

25 molecular weight is produced by reductive amination chemistry (RAC). In another embodiment, the glycoconjugate having such a molecular weight is produced by reductive amination chemistry (RAC) prepared in DMSO. In another embodiment, the glycoconjugate having such a molecular weight is produced by eTEC conjugation described herein.

30 In further embodiments, the glycoconjugate of the invention has a molecular weight of between 5,000 kDa and 20,000 kDa; between 5,000 kDa and 15,000 kDa; between 5,000 kDa and 10,000 kDa; between 5,000 kDa and 7,500 kDa; between 6,000 kDa and 20,000 kDa; between 6,000 kDa and 15,000 kDa; between 6,000 kDa and 12,500 kDa; between 6,000 kDa and 10,000 kDa or between 6,000 kDa and 7,500 kDa.

35 The molecular weight of the glycoconjugate may be measured by SEC-MALLS. Any whole number integer within any of the above ranges is contemplated as an

embodiment of the disclosure. The glycoconjugates of the invention may also be characterized by the ratio (weight/weight) of saccharide to carrier protein. In some embodiments, the ratio of polysaccharide to carrier protein in the glycoconjugate (w/w) is between 0.5 and 3 (e.g., about 0.5, about 0.6, about 0.7, about 0.8, about 0.9, about 1.0, about 1.1, about 1.2, about 1.3, about 1.4, about 1.5, about 1.6, about 1.7, about 1.8, about 1.9, about 2.0, about 2.1, about 2.2, about 2.3, about 2.4, about 2.5, about 2.6, about 2.7, about 2.8, about 2.9, or about 3.0). In other embodiments, the saccharide to carrier protein ratio (w/w) is between 0.5 and 2.0, between 0.5 and 1.5, between 0.8 and 1.2, between 0.5 and 1.0, between 1.0 and 1.5 or between 1.0 and 2.0. In further embodiments, the saccharide to carrier protein ratio (w/w) is between 0.8 and 1.2. In a preferred embodiment, the ratio of polysaccharide to carrier protein in the conjugate is between 0.9 and 1.1. In some such embodiments, the carrier protein is CRM₁₉₇.

The glycoconjugates may also be characterized by their molecular size distribution (K_d). Size exclusion chromatography media (CL-4B) can be used to determine the relative molecular size distribution of the conjugate. Size Exclusion Chromatography (SEC) is used in gravity fed columns to profile the molecular size distribution of conjugates. Large molecules excluded from the pores in the media elute more quickly than small molecules. Fraction collectors are used to collect the column eluate. The fractions are tested colorimetrically by saccharide assay. For the determination of K_d , columns are calibrated to establish the fraction at which molecules are fully excluded (V_0), ($K_d=0$), and the fraction representing the maximum retention (V_i), ($K_d=1$). The fraction at which a specified sample attribute is reached (V_e), is related to K_d by the expression, $K_d = (V_e - V_0) / (V_i - V_0)$.

Free saccharide. The glycoconjugates and immunogenic compositions of the invention may include free saccharide that is not covalently conjugated to the carrier protein, but is nevertheless present in the glycoconjugate composition. The free saccharide may be non-covalently associated with (i.e., non-covalently bound to, adsorbed to, or entrapped in or with) the glycoconjugate. In a preferred embodiment, the glycoconjugate comprises at most 50%, 45%, 40%, 35%, 30%, 25%, 20% or 15% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the glycoconjugate comprises less than about 25% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the glycoconjugate comprises at most about 20% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the glycoconjugate comprises at most about 15% of free polysaccharide compared to the total amount of polysaccharide. In another preferred embodiment, the glycoconjugate comprises at most about 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the

glycoconjugate comprises less than about 8% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the glycoconjugate comprises at most about 6% of free polysaccharide compared to the total amount of polysaccharide. In a preferred embodiment the glycoconjugate comprises at most about 5% of free polysaccharide compared to the total amount of polysaccharide. See, for example, **Table 19, Table 20, Table 21, Table 22, Table 23, Table 24, and Table 25.**

Covalent linkage. In other embodiments, the conjugate comprises at least one covalent linkage between the carrier protein and saccharide for every 5 to 10 saccharide repeat units; every 2 to 7 saccharide repeat units; every 3 to 8 saccharide repeat units; every 4 to 9 saccharide repeat units; every 6 to 11 saccharide repeat units; every 7 to 12 saccharide repeat units; every 8 to 13 saccharide repeat units; every 9 to 14 saccharide repeat units; every 10 to 15 saccharide repeat units; every 2 to 6 saccharide repeat units; every 3 to 7 saccharide repeat units; every 4 to 8 saccharide repeat units; every 6 to 10 saccharide repeat units; every 7 to 11 saccharide repeat units; every 8 to 12 saccharide repeat units; every 9 to 13 saccharide repeat units; every 10 to 14 saccharide repeat units; every 10 to 20 saccharide repeat units; every 4 to 25 saccharide repeat units or every 2 to 25 saccharide repeat units. In frequent embodiments, the carrier protein is CRM₁₉₇. In another embodiment, at least one linkage between carrier protein and saccharide occurs for every 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 saccharide repeat units of the polysaccharide. In one embodiment, the carrier protein is CRM₁₉₇. Any whole number integer within any of the above ranges is contemplated as an embodiment of the disclosure.

Lysine residues. Another way to characterize the glycoconjugates of the invention is by the number of lysine residues in the carrier protein (e.g., CRM₁₉₇) that become conjugated to the saccharide which can be characterized as a range of conjugated lysines (degree of conjugation). The evidence for lysine modification of the carrier protein, due to covalent linkages to the polysaccharides, can be obtained by amino acid analysis using routine methods known to those of skill in the art. Conjugation results in a reduction in the number of lysine residues recovered, compared to the carrier protein starting material used to generate the conjugate materials. In a preferred embodiment, the degree of conjugation of the glycoconjugate of the invention is between 2 and 15, between 2 and 13, between 2 and 10, between 2 and 8, between 2 and 6, between 2 and 5, between 2 and 4, between 3 and 15, between 3 and 13, between 3 and 10, between 3 and 8, between 3 and 6, between 3 and 5, between 3 and 4, between 5 and 15, between 5 and 10, between 8 and 15, between 8 and 12, between

10 and 15 or between 10 and 12. In one embodiment, the degree of conjugation of the glycoconjugate of the invention is about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14 or about 15. In a preferred embodiment, the degree of conjugation of the glycoconjugate of the invention
 5 is between 4 and 7. In some such embodiments, the carrier protein is CRM₁₉₇.

The frequency of attachment of the saccharide chain to a lysine on the carrier protein is another parameter for characterizing the glycoconjugates of the invention. For example, in some embodiments, at least one covalent linkage between the carrier protein and the polysaccharide for every 4 saccharide repeat units of the polysaccharide. In another
 10 embodiment, the covalent linkage between the carrier protein and the polysaccharide occurs at least once in every 10 saccharide repeat units of the polysaccharide. In another embodiment, the covalent linkage between the carrier protein and the polysaccharide occurs at least once in every 15 saccharide repeat units of the polysaccharide. In a further embodiment, the covalent
 15 linkage between the carrier protein and the polysaccharide occurs at least once in every 25 saccharide repeat units of the polysaccharide.

O-acetylation. In some embodiments, the saccharides of the invention are O-acetylated. In some embodiments, the glycoconjugate comprises a saccharide which has a degree of O-acetylation of between 10-100%, between 20-100%, between 30-100%, between 40-100%, between 50-100%, between 60-100%, between 70-100%, between 75-100%, 80-
 20 100%, 90-100%, 50-90%, 60-90%, 70-90% or 80-90%. In other embodiments, the degree of O-acetylation is $\geq 10\%$, $\geq 20\%$, $\geq 30\%$, $\geq 40\%$, $\geq 50\%$, $\geq 60\%$, $\geq 70\%$, $\geq 80\%$, or $\geq 90\%$, or about 100%. By % of O-acetylation it is meant the percentage of a given saccharide relative to 100% (where each repeat unit is fully acetylated relative to its acetylated structure).

In some embodiments, the glycoconjugate is prepared by reductive amination. In some
 25 embodiments, the glycoconjugate is a single-end-linked conjugated saccharide, wherein the saccharide is covalently bound to a carrier protein directly. In some embodiments, the glycoconjugate is covalently bound to a carrier protein through a (2-((2-oxoethyl)thio)ethyl) carbamate (eTEC) spacer.

REDUCTIVE AMINATION. In one embodiment, the saccharide is conjugated to the
 30 carrier protein by reductive amination (such as described in U.S. Patent Appl. Pub. Nos. 2006/0228380, 2007/0231340, 2007/0184071 and 2007/0184072, WO 2006/110381, WO 2008/079653, and WO 2008/143709).

Reductive amination includes (1) oxidation of the saccharide, (2) reduction of the activated saccharide and a carrier protein to form a conjugate. Before oxidation, the saccharide
 35 is optionally hydrolyzed. Mechanical or chemical hydrolysis may be employed. Chemical hydrolysis may be conducted using acetic acid.

The oxidation step may involve reaction with periodate. The term "periodate" as used herein refers to both periodate and periodic acid. The term also includes both metaperiodate (IO_4^-) and orthoperiodate (IO_6^{5-}) and the various salts of periodate (e.g., sodium periodate and potassium periodate). In one embodiment the polysaccharide is

5 oxidized in the presence of metaperiodate, preferably in the presence of sodium periodate (NaIO_4). In another embodiment the polysaccharide is oxidized in the presence of orthoperiodate, preferably in the presence of periodic acid.

In one embodiment, the oxidizing agent is a stable nitroxyl or nitroxide radical compound, such as piperidine-N-oxy or pyrrolidine-N-oxy compounds, in the presence

10 of an oxidant to selectively oxidize primary hydroxyls. In said reaction, the actual oxidant is the N-oxoammonium salt, in a catalytic cycle. In an aspect, said stable nitroxyl or nitroxide radical compound are piperidine-N-oxy or pyrrolidine-N-oxy compounds. In an aspect, said stable nitroxyl or nitroxide radical compound bears a TEMPO (2,2,6,6-tetramethyl-1-piperidinyloxy) or a PROXYL (2,2,5,5-tetramethyl-1-pyrrolidinyloxy)

15 moiety. In an aspect, said stable nitroxyl radical compound is TEMPO or a derivative thereof. In an aspect, said oxidant is a molecule bearing a N-halo moiety. In an aspect, said oxidant is selected from any one of N-Chlorosuccinimide, N-Bromosuccinimide, N-Iodosuccinimide, Dichloroisocyanuric acid, 1,3,5-trichloro-1,3,5-triazine-2,4,6-trione, Dibromoisocyanuric acid, 1,3,5-tribromo-1,3,5-triazine-2,4,6-trione, Diiodoisocyanuric

20 acid and 1,3,5-triiodo-1,3,5-triazine-2,4,6-trione. Preferably said oxidant is N-Chlorosuccinimide.

Following the oxidation step of the saccharide, the saccharide is said to be activated and is referred to as "activated" herein below. The activated saccharide and the carrier protein may be lyophilized (freeze-dried), either independently (discrete

25 lyophilization) or together (co-lyophilized). In one embodiment the activated saccharide and the carrier protein are co-lyophilized. In another embodiment the activated polysaccharide and the carrier protein are lyophilized independently.

In one embodiment the lyophilization takes place in the presence of a non-reducing sugar, possible non-reducing sugars include sucrose, trehalose, raffinose,

30 stachyose, melezitose, dextran, mannitol, lactitol and palatinin.

The next step of the conjugation process is the reduction of the activated saccharide and a carrier protein to form a conjugate (so-called reductive amination), using a reducing agent. Suitable reducing agents include the cyanoborohydrides, such as sodium cyanoborohydride, sodium triacetoxymethylborohydride or sodium or zinc

35 borohydride in the presence of Bronsted or Lewis acids), amine boranes such as pyridine borane, 2-Picoline Borane, 2,6-diborane-methanol, dimethylamine-borane, t-

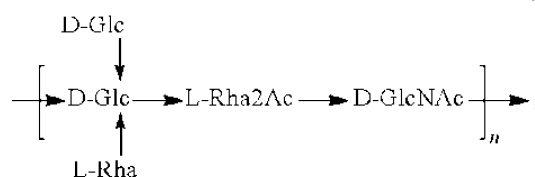
BuMe'PrN-BH₃, benzylamine-BH₃ or 5-ethyl-2-methylpyridine borane (PEMB), borane-pyridine, or borohydride exchange resin. In one embodiment the reducing agent is sodium cyanoborohydride.

In an embodiment, the reduction reaction is carried out in aqueous solvent (e.g., selected from PBS, MES, HEPES, Bis-tris, ADA, PIPES, MOPSO, BES, MOPS, DIPSO, MOBS, HEPPSO, POPSO, TEA, EPPS, Bicine or HEPB, at a pH between 6.0 and 8.5, 7.0 and 8.0, or 7.0 and 7.5), in another embodiment the reaction is carried out in aprotic solvent. In an embodiment, the reduction reaction is carried out in DMSO (dimethylsulfoxide) or in DMF (dimethylformamide) solvent. The DMSO or DMF solvent may be used to reconstitute the activated polysaccharide and carrier protein which has been lyophilized.

At the end of the reduction reaction, there may be unreacted aldehyde groups remaining in the conjugates, these may be capped using a suitable capping agent. In one embodiment this capping agent is sodium borohydride (NaBH₄). Following the conjugation (the reduction reaction and optionally the capping), the glycoconjugates may be purified (enriched with respect to the amount of polysaccharide-protein conjugate) by a variety of techniques known to the skilled person. These techniques include dialysis, concentration/diafiltration operations, tangential flow filtration precipitation/elution, column chromatography (DEAE or hydrophobic interaction chromatography), and depth filtration. The glycoconjugates may be purified by diafiltration and/or ion exchange chromatography and/or size exclusion chromatography. In an embodiment, the glycoconjugates are purified by diafiltration or ion exchange chromatography or size exclusion chromatography. In one embodiment the glycoconjugates are sterile filtered.

In a preferred embodiment, a glycoconjugate from an *E. coli* serotype is selected from any one of O25B, O1, O2, and O6 is prepared by reductive amination. In a preferred embodiment, the glycoconjugates from *E. coli* serotypes O25B, O1, O2, and O6 are prepared by reductive amination.

In one aspect, the invention relates to a conjugate that includes a carrier protein, e.g., CRM₁₉₇, linked to a saccharide of Formula O25B, presented by



, wherein n is any integer greater than or equal to 1. In a preferred embodiment, n is an integer of at least 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, and at most 200, 100, 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 68, 67, 66, 65, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51, or 50. Any minimum value and any maximum value may be combined to define a range.

Exemplary ranges include, for example, at least 1 to at most 1000; at least 10 to at most 500;

and at least 20 to at most 80. In one preferred embodiment, n is at least 31 to at most 90, more preferably 40 to 90, most preferably 60 to 85.

In another aspect, the invention relates to a conjugate that includes a carrier protein, e.g., CRM₁₉₇, linked to a saccharide having any one of the following structures shown in **Table 1** (see also **FIG. 9A-9C** and **FIG. 10A-10B**), wherein n is an integer greater than or equal to 1.

Without being bound by theory or mechanism, in some embodiments, a stable conjugate is believed to require a level of saccharide antigen modification that is balanced against preserving the structural integrity of the critical immunogenic epitopes of the antigen.

Activation and formation of an Aldehyde. In some embodiments, the saccharide of the invention is activated and results in the formation of an aldehyde. In such embodiments wherein the saccharide is activated, the percentage (%) of activation (or degree of oxidation (DO)) (see, e.g., **Example 31**) refers to moles of a saccharide repeat unit per moles of aldehyde of the activated polysaccharide. For example, in some embodiments, the saccharide is activated by periodate oxidation of vicinal diols on a repeat unit of the polysaccharide, resulting in the formation of an aldehyde. Varying the molar equivalents (meq) of sodium periodate relative to the saccharide repeat unit and temperature during oxidation results in varying levels of degree of oxidation (DO).

The saccharide and aldehyde concentrations are typically determined by colorimetric assays. An alternative reagent is TEMPO (2,2,6,6-tetramethylpiperidine 1-oxyl radical)–N-chlorosuccinimide (NCS) combination, which results in the formation of aldehydes from primary alcohol groups.

In some embodiments, the activated saccharide has a degree of oxidation wherein the moles of a saccharide repeat unit per moles of aldehyde of the activated saccharide is between 1-100, such as, for example, between 2-80, between 2-50, between 3-30, and between 4-25. The degree of activation is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ≥ 20 , ≥ 30 , ≥ 40 , ≥ 50 , ≥ 60 , ≥ 70 , ≥ 80 , or ≥ 90 , or about 100. Preferably, the degree of oxidation (DO) is at least 5 and at most 50, more preferably at least 10 and at most 25. In one embodiment, the degree of activation is at least 10 and at most 25. Any minimum value and any maximum value may be combined to define a range. A degree of oxidation value may be represented as percentage (%) of activation. For example, in one embodiment, a DO value of 10 refers to one activated saccharide repeat unit out of a total of 10 saccharide repeat units in the activated saccharide, in which case the DO value of 10 may be represented as 10% activation.

In some embodiments, the conjugate prepared by reductive amination chemistry includes a carrier protein and a saccharide, wherein the saccharide includes a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185,

Formula O186, and Formula O187. In some embodiments, the saccharide in the conjugate includes a Formula, wherein n is an integer from 1 to 1000, from 5 to 1000, preferably 31 to 100, more preferably 35 to 90, most preferably 35 to 65.

SINGLE-END LINKED CONJUGATES. In some embodiments, the conjugate is
 5 single-end-linked conjugated saccharide, wherein the saccharide is covalently bound at one end of the saccharide to a carrier protein. In some embodiments, the single-end-linked conjugated polysaccharide has a terminal saccharide. For example, a conjugate is single-end linked if one of the ends (a terminal saccharide residue) of the polysaccharide is covalently bound to a carrier protein. In some embodiments, the
 10 conjugate is single-end linked if a terminal saccharide residue of the polysaccharide is covalently bound to a carrier protein through a linker. Such linkers may include, for example, a cystamine linker (A1), a 3,3'-dithio bis(propanoic dihydrazide) linker (A4), and a 2,2'-dithio-N,N'-bis(ethane-2,1-diyl)bis(2-(aminooxy)acetamide) linker (A6).

In some embodiments, the saccharide is conjugated to the carrier protein
 15 through a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) residue to form a single-end linked conjugate. See, for example, **Example 26**, **Example 27**, **Example 28**, and **FIG. 17**.

In some embodiments, the conjugate is preferably not a bioconjugate. The term
 "bioconjugate" refers to a conjugate between a protein (e.g., a carrier protein) and an
 20 antigen, e.g., an O antigen (e.g., O25B) prepared in a host cell background, wherein host cell machinery links the antigen to the protein (e.g., N-links). Glycoconjugates include bioconjugates, as well as sugar antigen (e.g., oligo- and polysaccharides)-protein conjugates prepared by means that do not require preparation of the conjugate in a host cell, e.g., conjugation by chemical linkage of the protein and saccharide.

Thiol Activated Saccharides. In some embodiments, the saccharide of the
 25 invention is thiol activated. In such embodiments wherein the saccharide is thiol activated, the percentage (%) of activation refers to moles of thiol per saccharide repeat unit of the activated polysaccharide. The saccharide and thiol concentrations are typically determined by Ellman's assay for quantitation of sulfhydryls. For example, in
 30 some embodiments, the saccharide includes activation of 2-Keto-3-deoxyoctanoic acid (KDO) with a disulfide amine linker. See, for example, **Example 10** and **FIG. 31**. In some embodiments, the saccharide is covalently bound to a carrier protein through a bivalent, heterobifunctional linker (also referred to herein as a "spacer"). The linker preferably provides a thioether bond between the saccharide and the carrier protein,
 35 resulting in a glycoconjugate referred to herein as a "thioether glycoconjugate." In some

embodiments, the linker further provides carbamate and amide bonds, such as, for example, (2-((2-oxoethyl)thio)ethyl) carbamate (eTEC). See, for example, **Example 21**.

In some embodiments, the single-end linked conjugate includes a carrier protein and a saccharide, wherein the saccharide includes a structure selected from any one of

5 Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula

10 O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40,

15 Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula

20 O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula

25 O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula

30 O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula

35 O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula

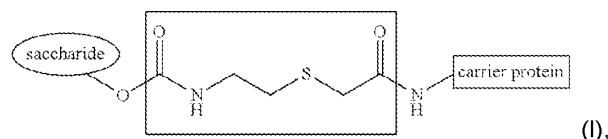
O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187. In some embodiments, the saccharide in the conjugate includes a Formula, wherein n is an integer from 1 to 1000, from 5 to 1000, preferably 31 to 100, more preferably 35 to 90, most preferably 35 to 65.

For example, in one embodiment, the single-end linked conjugate includes a carrier protein and a saccharide having a structure selected from Formula O8, Formula O9a, Formula O9, Formula O20ab, Formula O20ac, Formula O52, Formula O97, and Formula O101, wherein n is an integer from 1 to 10.

10 F. eTEC CONJUGATES

In one aspect, the invention relates generally to glycoconjugates comprising a saccharide derived from *E. coli* described above covalently conjugated to a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer (as described, for example, in US Patent 9517274 and International Patent Application Publication
15 WO2014027302, incorporated by reference herein in their entireties), including immunogenic compositions comprising such glycoconjugates, and methods for the preparation and use of such glycoconjugates and immunogenic compositions. Said glycoconjugates comprise a saccharide covalently conjugated to a carrier protein through one or more eTEC spacers, wherein the saccharide is covalently conjugated to
20 the eTEC spacer through a carbamate linkage, and wherein the carrier protein is covalently conjugated to the eTEC spacer through an amide linkage. The eTEC spacer includes seven linear atoms (i.e., $-\text{C}(\text{O})\text{NH}(\text{CH}_2)_2\text{SCH}_2\text{C}(\text{O})-$) and provides stable thioether and amide bonds between the saccharide and carrier protein.

The eTEC linked glycoconjugates of the invention may be represented by the
25 general formula (I):



where the atoms that comprise the eTEC spacer are contained in the central box.

In said glycoconjugates of the invention, the saccharide may be a polysaccharide or an oligosaccharide.

30 The carrier proteins incorporated into the glycoconjugates of the invention are selected from the group of carrier proteins generally suitable for such purposes, as further described herein or known to those of skill in the art. In particular embodiments, the carrier protein is CRM₁₉₇.

In another aspect, the invention provides a method of making a glycoconjugate comprising a saccharide described herein conjugated to a carrier protein through an eTEC spacer, comprising the steps of a) reacting a saccharide with a carbonic acid derivative in an organic solvent to produce an activated saccharide; b) reacting the
5 activated saccharide with cystamine or cysteamine or a salt thereof, to produce a thiolated saccharide; c) reacting the thiolated saccharide with a reducing agent to produce an activated thiolated saccharide comprising one or more free sulfhydryl residues; d) reacting the activated thiolated saccharide with an activated carrier protein comprising one or more α -haloacetamide groups, to produce a thiolated saccharide-carrier protein conjugate; and e) reacting the
10 thiolated saccharide-carrier protein conjugate with (i) a first capping reagent capable of capping unconjugated α -haloacetamide groups of the activated carrier protein; and/or (ii) a second capping reagent capable of capping unconjugated free sulfhydryl residues of the activated thiolated saccharide; whereby an eTEC linked glycoconjugate is produced.

In frequent embodiments, the carbonic acid derivative is 1,1'-carbonyl-di-(1,2,4-triazole) (CDT) or 1,1'-carbonyldiimidazole (CDI). Preferably, the carbonic acid derivative is CDT and
15 the organic solvent is a polar aprotic solvent, such as dimethylsulfoxide (DMSO). In preferred embodiments, the thiolated saccharide is produced by reaction of the activated saccharide with the bifunctional symmetric thioalkylamine reagent, cystamine or a salt thereof. Alternatively, the thiolated saccharide may be formed by reaction of the activated saccharide with cysteamine or
20 a salt thereof. The eTEC linked glycoconjugates produced by the methods of the invention may be represented by general Formula (I).

In frequent embodiments, the first capping reagent is N-acetyl-L-cysteine, which reacts with unconjugated α -haloacetamide groups on lysine residues of the carrier protein to form an S-carboxymethylcysteine (CMC) residue covalently linked to the activated lysine residue
25 through a thioether linkage.

In other embodiments, the second capping reagent is iodoacetamide (IAA), which reacts with unconjugated free sulfhydryl groups of the activated thiolated saccharide to provide a capped thioacetamide. Frequently, step e) comprises capping with both a first capping reagent and a second capping reagent. In certain embodiments, step e) comprises capping with N-
30 acetyl-L-cysteine as the first capping reagent and IAA as the second capping reagent.

In some embodiments, the capping step e) further comprises reaction with a reducing agent, for example, DTT, TCEP, or mercaptoethanol, after reaction with the first and/or second capping reagent.

The eTEC linked glycoconjugates and immunogenic compositions of the invention may
35 include free sulfhydryl residues. In some instances, the activated thiolated saccharides formed by the methods provided herein will include multiple free sulfhydryl residues, some of which

may not undergo covalent conjugation to the carrier protein during the conjugation step. Such residual free sulfhydryl residues are capped by reaction with a thiol-reactive capping reagent, for example, iodoacetamide (IAA), to cap the potentially reactive functionality. Other thiol-reactive capping reagents, e.g., maleimide containing reagents and the like are also contemplated.

In addition, the eTEC linked glycoconjugates and immunogenic compositions of the invention may include residual unconjugated carrier protein, which may include activated carrier protein which has undergone modification during the capping process steps.

In some embodiments, step d) further comprises providing an activated carrier protein comprising one or more α -haloacetamide groups prior to reacting the activated thiolated saccharide with the activated carrier protein. In frequent embodiments, the activated carrier protein comprises one or more α -bromoacetamide groups.

In another aspect, the invention provides an eTEC linked glycoconjugate comprising a saccharide described herein conjugated to a carrier protein through an eTEC spacer produced according to any of the methods disclosed herein.

In some embodiments, the carrier protein is CRM₁₉₇ and the covalent linkage via an eTEC spacer between the CRM₁₉₇ and the polysaccharide occurs at least once in every 4, 10, 15 or 25 saccharide repeat units of the polysaccharide.

For each of the aspects of the invention, in particular embodiments of the methods and compositions described herein, the eTEC linked glycoconjugate comprises a saccharide described herein, such as, a saccharide derived from *E. coli*.

In another aspect, the invention provides a method of preventing, treating or ameliorating a bacterial infection, disease or condition in a subject, comprising administering to the subject an immunologically effective amount of an immunogenic composition of the invention, wherein said immunogenic composition comprises an eTEC linked glycoconjugate comprising a saccharide described herein. In some embodiments, the saccharide is derived from *E. coli*.

In some embodiments, the eTEC linked glycoconjugate comprises a carrier protein and a saccharide, in which said saccharide comprises a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula

O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, 5 Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula 10 O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, 15 Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, 20 Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, 25 Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187. In some embodiments, the saccharide in the conjugate 30 includes a Formula, wherein n is an integer from 1 to 1000, from 5 to 1000, preferably 31 to 100, more preferably 35 to 90, most preferably 35 to 65.

The number of lysine residues in the carrier protein that become conjugated to the saccharide can be characterized as a range of conjugated lysines. For example, in some 35 embodiments of the immunogenic compositions, the CRM₁₉₇ may comprise 4 to 16 lysine residues out of 39 covalently linked to the saccharide. Another way to express this parameter is that about 10% to about 41% of CRM₁₉₇ lysines are covalently linked to the saccharide. In

other embodiments, the CRM₁₉₇ may comprise 2 to 20 lysine residues out of 39 covalently linked to the saccharide. Another way to express this parameter is that about 5% to about 50% of CRM₁₉₇ lysines are covalently linked to the saccharide.

5 In frequent embodiments, the carrier protein is CRM₁₉₇ and the covalent linkage via an eTEC spacer between the CRM₁₉₇ and the polysaccharide occurs at least once in every 4, 10, 15 or 25 saccharide repeat units of the polysaccharide.

In other embodiments, the conjugate comprises at least one covalent linkage between the carrier protein and saccharide for every 5 to 10 saccharide repeat units; every 2 to 7 saccharide repeat units; every 3 to 8 saccharide repeat units; every 4 to 9
10 saccharide repeat units; every 6 to 11 saccharide repeat units; every 7 to 12 saccharide repeat units; every 8 to 13 saccharide repeat units; every 9 to 14 saccharide repeat units; every 10 to 15 saccharide repeat units; every 2 to 6 saccharide repeat units, every 3 to 7 saccharide repeat units; every 4 to 8 saccharide repeat units; every 6 to 10
15 saccharide repeat units; every 7 to 11 saccharide repeat units; every 8 to 12 saccharide repeat units; every 9 to 13 saccharide repeat units; every 10 to 14 saccharide repeat units; every 10 to 20 saccharide repeat units; or every 4 to 25 saccharide repeat units.

In another embodiment, at least one linkage between carrier protein and saccharide occurs for every 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 saccharide repeat units of the polysaccharide.

20 G. CARRIER PROTEINS

A component of the glycoconjugate of the invention is a carrier protein to which the saccharide is conjugated. The terms "protein carrier" or "carrier protein" or "carrier" may be used interchangeably herein. Carrier proteins should be amendable to standard conjugation procedures.

25 One component of the conjugate is a carrier protein to which the O-polysaccharide is conjugated. In one embodiment, the conjugate includes a carrier protein conjugated to the core oligosaccharide of the O-polysaccharide (see FIG. 24). In one embodiment, the conjugate includes a carrier protein conjugated to the O-antigen of the O-polysaccharide.

30 The terms "protein carrier" or "carrier protein" or "carrier" may be used interchangeably herein. Carrier proteins should be amendable to standard conjugation procedures.

In a preferred embodiment, the carrier protein of the conjugates is independently selected from any one of TT, DT, DT mutants (such as CRM₁₉₇), *H. influenzae* protein D, PhtX, PhtD, PhtDE fusions (particularly those described in WO 01/98334 and WO
35

03/54007), detoxified pneumolysin, PorB, N19 protein, PspA, OMPC, toxin A or B of *C. Difficile* and PsaA. In an embodiment, the carrier protein of the conjugates of the invention is DT (Diphtheria toxoid). In another embodiment, the carrier protein of the conjugates of the invention is TT (tetanus toxoid). In another embodiment, the carrier protein of the conjugates of the invention is PD (*Haemophilus influenzae* protein D - see, e.g., EP 0 594 610 B). In some embodiments, the carrier protein includes poly(L-lysine) (PLL).

In a preferred embodiment, the saccharides are conjugated to CRM₁₉₇ protein. The CRM₁₉₇ protein is a nontoxic form of diphtheria toxin but is immunologically indistinguishable from the diphtheria toxin. CRM₁₉₇ is produced by *C. diphtheriae* infected by the nontoxigenic phage β197tox created by nitrosoguanidine mutagenesis of the toxigenic corynebacteriophage beta. The CRM₁₉₇ protein has the same molecular weight as the diphtheria toxin but differs therefrom by a single base change (guanine to adenine) in the structural gene. This single base change causes an amino acid substitution glutamic acid for glycine) in the mature protein and eliminates the toxic properties of diphtheria toxin. The CRM₁₉₇ protein is a safe and effective T-cell dependent carrier for saccharides.

Accordingly, in some embodiments, the conjugates of the invention include CRM₁₉₇ as the carrier protein, wherein the saccharide is covalently linked to CRM₁₉₇.

In a preferred embodiment, the carrier protein of the glycoconjugates is selected in the group consisting of DT (Diphtheria toxin), TT (tetanus toxoid) or fragment C of TT, CRM₁₉₇ (a nontoxic but antigenically identical variant of diphtheria toxin), other DT mutants (such as CRM176, CRM228, CRM 45 (Uchida et al J. Biol. Chem. 218; 3838-3844, 1973), CRM9, CRM45, CRM102, CRM103 or CRM107; and other mutations described by Nicholls and Youle in Genetically Engineered Toxins, Ed: Frankel, Maecel Dekker Inc, 1992; deletion or mutation of Glu-148 to Asp, Gln or Ser and/or Ala 158 to Gly and other mutations disclosed in US 4709017 or US 4950740; mutation of at least one or more residues Lys 516, Lys 526, Phe 530 and/or Lys 534 and other mutations disclosed in US 5917017 or US 6455673; or fragment disclosed in US 5843711), pneumococcal pneumolysin (Kuo et al (1995) Infect Immun 63; 2706-13) including ply detoxified in some fashion for example dPLY-GMBS (WO 04081515, PCT/EP2005/010258) or dPLY-formol, PhtX, including PhtA, PhtB, PhtD, PhtE (sequences of PhtA, PhtB, PhtD or PhtE are disclosed in WO 00/37105 or WO 00/39299) and fusions of Pht proteins for example PhtDE fusions, PhtBE fusions, Pht A-E (WO 01/98334, WO 03/54007, WO2009/000826), OMPC (meningococcal outer membrane protein - usually extracted from *N. meningitidis* serogroup B - EP0372501), PorB (from *N. meningitidis*), PD (*Haemophilus influenzae* protein D - see, e.g., EP 0 594 610 B), or immunologically functional equivalents thereof, synthetic peptides (EP0378881, EP0427347), heat shock proteins (WO 93/17712, WO 94/03208), pertussis proteins (WO 98/58668, EP0471 177), cytokines, lymphokines, growth

factors or hormones (WO 91/01146), artificial proteins comprising multiple human CD4+ T cell epitopes from various pathogen derived antigens (Falugi et al (2001) Eur J Immunol 31 ; 3816-3824) such as N19 protein (Baraldoi et al (2004) Infect Immun 72; 4884-7) pneumococcal surface protein PspA (WO 02/091998), iron uptake proteins (WO 01/72337), toxin A or B of *C. difficile* (WO 00/61761), transferrin binding proteins, pneumococcal adhesion protein (PsaA), recombinant *Pseudomonas aeruginosa* exotoxin A (in particular non-toxic mutants thereof (such as exotoxin A bearing a substitution at glutamic acid 553 (Uchida Cameron DM, RJ Collier. 1987. J. Bacteriol. 169:4967-4971)). Other proteins, such as ovalbumin, keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) or purified protein derivative of tuberculin (PPD) also can be used as carrier proteins. Other suitable carrier proteins include inactivated bacterial toxins such as cholera toxoid (e.g., as described in Int'l Patent Application No. WO 2004/083251), *E. coli* LT, *E. coli* ST, and exotoxin A from *Pseudomonas aeruginosa*.

In some embodiments, the carrier protein is selected from any one of, for example, CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), flagellin, detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP). In one embodiment, the carrier protein is detoxified *Pseudomonas* exotoxin (EPA). In another embodiment, the carrier protein is not detoxified *Pseudomonas* exotoxin (EPA). In one embodiment, the carrier protein is flagellin. In another embodiment, the carrier protein is not flagellin.

In a preferred embodiment, the carrier protein of the glycoconjugates is independently selected from the group consisting of TT, DT, DT mutants (such as CRM₁₉₇), *H. influenzae* protein D, PhtX, PhtD, PhtDE fusions (particularly those described in WO 01/98334 and WO 03/54007), detoxified pneumolysin, PorB, N19 protein, PspA, OMPC, toxin A or B of *C. Difficile* and PsaA. In an embodiment, the carrier protein of the glycoconjugates of the invention is DT (Diphtheria toxoid). In another embodiment, the carrier protein of the glycoconjugates of the invention is TT (tetanus toxoid). In another embodiment, the carrier protein of the glycoconjugates of the invention is PD (*Haemophilus influenzae* protein D - see, e.g., EP 0 594 610 B).

In a preferred embodiment, the capsular saccharides of the invention are conjugated to CRM₁₉₇ protein. The CRM₁₉₇ protein is a nontoxic form of diphtheria toxin but is immunologically indistinguishable from the diphtheria toxin. CRM₁₉₇ is produced by

C. diphtheriae infected by the nontoxigenic phage β 197tox- created by nitrosoguanidine mutagenesis of the toxigenic corynebacteriophage beta (Uchida, T. et al. 1971, Nature New Biology 233:8-11). The CRM₁₉₇ protein has the same molecular weight as the diphtheria toxin but differs therefrom by a single base change (guanine to adenine) in the structural
 5 gene. This single base change causes an amino acid substitution glutamic acid for glycine) in the mature protein and eliminates the toxic properties of diphtheria toxin. The CRM₁₉₇ protein is a safe and effective T-cell dependent carrier for saccharides. Further details about CRM₁₉₇ and production thereof can be found e.g. in US 5,614,382

Accordingly, in frequent embodiments, the glycoconjugates of the invention comprise
 10 CRM₁₉₇ as the carrier protein, wherein the capsular polysaccharide is covalently linked to CRM₁₉₇.

In a further embodiment, the carrier protein of the glycoconjugates is SCP (Streptococcal C5a Peptidase). All human isolates of β -hemolytic streptococci produce a highly conserved cell-wall protein SCP (Streptococcal C5a Peptidase) that specifically inactivates C5a.
 15 The *scp* genes encode a polypeptide containing between 1,134 and 1,181 amino acids (Brown *et al.*, PNAS, 2005, vol. 102, no. 51 pages 18391–18396). The first 31 residues are the export signal presequence and are removed upon passing through the cytoplasmic membrane. The next 68 residues serve as a pro-sequence and must be removed to produce active SCP. The next 10 residues can be removed without loss of protease activity. At the other end, starting
 20 with Lys-1034, are four consecutive 17-residue motifs followed by a cell sorting and cell-wall attachment signal. This combined signal is composed of a 20-residue hydrophilic sequence containing an LPTTND sequence, a 17-residue hydrophobic sequence, and a short basic carboxyl terminus.

SCP can be divided in domains (see figure 1B of Brown *et al.*, PNAS, 2005, vol. 102, no.
 25 51 pages 18391–18396). These domains are the Pre/Pro domain (which comprises the export signal presequence (commonly the first 31 residues) and the pro-sequence (commonly the next 68 residues)), the protease domain (which is splitted in two part (protease part 1 commonly residues 89–333/334 and protease domain part 2 and commonly residues 467/468–583/584), the protease-associated domain (PA domain) (commonly residues 333/334–467/468), three
 30 fibronectin type III (Fn) domains (Fn1, commonly residues 583/584–712/713; Fn2, commonly residues 712/713–928/929/930; commonly Fn3, residues 929/930–1029/1030/1031) and a cell wall anchor domain (commonly residues 1029/1030/1031 to the C-terminus).

In an embodiment, the carrier protein of the glycoconjugates of the invention is an SCP from GBS (SCP_B). An example of SCP_B is provided at SEQ. ID.NO: 3 of WO97/26008. See
 35 also SEQ ID NO: 3 of WO00/34487.

In another embodiment, the carrier protein of the glycoconjugate of the invention is an SCP from GAS (SCPA). Examples of SCPA can be found at SEQ.ID.NO:1 and SEQ.ID.NO:2 of WO97/26008. See also SEQ ID NOs: 1, 2 and 23 of WO00/34487.

In a further embodiment, the carrier protein of the glycoconjugate of the invention is an SCP as set forth in SEQ ID NO: 150 or 151 of WO2014/136064.

H. DOSAGES OF THE COMPOSITIONS

Dosage regimens may be adjusted to provide the optimum desired response. For example, a single dose of the polypeptide derived from *E. coli* or fragment thereof may be administered, several divided doses may be administered over time, or the dose may be proportionally reduced or increased as indicated by the exigencies of the situation. It is to be noted that dosage values may vary with the type and severity of the condition to be alleviated, and may include single or multiple doses. It is to be further understood that for any particular subject, specific dosage regimens should be adjusted over time according to the individual need and the professional judgment of the person administering or supervising the administration of the compositions, and that dosage ranges set forth herein are exemplary only and are not intended to limit the scope or practice of the claimed composition. Determining appropriate dosages and regimens for administration of the therapeutic protein are well-known in the relevant art and would be understood to be encompassed by the skilled artisan once provided the teachings disclosed herein.

In some embodiments, the amount of the polypeptide derived from *E. coli* or fragment thereof in the composition, may range from about 10 µg to about 300 µg of each protein antigen. In some embodiments, the amount of the polypeptide derived from *E. coli* or fragment thereof in the composition may range from about 20 µg to about 200 µg of each protein antigen.

The amount of glycoconjugate(s) in each dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccines. Such amount will vary depending upon which specific immunogen is employed and how it is presented.

The amount of a particular glycoconjugate in an immunogenic composition can be calculated based on total polysaccharide for that conjugate (conjugated and non-conjugated). For example, a glycoconjugate with 20% free polysaccharide will have about 80 µg of conjugated polysaccharide and about 20 µg of non-conjugated polysaccharide in a 100 µg polysaccharide dose. The amount of glycoconjugate can

vary depending upon the *E. coli* serotype. The saccharide concentration can be determined by the uronic acid assay.

The "immunogenic amount" of the different polysaccharide components in the immunogenic composition, may diverge and each may comprise about 1.0 µg, about 2.0 µg, about 3.0 µg, about 4.0 µg, about 5.0 µg, about 6.0 µg, about 7.0 µg, about 8.0 µg, about 9.0 µg, about 10.0 µg, about 15.0 µg, about 20.0 µg, about 30.0 µg, about 40.0 µg, about 50.0 µg, about 60.0 µg, about 70.0 µg, about 80.0 µg, about 90.0 µg, or about 100.0 µg of any particular polysaccharide antigen. Generally, each dose will comprise 0.1 µg to 100 µg of polysaccharide for a given serotype, particularly 0.5 µg to 20 µg, more particularly 1 µg to 10 µg, and even more particularly 2 µg to 5 µg. Any whole number integer within any of the above ranges is contemplated as an embodiment of the disclosure. In one embodiment, each dose will comprise 1 µg, 2 µg, 3 µg, 4 µg, 5 µg, 6 µg, 7 µg, 8 µg, 9 µg, 10 µg, 15 µg or 20 µg of polysaccharide for a given serotype.

Carrier protein amount. Generally, each dose will comprise 5 µg to 150 µg of carrier protein, particularly 10 µg to 100 µg of carrier protein, more particularly 15 µg to 100 µg of carrier protein, more particularly 25 µg to 75 µg of carrier protein, more particularly 30 µg to 70 µg of carrier protein, more particularly 30 µg to 60 µg of carrier protein, more particularly 30 µg to 50 µg of carrier protein and even more particularly 40 µg to 60 µg of carrier protein. In one embodiment, said carrier protein is CRM₁₉₇. In one embodiment, each dose will comprise about 25 µg, about 26 µg, about 27 µg, about 28 µg, about 29 µg, about 30 µg, about 31 µg, about 32 µg, about 33 µg, about 34 µg, about 35 µg, about 36 µg, about 37 µg, about 38 µg, about 39 µg, about 40 µg, about 41 µg, about 42 µg, about 43 µg, about 44 µg, about 45 µg, about 46 µg, about 47 µg, about 48 µg, about 49 µg, about 50 µg, about 51 µg, about 52 µg, about 53 µg, about 54 µg, about 55 µg, about 56 µg, about 57 µg, about 58 µg, about 59 µg, about 60 µg, about 61 µg, about 62 µg, about 63 µg, about 64 µg, about 65 µg, about 66 µg, about 67 µg, about 68 µg, about 69 µg, about 70 µg, about 71 µg, about 72 µg, about 73 µg, about 74 µg or about 75 µg of carrier protein. In one embodiment, said carrier protein is CRM₁₉₇. In another embodiment, said carrier protein is SCP.

I. ADJUVANT

In some embodiments, the immunogenic compositions disclosed herein may further comprise at least one, two or three adjuvants. In some embodiments, the immunogenic compositions disclosed herein may further comprise at least one adjuvant. In some embodiments, the immunogenic compositions disclosed herein may further comprise one adjuvant. In some embodiments, the immunogenic compositions disclosed herein may further comprise two adjuvants. The term "adjuvant" refers to a compound or mixture that enhances the

immune response to an antigen. Antigens may act primarily as a delivery system, primarily as an immune modulator or have strong features of both. Suitable adjuvants include those suitable for use in mammals, including humans.

5 Examples of known suitable delivery-system type adjuvants that can be used in humans include, but are not limited to, alum (e.g., aluminum phosphate, aluminum sulfate or aluminum hydroxide), calcium phosphate, liposomes, oil-in-water emulsions such as MF59 (4.3% w/v squalene, 0.5% w/v polysorbate 80 (Tween 80), 0.5% w/v sorbitan trioleate (Span 85)), water-in-oil emulsions such as Montanide, and poly(D,L-lactide-co-glycolide) (PLG) microparticles or nanoparticles.

10 In an embodiment, the immunogenic compositions disclosed herein comprise aluminum salts (alum) as adjuvant (e.g., aluminum phosphate, aluminum sulfate or aluminum hydroxide). In a preferred embodiment, the immunogenic compositions disclosed herein comprise aluminum phosphate or aluminum hydroxide as adjuvant. In an embodiment, the immunogenic compositions disclosed herein comprise from 0.1 mg/mL
15 to 1 mg/mL or from 0.2 mg/mL to 0.3 mg/mL of elemental aluminum in the form of aluminum phosphate. In an embodiment, the immunogenic compositions disclosed herein comprise about 0.25 mg/mL of elemental aluminum in the form of aluminum phosphate.

Examples of known suitable immune modulatory type adjuvants that can be used in humans include, but are not limited to, saponin extracts from the bark of the Aquilla tree
20 (QS21, Quil A), TLR4 agonists such as MPLA (Monophosphoryl Lipid A), 3DMPL (3-O-deacylated MPL) or GLA-AQ, LT/CT mutants, cytokines such as the various interleukins (e.g., IL-2, IL-12) or GM-CSF, AS01, and the like.

Examples of known suitable immune modulatory type adjuvants with both delivery and immune modulatory features that can be used in humans include, but are not limited
25 to, ISCOMS (see, e.g., Sjölander et al. (1998) J. Leukocyte Biol. 64:713; WO 90/03184, WO 96/11711, WO 00/48630, WO 98/36772, WO 00/41720, WO 2006/134423 and WO 2007/026190) or GLA-EM which is a combination of a TLR4 agonist and an oil-in-water emulsion.

For veterinary applications including but not limited to animal experimentation, one
30 can use Complete Freund's Adjuvant (CFA), Freund's Incomplete Adjuvant (IFA), Emulsigen, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI,
35 which contains three components extracted from bacteria, monophosphoryl lipid A,

trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion.

Further exemplary adjuvants to enhance effectiveness of the immunogenic compositions disclosed herein include, but are not limited to (1) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (b) RIBI™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, Mont.) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components such as monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL+CWS (DETOX™); (2) saponin adjuvants, such as QS21, STIMULON™ (Cambridge Bioscience, Worcester, Mass.), ABISCO® (Isconova, Sweden), or ISCOMATRIX® (Commonwealth Serum Laboratories, Australia), may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), which ISCOMs may be devoid of additional detergent (e.g., WO 00/07621); (3) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (4) cytokines, such as interleukins (e.g., IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 (e.g., WO 99/44636)), interferons (e.g., gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (5) monophosphoryl lipid A (MPL) or 3-O-deacylated MPL (3dMPL) (see, e.g., GB2220211, EP0689454) (see, e.g., WO 00/56358); (6) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (see, e.g., EP0835318, EP0735898, EP0761231); (7) a polyoxyethylene ether or a polyoxyethylene ester (see, e.g., WO 99/52549); (8) a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol (e.g., WO 01/21207) or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol (e.g., WO 01/21152); (9) a saponin and an immunostimulatory oligonucleotide (e.g., a CpG oligonucleotide) (e.g., WO 00/62800); (10) an immunostimulant and a particle of metal salt (see, e.g., WO 00/23105); (11) a saponin and an oil-in-water emulsion (e.g., WO 99/11241); (12) a saponin (e.g., QS21)+3dMPL+IM2 (optionally+a sterol) (e.g., WO 98/57659); (13) other substances that act as immunostimulating agents to enhance the efficacy of the composition. Muramyl peptides include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-25 acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutarninyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE), etc.

In another embodiment, the adjuvant is a liposomal QS21 formulation as set forth in Example 35. In a further embodiment, the adjuvant is a liposomal MPLA formulation as set forth

in Example 35. In a still further embodiment, the adjuvant is a liposomal MPLA/QS21 formulation as set forth in Example 35.

In an embodiment of the present invention, the immunogenic compositions as disclosed herein comprise a CpG Oligonucleotide as adjuvant. A CpG oligonucleotide as used herein refers to an immunostimulatory CpG oligodeoxynucleotide (CpG ODN), and accordingly these terms are used interchangeably unless otherwise indicated. Immunostimulatory CpG oligodeoxynucleotides contain one or more immunostimulatory CpG motifs that are unmethylated cytosine-guanine dinucleotides, optionally within certain preferred base contexts. The methylation status of the CpG immunostimulatory motif generally refers to the cytosine residue in the dinucleotide. An immunostimulatory oligonucleotide containing at least one unmethylated CpG dinucleotide is an oligonucleotide which contains a 5' unmethylated cytosine linked by a phosphate bond to a 3' guanine, and which activates the immune system through binding to Toll-like receptor 9 (TLR-9). In another embodiment the immunostimulatory oligonucleotide may contain one or more methylated CpG dinucleotides, which will activate the immune system through TLR9 but not as strongly as if the CpG motif(s) was/were unmethylated. CpG immunostimulatory oligonucleotides may comprise one or more palindromes that in turn may encompass the CpG dinucleotide. CpG oligonucleotides have been described in a number of issued patents, published patent applications, and other publications, including U.S. Pat. Nos. 6,194,388; 6,207,646; 6,214,806; 6,218,371; 6,239,116; and 6,339,068.

In an embodiment of the present invention, the immunogenic compositions as disclosed herein comprise any of the CpG Oligonucleotide described at page 3, line 22, to page 12, line 36, of WO 2010/125480.

Different classes of CpG immunostimulatory oligonucleotides have been identified. These are referred to as A, B, C and P class, and are described in greater detail at page 3, line 22, to page 12, line 36, of WO 2010/125480. Methods of the invention embrace the use of these different classes of CpG immunostimulatory oligonucleotides.

VII. Nanoparticles

In another aspect, disclosed herein is an immunogenic complex that includes 1) a nanostructure; and 2) at least one fimbrial polypeptide antigen or fragment thereof. Preferably, the fimbrial polypeptide or fragment thereof is derived from *E. coli* fimbrial H (fimH). In a preferred embodiment, the fimbrial polypeptide is selected from any one of the fimbrial polypeptides described above. For example, the fimbrial polypeptide may comprise any one amino acid sequence selected from SEQ ID NOs:1-10, 18, 20, 21, 23, 24, 26-29 and 110-113.

In some embodiments, the antigen is fused or conjugated to the nanostructure exterior to stimulate development of adaptive immune responses to the displayed epitopes. In some embodiments, the immunogenic complex further includes an adjuvant or other immunomodulatory compounds attached to the exterior and/or encapsulated in the cage interior to help tailor the type of immune response generated for each pathogen.

In some embodiments, the nanostructure includes a single assembly including a plurality of identical first nanostructure-related polypeptides.

In alternative embodiments, the nanostructure includes a plurality assembly, including a plurality of identical first nanostructure-related polypeptides and a plurality of second assemblies, each second assembly comprising a plurality of identical second nanostructure-related polypeptides.

Various nanostructure platforms can be employed in generating the immunogenic compositions described herein. In some embodiments, the nanostructures employed are formed by multiple copies of a single subunit. In some embodiments, the nanostructures employed are formed by multiple copies of multiple different subunits.

The nanostructures are typically ball-like shaped, and/or have rotational symmetry (e.g., with 3-fold and 5-fold axis), e.g., with an icosahedral structure exemplified herein.

In some embodiments, the antigen is presented on self-assembling nanoparticles such as self-assembling nanostructures derived from ferritin (FR), E2p, Q β , and I3-01. E2p is a redesigned variant of dihydrolipoyl acyltransferase from *Bacillus stearothermophilus*. I3-01 is an engineered protein that may self-assemble into hyperstable nanoparticles. Sequences of the subunits of these proteins are known in the art. In a first aspect, disclosed herein is a nanostructure-related polypeptide comprising an amino acid sequence that is at least 75% identical over its length, and identical at least at one identified interface position, to the amino acid sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-92. The nanostructure-related polypeptides can be used, for example, to prepare the nanostructures. The nanostructure-related polypeptides were designed for their ability to self-assemble in pairs to form nanostructures, such as icosahedral nanostructures.

In some embodiments, the nanostructure includes (a) a plurality of first assemblies, each first assembly comprising a plurality of identical first nanostructure-related polypeptides, wherein the first nanostructure-related polypeptides comprise the amino acid sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-92; and (b) a plurality of second assemblies, each second assembly comprising a plurality of identical second nanostructure-related polypeptides, wherein the second nanostructure-related polypeptides comprise the amino acid sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-92, and wherein the second

nanostructure-related polypeptide differs from the first nanostructure-related polypeptide; wherein the plurality of first assemblies non-covalently interact with the plurality of second assemblies to form a nanostructure.

5 The nanostructures include symmetrically repeated, non-natural, non-covalent polypeptide-polypeptide interfaces that orient a first assembly and a second assembly into a nanostructure, such as one with an icosahedral symmetry.

SEQ ID NOS: 59-92 provide the amino acid sequence of exemplary nanostructure-related polypeptides. The number of interface residues for the exemplary nanostructure-related polypeptides of SEQ ID NO:59-92 range from 4-13
10 residues. In various embodiments, the nanostructure-related polypeptides comprise an amino acid sequence that is at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical over its length, and identical at least at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 identified interface positions (depending on the number of interface residues for a given nanostructure-related polypeptide), to the amino acid
15 sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-92. In other embodiments, the nanostructure-related polypeptides comprise an amino acid sequence that is at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical over its length, and identical at least at 20%, 25%, 33%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 100% of the identified
20 interface positions, to the amino acid sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-92. In further embodiments, the nanostructure-related polypeptides include a nanostructure-related polypeptide having the amino acid sequence of a nanostructure-related polypeptide selected from the group consisting of SEQ ID NOS: 59-98.

25 In one non-limiting embodiment, the nanostructure-related polypeptides can be modified to facilitate covalent linkage to a "cargo" of interest. In one non-limiting example, the nanostructure-related polypeptides can be modified, such as by introduction of various cysteine residues at defined positions to facilitate linkage to one or more antigens of interest, such that a nanostructure of the nanostructure-related
30 polypeptides would provide a scaffold to provide a large number of antigens for delivery as a vaccine to generate an improved immune response.

In some embodiments, some or all native cysteine residues that are present in the nanostructure-related polypeptides but not intended to be used for conjugation may be mutated to other amino acids to facilitate conjugation at defined positions. In another
35 non-limiting embodiment, the nanostructure-related polypeptides may be modified by linkage (covalent or non-covalent) with a moiety to help facilitate "endosomal escape."

For applications that involve delivering molecules of interest to a target cell, such as targeted delivery, a critical step can be escape from the endosome—a membrane-bound organelle that is the entry point of the delivery vehicle into the cell. Endosomes mature into lysosomes, which degrade their contents. Thus, if the delivery vehicle does not somehow “escape” from the endosome before it becomes a lysosome, it will be degraded and will not perform its function. There are a variety of lipids or organic polymers that disrupt the endosome and allow escape into the cytosol. Thus, in this embodiment, the nanostructure-related polypeptides can be modified, for example, by introducing cysteine residues that will allow chemical conjugation of such a lipid or organic polymer to the monomer or resulting assembly surface. In another non-limiting example, the nanostructure-related polypeptides can be modified, for example, by introducing cysteine residues that will allow chemical conjugation of fluorophores or other imaging agents that allow visualization of the nanostructures in vitro or in vivo.

Surface amino acid residues on the nanostructure-related polypeptides can be mutated in order to improve the stability or solubility of the protein subunits or the assembled nanostructures. As will be known to one of skill in the art, if the nanostructure-related polypeptide has significant sequence homology to an existing protein family, a multiple sequence alignment of other proteins from that family can be used to guide the selection of amino acid mutations at non-conserved positions that can increase protein stability and/or solubility, a process referred to as consensus protein design (9).

Surface amino acid residues on the nanostructure-related polypeptides can be mutated to positively charged (Arg, Lys) or negatively charged (Asp, Glu) amino acids in order to endow the protein surface with an overall positive or overall negative charge. In one non-limiting embodiment, surface amino acid residues on the nanostructure-related polypeptides can be mutated to endow the interior surface of the self-assembling nanostructure with a high net charge. Such a nanostructure can then be used to package or encapsulate a cargo molecule with the opposite net charge due to the electrostatic interaction between the nanostructure interior surface and the cargo molecule. In one non-limiting embodiment, surface amino acid residues on the nanostructure-related polypeptides can be mutated primarily to Arginine or Lysine residues in order to endow the interior surface of the self-assembling nanostructure with a net positive charge. Solutions containing the nanostructure-related polypeptides can then be mixed in the presence of a nucleic acid cargo molecule such as a dsDNA, ssDNA, dsRNA, ssRNA, cDNA, miRNA, siRNA, shRNA, piRNA, or other nucleic acid in order to encapsulate the nucleic acid inside the self-assembling nanostructure. Such a nanostructure could be used, for example, to protect, deliver, or concentrate nucleic acids.

In one embodiment, the nanostructure has icosahedral symmetry. In this embodiment, the nanostructure may comprise 60 copies of the first nanostructure-related polypeptide and 60

copies of the second nanostructure-related polypeptide. In one such embodiment, the number of identical first nanostructure-related polypeptides in each first assembly is different than the number of identical second nanostructure-related polypeptides in each second assembly. For example, in one embodiment, the nanostructure comprises
5 twelve first assemblies and twenty second assemblies; in this embodiment, each first assembly may, for example, comprise five copies of the identical first nanostructure-related polypeptide, and each second assembly may, for example, comprise three copies of the identical second nanostructure-related polypeptide. In another
10 embodiment, the nanostructure comprises twelve first assemblies and thirty second assemblies; in this embodiment, each first assembly may, for example, comprise five copies of the identical first nanostructure-related polypeptide, and each second assembly may, for example, comprise two copies of the identical second nanostructure-related polypeptide. In a further embodiment, the nanostructure comprises twenty first
15 assemblies and thirty second assemblies; in this embodiment, each first assembly may, for example, comprise three copies of the identical first nanostructure-related polypeptide, and each second assembly may, for example, comprise two copies of the identical second nanostructure-related polypeptide. All of these embodiments are capable of forming synthetic nanomaterials with regular icosahedral symmetry.

20 **VIII. COMBINATION WITH A SACCHARIDE AND/OR POLYPEPTIDE OR FRAGMENT THEREOF DERIVED FROM *KLEBSIELLA PNEUMONIAE***

Klebsiella pneumoniae (*K. pneumoniae*) is a Gram-negative pathogen, known to cause urinary tract infections, bacteremia, and sepsis. Multidrug-resistant *Klebsiella pneumoniae* infections are an increasing cause of mortality in vulnerable populations at
25 risk. The O-antigen serotypes are highly prevalent among strains causing invasive disease globally and derived O-antigen glycoconjugates are attractive as vaccine antigens.

In one aspect, any of the compositions disclosed herein may further comprise at least one saccharide that is, or is derived from, at least one *K. pneumoniae* serotype
30 selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12. In a preferred embodiment, any of the compositions disclosed herein may further comprise a polypeptide derived from *K. pneumoniae* selected from a polypeptide derived from *K. pneumoniae* Type I fimbrial protein or an immunogenic fragment thereof; or a polypeptide derived from *K. pneumoniae* Type III fimbrial protein
35 or an immunogenic fragment thereof; or a combination thereof.

As is known in the art, *K. pneumoniae* O1 and O2 O-antigens and their corresponding v1 and v2 subtypes are polymeric galactans that differ in the structures of their repeat units. *K. pneumoniae* O1 and O2 antigens contain homopolymer galactose units (or galactans). *K. pneumoniae* O1 and O2 antigens each contain D-galactan I units (sometimes referred to as the O2a repeat unit), but O1 antigens differ in that O1 antigens have a D-galactan II cap structure. D-galactan III (d-Gal-III) is a variant of D-galactan I. Structures of the base galactans I and III that define the two distinct serotype O2 subtypes, O2v1 and O2v2; and structures of the derived chimeras resulting from capping by galactan II which yields subtypes O1v1 and O1v2, are shown in Kelly SD, et al. *J Biol Chem* 2019; 294:10863-76; and Clarke BR, et al. *J Biol Chem* 2018; 293:4666-79.

In some embodiments, the saccharide derived from *K. pneumoniae* O1 includes a repeat unit of $[\rightarrow 3)\text{-}\beta\text{-D-Galf}\text{-(1}\rightarrow 3)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow]$. In some embodiments, the saccharide derived from *K. pneumoniae* O1 includes a repeat unit of $[\rightarrow 3)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow 3)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow]$. In some embodiments, the saccharide derived from *K. pneumoniae* O1 includes a repeat unit of $[\rightarrow 3)\text{-}\beta\text{-D-Galf}\text{-(1}\rightarrow 3)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow]$, and a repeat unit of $[\rightarrow 3)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow 3)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow]$. In some embodiments, the saccharide derived from *K. pneumoniae* O1 includes a repeat unit of $\rightarrow 3)\text{-}\beta\text{-D-Galf}\text{-(1}\rightarrow 3)\text{-}[\alpha\text{-D-Galp}\text{-(1}\rightarrow 4)]\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow]$ (referred to as the D-Gal-III repeat unit). (Kol O., et al. (1992) *Carbohydr. Res.* 236, 339–344; Whitfield C., et al. (1991) *J. Bacteriol.* 173, 1420–1431).

In some embodiments, the saccharide derived from *K. pneumoniae* O2 includes a repeat unit of $[\rightarrow 3)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow 3)\text{-}\beta\text{-D-Galf}\text{-(1}\rightarrow]$ (which may be an element of *K. pneumoniae* serotype O2a antigen). In some embodiments, the saccharide derived from *K. pneumoniae* O2 includes a repeat unit of $[\rightarrow 3)\text{-}\beta\text{-D-GlcNAc}\text{-(1}\rightarrow 5)\text{-}\beta\text{-D-Galf}\text{-(1}\rightarrow]$ (which may be an element of *K. pneumoniae* serotype O2c antigen). In some embodiments, the saccharide derived from *K. pneumoniae* O2 includes a modification of the O2a repeat unit by side chain addition of (1→4)-linked Galp residues (which may be an element of the *K. pneumoniae* O2afg antigen). In some embodiments, the saccharide derived from *K. pneumoniae* O2 includes a modification of the O2a repeat unit by side chain addition of (1→2)-linked Galp residues (which may be an element of the *K. pneumoniae* O2aeh antigen). (Whitfield C., et al. (1992) *J. Bacteriol.* 174, 4913–4919).

Without being bound by mechanism or theory, O-antigen polysaccharide structure of *K. pneumoniae* serotypes O3 and O5 are disclosed in the art to be identical to those of *E. coli* serotypes O9a (Formula O9a) and O8 (Formula O8), respectively.

In some embodiments, the saccharide derived from *K. pneumoniae* O4 includes a repeat unit of $[\rightarrow 4)\text{-}\alpha\text{-D-Galp}\text{-(1}\rightarrow 2)\text{-}\beta\text{-D-Ribf}\text{-(1}\rightarrow]$. In some embodiments, the saccharide derived from *K. pneumoniae* O7 includes a repeat unit of $[\rightarrow 2)\text{-}\alpha\text{-L-Rhap}\text{-(1}\rightarrow 2)\text{-}\beta\text{-D-Ribf}\text{-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rhap}\text{-(1}\rightarrow 3)\text{-}\alpha\text{-L-Rhap}\text{-(1}\rightarrow]$. In some embodiments, the saccharide derived from *K.*

pneumoniae O8 serotype includes the same repeat-unit structure as *K. pneumoniae* O2a, but is nonstoichiometrically O-acetylated. In some embodiments, the saccharide derived from *K. pneumoniae* O12 serotype includes a repeat unit of [α -Rhap-(1 $\rightarrow$ 3)- β -GlcNAc] disaccharide repeat unit.

5 In one aspect, the invention includes a composition including a polypeptide derived from *E. coli* FimH or a fragment thereof; and at least one saccharide that is, or derived from, at least one *K. pneumoniae* serotype selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12. In some
10 embodiments, the composition includes saccharides from or derived from one or more of serotypes O1, O2, O3, and O5, or a combination thereof. In some embodiments, the composition includes saccharides from or derived from each of serotypes O1, O2, O3, and O5.

In another aspect, the invention includes a composition including at least one
15 saccharide that is, or is derived from, at least one *K. pneumoniae* serotype selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12; and a saccharide derived from an *E. Coli* O-antigen having a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C),
Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6),
Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g.,
20 Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula
25 O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula
30 O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D1, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79,
35 Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91,

Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 1 to 100. In some embodiments, the composition includes a saccharide from or derived from one or more of *K. pneumoniae* serotypes O1, O2, O3 and O5, or a combination thereof. In some embodiments, the composition includes a saccharide from or derived from each of *K. pneumoniae* serotypes O1, O2, O3 and O5. In some embodiments, the composition includes a saccharide derived from an *E. Coli* O-antigen having Formula O9 and does not include a saccharide derived from *K. pneumoniae* serotype O3. In some embodiments, the composition includes a saccharide derived from an *E. Coli* O-antigen having Formula O8 and does not include a saccharide derived from *K. pneumoniae* serotype O5.

In another aspect, the invention relates to a composition including a polypeptide derived from *E. coli* FimH or a fragment thereof; at least one saccharide that is, or derived from, at least one *K. pneumoniae* serotype selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12; and a saccharide having a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26,

Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49,

5 Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula

10 O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula

15 O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136,

20 Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula

25 O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein *n* is an integer from 1 to 100, preferably from 31 to

30 90. In some embodiments, the composition includes a saccharide derived from an *E. Coli* O-antigen having Formula O9 and does not include a saccharide derived from *K. pneumoniae* serotype O3. In some embodiments, the composition includes a saccharide derived from an *E. Coli* O-antigen having Formula O8 and does not include a saccharide derived from *K. pneumoniae* serotype O5.

In some embodiments, the composition includes at least one saccharide derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2, O3, and O5.

In some embodiments, the composition includes at least one saccharide derived from *K. pneumoniae* type O1. In one aspect of this embodiment, the *K. pneumoniae* O-antigen is selected from subtype v1 (O1v1) or subtype v2 (O1v2). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is selected from subtype v1 (O1v1) and subtype v2 (O1v2). In some embodiments, the composition includes at least one saccharide derived from *K. pneumoniae* type O2. In one aspect of this embodiment, the *K. pneumoniae* O-antigen is selected from subtype v1 (O2v1) or subtype v2 (O2v2). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is selected from subtype v1 (O2v1) and subtype v2 (O2v2). In another aspect, the *K. pneumoniae* O-antigen is selected from the group consisting of: a) serotype O1 subtype v1 (O1v1), b) serotype O1 subtype v2 (O1v2), c) serotype O2 subtype v1 (O2v1), and d) serotype O2 subtype v2 (O2v2). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is subtype v1 (O1v1). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is subtype v2 (O1v2). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is subtype v1 (O2v1). In one aspect of this embodiment, the *K. pneumoniae* O-antigen is subtype v2 (O2v2). In another aspect of this embodiment, the composition comprises one, two, three or four *K. pneumoniae* O-antigen selected from the group consisting of: a) serotype O1 subtype v1 (O1v1), b) serotype O1 subtype v2 (O1v2), c) serotype O2 subtype v1 (O2v1), and d) serotype O2 subtype v2 (O2v2). In some embodiments, the composition includes a combination of saccharides derived from *K. pneumoniae*, wherein a first saccharide is derived from any one of *K. pneumoniae* types selected from the group consisting of O1, O2, O3, and O5; and a second saccharide is derived from a saccharide is derived from any one of *K. pneumoniae* types selected from the group consisting of O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12. For example, in some embodiments, the composition includes at least one saccharide derived from *K. pneumoniae* type O1 and at least one saccharide derived from *K. pneumoniae* type O2. In a preferred embodiment, the saccharide derived from *K. pneumoniae* is conjugated to a carrier protein; and the saccharide derived from *E. coli* is conjugated to a carrier protein.

In another aspect, the invention includes a composition including a polypeptide derived from *E. coli* FimH or a fragment thereof; and at least one saccharide derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2, O3, and O5.

In another aspect, the invention includes at least one saccharide derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2, O3, and O5; and at least one saccharide derived from *E. coli* having a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4

(e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, 5 Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, 10 Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, 15 Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula 20 O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula 25 O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, 30 Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula 35 O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187. In some embodiments, the

- composition includes a saccharide derived from an *E. coli* O-antigen having Formula O9 and does not include a saccharide derived from *K. pneumoniae* serotype O3. In some embodiments, the composition includes a saccharide derived from an *E. coli* O-antigen having Formula O8 and does not include a saccharide derived from *K. pneumoniae* serotype O5.
- 5 In some embodiments, the composition includes at least one saccharide derived from *K. pneumoniae* type O1; and at least one saccharide derived from *E. coli* having a structure selected from the group consisting of Formula O8 and Formula O9. In another embodiment, the composition includes at least one saccharide derived from *K. pneumoniae* type O2; and at least one saccharide derived from *E. coli* having a structure selected from the group consisting of Formula O8 and Formula O9. In another embodiment, the composition includes at least one saccharide derived from *K. pneumoniae* type O1; at least one saccharide derived from *K. pneumoniae* type O2; and at least one saccharide derived from *E. coli* having a structure selected from the group consisting of Formula O8 and Formula O9.
- 10 In one embodiment, the invention provides a method of inducing an immune response to *K. pneumoniae* in a subject comprising administering to the subject an immunologically effective amount of an immunogenic composition comprising at least one glycoconjugate from *E. coli* serotype O8 or O9, wherein said immunogenic composition does not comprise glycoconjugates from *K. pneumoniae* serotype O5 or O3. In one aspect, the composition includes a saccharide derived from an *E. coli* O-antigen having Formula O8 and does not include a saccharide derived from *K. pneumoniae* serotype O5. In another aspect, the composition includes a saccharide derived from an *E. coli* O-antigen having Formula O9 and does not include a saccharide derived from *K. pneumoniae* serotype O3.
- 15 In another embodiment, the invention provides a method of inducing an immune response to *E. coli* in a subject comprising administering to the subject an immunologically effective amount of an immunogenic composition comprising at least one glycoconjugate from *K. pneumoniae* serotype O5 or O3, or a variant thereof, wherein said immunogenic composition does not comprise glycoconjugates from *E. coli* serotype O8 or O9. In one aspect, the composition includes a saccharide derived from *K. pneumoniae* serotype O5 and does not include a saccharide derived from an *E. coli* O-antigen having Formula O8. In another aspect, the composition includes a saccharide derived from *K. pneumoniae* serotype O3 and does not include a saccharide derived from an *E. coli* O-antigen having Formula O9.
- 20 In some embodiments, the composition includes at least one saccharide that is, or is derived from, at least one *K. pneumoniae* serotype selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12; at least one saccharide derived from *E. coli* having a structure selected from the group consisting of Formula O8 and
- 25 30 35

Formula O9. In some embodiments, the composition includes at least one saccharide that is, or derived from, at least one *K. pneumoniae* serotype selected from O1 (and d-Gal-III variants), O2 (and d-Gal-III variants), O2ac, O3, O4, O5, O7, O8, and O12; at least one saccharide derived from *E. coli* having a structure selected from the group consisting of Formula O1A, Formula O1B, Formula O2, Formula O6, and Formula O25B.

In some embodiments, the composition further includes a polypeptide derived from *K. pneumoniae* selected from a polypeptide derived from *K. pneumoniae* Type I fimbrial protein or an immunogenic fragment thereof; or a polypeptide derived from *K. pneumoniae* Type III fimbrial protein or an immunogenic fragment thereof, or a combination thereof. The sequences of said polypeptides are known in the art.

IX. Uses of the Compositions

In one aspect, the disclosure provides use of the compositions described herein, nucleic acids encoding *E. coli* FimH polypeptides or vectors for expressing them, or compositions comprising a polypeptide or nucleic acids as a medicament, or in the manufacture of a medicament, for eliciting an immune response against *E. coli* or *K. pneumoniae* or for preventing *E. coli* or *K. pneumoniae* infection in a subject.

In other aspects, the present disclosure provides a method of eliciting an immune response against *E. coli* or *K. pneumoniae* in a subject, such as a human, comprising administering to the subject an effective amount of a composition described herein or a nucleic acid molecule encoding an *E. coli* or *K. pneumoniae* polypeptide. The present disclosure also provides a method of preventing *E. coli* or *K. pneumoniae* infection in a subject, comprising administering to the subject an effective amount of a pharmaceutical composition, such as a vaccine, comprising a *E. coli* or *K. pneumoniae* composition described herein.

As used herein, "subject" means a mammal, preferably a human. In some particular embodiments, the human is a child, such as an infant. In some other particular embodiments, the human is a woman, particularly a pregnant woman. The compositions of the invention may be administered to the subject with or without administration of an adjuvant. The effective amount administered to the subject is an amount that is sufficient to elicit an immune response against an *E. coli* or *K. pneumoniae* antigen in the subject. Subjects that can be selected for treatment include those that are at risk for developing an *E. coli* or *K. pneumoniae* infection because of exposure or the possibility of exposure to *E. coli* or *K. pneumoniae*. Because humans may be infected with *E. coli* or *K. pneumoniae* by the age of 2, the entire birth cohort is included as a relevant population

for immunization. This could be done, for example, by beginning an immunization regimen anytime from birth to 6 months of age, from 6 months of age to 5 years of age, in pregnant women (or women of child-bearing age) to protect their infants by passive transfer of antibody, infants still *in utero*, and subjects greater than 50 years of age.

5 Administration of the compositions provided by the present disclosure, such as pharmaceutical compositions, can be carried out using standard routes of administration. Non-limiting embodiments include parenteral administration, such as intradermal, intramuscular, subcutaneous, transcutaneous, mucosal, or oral administration.

The total dose of the composition provided to a subject during one administration can be
10 varied as is known to the skilled practitioner.

It is also possible to provide one or more booster administrations of one or more of the vaccine compositions. If a boosting vaccination is performed, typically, such a boosting vaccination will be administered to the same subject at a moment between one week and 10 years, preferably between two weeks and six months, after administering the composition to the
15 subject for the first time (which is in such cases referred to as "priming vaccination"). In alternative boosting regimens, it is also possible to administer different vectors, e.g., one or more adenovirus, or other vectors such as modified vaccinia virus of Ankara (MVA), or DNA, or protein, to the subject after the priming vaccination. It is, for instance, possible to administer to the subject a recombinant viral vector hereof as a prime, and boosting with a composition
20 described herein.

In certain embodiments, the administration comprises a priming administration and at least one booster administration. In certain other embodiments, the administration is provided annually. In still other embodiments, the administration is provided annually together with an influenza vaccine.

25 The vaccines provided by the present disclosure may be used together with one or more other vaccines. For example, in adults they may be used together with an influenza vaccine, Prevnar, tetanus vaccine, diphtheria vaccine, and pertussis vaccine. For pediatric use, vaccines provided by the present disclosure may be used with any other vaccine indicated for pediatric patients.

30 EXAMPLES

In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only and are not to be construed as limiting the

scope of the invention in any manner. The following Examples illustrate some embodiments of the invention.

EXAMPLE 1

Bacterial fimbrial adhesins FimH and FmlH allow *Escherichia coli* to exploit distinct urinary tract microenvironments through recognition of specific host cell glycoproteins. FimH binds to mannosylated uroplakin receptors in the uroepithelium whereas FmlH binds to galactose or N-acetylgalactosamine O-glycans on epithelial surface proteins in the kidney and inflamed bladder. FimH fimbriae also play a role in colonization of enterotoxigenic *E. coli* (ETEC) and multidrug-resistant invasive *E. coli* in the gut through binding to highly mannosylated proteins on the intestinal epithelia.

Full length FimH is composed of two domains: the N-terminal lectin domain and the C-terminal pilin domain, which are connected by a short linker. The lectin domain of FimH contains the carbohydrate recognition domain, which is responsible for binding to the mannosylated uroplakin 1a on the urothelial cell surface. The pilin domain is anchored to the core of the pilus via a donor strand of the subsequent FimG subunit, which is a process termed donor strand complementation.

Conformation and ligand-binding properties of the lectin domain of FimH are under the allosteric control of the pilin domain of FimH. Under static conditions, the interaction of the two domains of full length FimH stabilizes the lectin domain in the low-affinity to monomannose (for example, $K_d \sim 300 \mu\text{M}$) state, which is characterized by a shallow binding pocket. Binding to a mannoside ligand induces a conformational change leading to a medium affinity state, where the lectin and pilin domains remain in close contact. However, upon shear stress, the lectin and pilin domains separate, thereby inducing the high-affinity state (for example, $K_d < 1.2 \mu\text{M}$).

Because of the absence of negative allosteric regulation exerted by the pilin domain, the isolated lectin domain of FimH is locked in the high-affinity state. The isolated, recombinant lectin domain, which is locked in the high-affinity state, exhibits high stability. Locking the adhesin in a low-binding conformation, however, induces the production of adhesion-inhibiting antibodies. Accordingly, there is an interest in stabilizing the lectin domain in the low-affinity state.

Table 3 sets forth the FimH constructs used to make various constructs to address these needs.

Table 3: Summary of FimH constructs

Construct	Plasmid	Signal sequence	Protein Description	Linker	Additional protein variant	Backbone	Mass
FimH lectin domain	pSB01877	FimH signal sequence	FimH J96 F22..G181	none	none	pcDNA3.1(+) or pCAG vector	
FimH lectin domain	pSB01878	mlgK signal sequence	FimH J96 F22..G181	none	none	pcDNA3.1(+) or pCAG vector	Fully reduced mass, with His-tag: 18117.48 Observed non-reduced mass, with His-tag: 18117.90 Mass without tag: 17022.08
FimH/C	pSB01879	FimH signal sequence	FimH J96 F22..Q300	none	none	pBudCE4.1 Dual promoter vector (CMV & EF1 α)	

FimH/C	pSB01880	mlgK signal sequence	FimH J96 F22..Q300	none	none	pBudCE4.1 Dual promoter vector (CMV & EF1 α)	
FimH/C	pSB01881	mlgK signal sequence	FimC G37..E241 (according to SEQ ID NO: 18)	none	none	pBudCE4.1 Dual promoter vector (CMV & EF1 α)	
FimH- dscG	pSB01882	FimH signal sequence	FimH J96 F22..Q300	DNKQ	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01883	FimH signal sequence	FimH J96 F22..Q300	GGSGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01884	FimH signal sequence	FimH J96 F22..Q300	GGSSGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01885	FimH signal sequence	FimH J96 F22..Q300	GGSSGGG	FimG A1..K14 (SEQ ID NO: 17)	N-terminus residue at W20, and therefore does not appear to have been processed at preferred position; a small amount of	

						protein present exhibiting preferred processing as indicated by the small amount of the FACK peptide being detected	
FimH- dscG	pSB01886	FimH signal sequence	FimH J96 F22..Q300	GGGSSGGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01887	FimH signal sequence	FimH J96 F22..Q300	GGGSGSGGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01888	FimH signal sequence	FimH J96 F22..Q300	GGGSGSGGGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01889	mlgK signal sequence	FimH J96 F22..Q300	DNKQ	FimG A1..K14 (SEQ ID NO: 17)		
FimH- dscG	pSB01890	mlgK signal sequence	FimH J96 F22..Q300	GGSGG	FimG A1..K14 (SEQ ID NO: 17)		

FimH-dscG	pSB01891	mlgK signal sequence	FimH J96 F22..Q300	GGSSGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH-dscG	pSB01892	mlgK signal sequence	FimH J96 F22..Q300	GGSSGGG	FimG A1..K14 (SEQ ID NO: 17)	appears to have had the signal peptide processed with the F22 being the preferred N-terminal residue; identity of the peptide was confirmed by MS/MS	
FimH-dscG	pSB01893	mlgK signal sequence	FimH J96 F22..Q300	GGGSSGGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH-dscG	pSB01894	mlgK signal sequence	FimH J96 F22..Q300	GGGSGSGGG	FimG A1..K14 (SEQ ID NO: 17)		
FimH-dscG	pSB01895	mlgK signal sequence	FimH J96 F22..Q300	GGGSGGSGGG	FimG A1..K14 (SEQ ID NO: 17)		

FimH lectin domain	pSB02081	mlgK signal sequence	F22..G181 J96 FimH N28Q N91S / His8 in pcDNA3.1(+)				
FimH lectin domain	pSB02082	mlgK signal sequence	F22..G181 J96 FimH N28Q N91S / His8 in pcDNA3.1(+)				
FimH lectin domain	pSB02083	mlgK signal sequence	F22..G181 J96 FimH N28S N91S / His8 in pcDNA3.1(+)				
FimH lectin domain	pSB02088	mlgK signal sequence	F22..G181 J96 FimH V48C L55C / His8 in pcDNA3.1(+)				
FimH lectin domain	pSB02089	mlgK signal sequence	F22..G181 J96 FimH N28Q V48C L55C N91S / His8 in pcDNA3.1(+)				
FimH lectin domain	pSB02158	mlgK signal sequence	F22..G181 J96 FimH N28S V48C L55C N91S / His8 in pcDNA3.1(+)				
FimH- dscG	pSB02159						
FimH- dscG	pSB02198	mlgK signal sequence	FimH mlgK signal pept / F22..Q300				

			J96 FimH N28S V48C L55C N91S N249Q / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02199	mlgK signal sequence	FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S N256Q / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02200	mlgK signal sequence	FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S N249Q N256Q / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02304	mlgK signal sequence	FimH mlgK signal pept / F22..Q300				

			J96 FimH N28S V48C L55C N91S T251A / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02305	mlgK signal sequence	FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S T258A / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02306	mlgK signal sequence	FimH mlgK signal pept / F22..Q300 J96 FimH N28S V48C L55C N91S T251A T258A / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02307	mlgK signal sequence	FimH mlgK signal pept / F22..Q300				

			J96 FimH N28S N91S N249Q / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				
FimH- dscG	pSB02308	mlgK signal sequence	FimH mlgK signal pept / F22..Q300 J96 FimH N28S N91S N256Q / 7 AA linker / FimG A1..K14 / GGHis8 in pcDNA3.1(+)				

All of the FimH constructs studied were monomeric proteins of expected molecular weight.

Table 4

Protein	Sedimentation Coefficient, S	M _{w,app}	M _{w, expected}	Homogeneity
E.coli expression				
<i>Cytosolic FimH-LD</i>	1.9 S	18 kDa	18 kDa	98%
<i>Periplasmic FimH-LD</i>	1.9 S	18 kDa	18 kDa	98%
<i>FimH-LD lock mutant</i>	2.0 S	19 kDa	18 kDa	97%
Mammalian expression				
<i>FimH-LD</i>	1.9 S	18 kDa	18 kDa	>99%
<i>FimH-LD lock mutant</i>	1.9 S	18 kDa	18 kDa	98%
<i>FimH wild type</i>	2.7 S	36 kDa	34 kDa	96%
<i>FimH lock mutant</i>	2.7 S	34 kDa	34 kDa	94%

Expected molecular weight of FimC-FimH complex is 53.1 kDa;

Expected molecular weight of FimC is 24 kDa.

EXAMPLE 2: Mammalian expression of FimH lectin binding domain

5 The present non-limiting example relates to producing a polypeptide derived from *E. coli* or a fragment thereof in a HEK cell line. The yields were relatively high, as compared to expression of the polypeptide derived from *E. coli* or a fragment thereof in an *E. coli* host cell.

 To accomplish the production of FimH variants from mammalian cells, a SignalP prediction algorithm was used to analyze different heterologous signal sequences for secretion
10 of proteins and fragments. The wild type FimH leader sequence was also analyzed. The predictions indicated that the wild type FimH leader sequence may work for secretion of the FimH variants in mammalian cells, however, the secreted variant was predicted to be cleaved at the W20 residue of the full-length wild type FimH (see SEQ ID NO: 1), rather than the F22 residue of the full-length wild type FimH (see SEQ ID NO: 1). A hemagglutinin signal sequence
15 was predicted not to work. The murine IgK signal sequence was predicted to produce an N-terminus of F22 of SEQ ID NO: 1, or F1 residue of the mature protein.

 Based on these analyses, DNA was synthesized and recombinantly produced constructs to express the FimH lectin binding domain with the wild-type FimH leader. Constructs were also prepared to express the FimH lectin binding domain with the mIgK signal
20 sequence. Affinity purification tags, such as His tag, were introduced to the C-terminus of the polypeptide derived from *E. coli* or a fragment thereof to facilitate purification.

 The expression plasmid was transfected into HEK host cells, namely EXP1293 mammalian cells.

 The polypeptides or fragments thereof derived from *E. coli* were successfully expressed.
25 For example, the preferred N-terminal processing using the mIgK signal sequence fused to the mature start of FimH at F22 was demonstrated for the pSB01892 FimHdscG construct by MS. The processing is believed correct for the lectin domain construct pSB01878 and the mass spec data supports this.

 The preferred N-terminal processing (i.e., processing at F22 of SEQ ID NO: 1) was not
30 shown with the native FimH leader peptide.

 pSB01877 and pSB01878 constructs are in pcDNA3.1(+) mammalian expression vectors. The cells were diluted and subsequently used in 20 ml transfections. 1 ug/ml DNA for each construct was used and transfected cells in 125ml flasks using Expifectamine protocol. After 72 hours, the cell viability was still good so the expression was allowed to continue until 96
35 hours. Samples were taken at 72 hours and ran 10 ul of each on SDS PAGE gels to check for expression.

After 96 hours, conditioned media was harvested and 0.25ml of Nickel Excel resin was added with batch binding O/N at 4°C with rotation. Eluted in TrisCl pH8.0, NaCl, imidazole. See **FIG. 4**.

5 pSB01878 has expected mass consistent with N-terminal F22. Glycosylation present on 1 or 2 sites (+1 mass from each deamidation of N-D).

Glycosylation mutants were constructed. See, for example, pSB02081, pSB02082, pSB02083, pSB02088, and pSB02089. The glycosylation mutants expressed the polypeptides of interest. See **FIG. 5** for results.

10 A FimH lectin domain lock mutant was also constructed. See, for example, pSB02158. Results of the expression of the pSB02158 construct is shown in **FIG. 6B**.

Fluorescence polarization assay using 0.5 pmoles fluorescein-conjugated aminophenyl-mannopyranoside (APMP). The assay was performed at room temperature, 300 RPM for 64 hrs. Results shown in **FIG. 6C**.

15 **EXAMPLE 3: Mammalian expression of FimH/C complex, pSB01879 and pSB01880**

For production of the FimH/C complex, dual expression constructs of the FimC under the EF1alpha promoter and the FimH with either the wild type or mlgK signal peptide were prepared. These were cloned into a pBudCE4.1 mammalian expression vector (ThermoFisher) and a C-term His tag was added to the FimC. The FimC variant
20 was designed for secretion using the mlgK signal peptide as it resulted in a positive prediction to yield the G37 FimC as the first residue of the mature protein based on SignalP analysis.

More specifically, these constructs were designed to have the FimC fragment under the EF1alpha promoter in the vector pBudCE4.1 and the FimH fragment inserts
25 under the CMV promoter in the same vector. The vector pBudCE4.1 is an expression vector from Thermo Fisher that has 2 promoters for expression in mammalian cells. The FimC fragment insert (pSB01881 insert) was subcloned by digesting with NotI and XhoI and subcloning into the pBudCE4.1 vector at the same sites. These were plated onto 2xYT zeocin 50 ug/ml plates. Colonies were inoculated into 2xYT with zeocin 50ug/ml,
30 grew overnight at 37°C and plasmid prepped. These were digested with NotI and XhoI to check for insert and all colonies had insert size of ~722 bp.

pSB01881 was digested with HindIII and BamHI and the pSB01879 insert and pSB01880 insert DNA was digested with HindIII and BamHI. These fragments were gel isolated and subcloned into the pSB01881 vector and plated onto 2xYTzeo50 ug/ml
35 plates. Colonies from each were inoculated into 2xYT zeo50ug/ml, grown overnight at 37°C, plasmid prepped and digested with NotI and XhoI to test for FimC insert and

HindIII and BamHI to test for FimH inserts. All clones had expected sized inserts at both cloning sites. The pSB01879-1 and pSB01880-1 clones were subsequently used for expression.

The FimH/FimC complex has been demonstrated to express in EXP1293 cells as well. Expression may be optimized by switching promoters, such as EF1 α , CAG, Ub, Tub, or other promoters.

The preferred N-terminal processing (i.e., processing at F22 of SEQ ID NO: 1) was not shown with the native FimH leader peptide.

Exemplary results from SignalP 4.1 (DTU Bioinformatics) used for signal peptide predictions are shown below. Additional signal peptides are predicted to produce the preferred N-terminus of Phe at position 1 of the mature FimH polypeptide or fragment thereof. The following is only a representative sample set of 4 common signal sequences.

The following signal peptide sequences were predicted to yield the preferred N-terminus of Phe at position 1 of the mature FimH polypeptide or fragment thereof:

15

Table 5

	signal peptide sequence	SEQ ID NO:
>sp P55899 FCGRN_HUMAN IgG receptor FcRn large subunit p51 OS=Homo sapiens OX=9606 GN=FCGRT PE=1 SV=1	MGVPRPQPWALGLLLFLLPGSLG	SEQ ID NO: 55
>tr Q6FGW4 Q6FGW4_HUMAN IL10 protein OS=Homo sapiens OX=9606 GN=IL10 PE=2 SV=1	MHSSALLCCLVLLTGVRA	SEQ ID NO: 56

The following signal peptide sequences were NOT predicted to yield the preferred N-terminus of Phe at position 1 of the mature FimH polypeptide or fragment thereof:

Table 6

	signal peptide sequence	SEQ ID NO:

>sp P03420 FUS_HRSVA Fusion glycoprotein F0 OS=Human respiratory syncytial virus A (strain A2) OX=11259 GN=F PE=1 SV=1	MELLILKANAITTILTAVTFCFASG	SEQ ID NO: 57
>sp P03451 HEMA_I57A0 Hemagglutinin OS=Influenza A virus (strain A/Japan/305/1957 H2N2) OX=387161 GN=HA PE=1 SV=1	MAIYLILLFTAVRG	SEQ ID NO: 58

Table 7

SignalP 4.1 used for predictions

	Fusion sequence
>sp P55899 FCGRN_HUMAN IgG receptor FcRn large subunit p51 OS=Homo sapiens OX=9606 GN=FCGRT PE=1 SV=1 MGVPRQPWALGLLLFLPGSLGAESHSLLY HLTAVSSPAPGTPAFWVSGWLGPPQYLS YNSLRGEAEPAGAWVWENQVSWYWEKETT DLRIKEKLFLEAFKALGGKGPYTLQGLLGCE LGPDNTSVPTAKFALNGEEFMNFDLKQGTWG GDWPEALAIQRWQQQDKAANKELTFLF SCPHRLREHLERGRGNLEWKEPPSMRLKARPS SPGFSVLTCASFYPPELQLRFLRNL AAGTGQGDGFPNSDGSFHASSSLTVKSGDEH HYCCIVQHAGLAQPLRVELESPAKSSVLV VGIVIGVLLLTAAAVGGALLWRRMRSGLPAP WISLRGDDTGVLPTPGEAQDADLKDVNV IPATA (SEQ ID NO: 102)	The signal peptide from the protein listed in the respective left column is shown below in CAPITAL LETTERS. The N-terminus of FimH is depicted in lower case. MGVPRQPWALGLLLFLPGSLGfacktangtaipiggg anvyvnlapvvvngqnlvldls (SEQ ID NO: 103)
	# Measure Position Value Cutoff signal peptide? max. C 24 0.664 max. Y 24 0.788 max. S 9 0.966 mean S 1-23 0.935 D 1-23 0.867 0.450 YES Name=Sequence SP='YES' Cleavage site between pos. 23 and 24: SLG-FA D=0.867 D-cutoff=0.450 Networks=SignalP-noTM

>sp P03420 FUS_HRSVA Fusion glycoprotein F0 OS=Human respiratory syncytial virus A (strain A2) OX=11259 GN=F PE=1 SV=1 MELLILKANAIITILTAVTFCFASGQNITEEFYQ STCSAVSKGYLSALRTGWYTSVITIE LSNIKENKCGNGTDAKVKLIKQELDKYKNAVTE LQLLMQSTPPTNNRARRELPRFMNYTLN NAKKTNTVLSKKRKRRLGFLGVSASIASG VAVSKVLHLEGEVNIKSALLSTNKAVVS LSNGVSVLTSKVLDLKNYIDKQLLPIVKNQSC SISNIETVIEFQQKNNRLEITREFSVN AGVTTPVSTYMLTNSELLSLINDMPITNDQKK LMSNNVQIVRQQSYSIMSIIKEEVLAYV VQLPLYGVIDTPCWKLHTSPLCTTNTKEGSNI CLTRTDRGWYCDNAGSVSFFPQAETCKV QSNRVFCDTMNSLTLPSEINLCNVDIFNPKYD CKIMTSKTDVSSSVITSLGAIVSCYGKT KCTASNKNRGIKTFSGNGCDYVSNGKMDTVS VGNTLYYVKNQEGKSLYVKGEPIINFYDP LVFPSDEFDASISQVNEKINQSLAFIRKSDELL HNVNAGKSTTNIMITIIIVIIIVILLS LIAVGLLLYCKARSTPVTLSKDQLSGINNIAFS N (SEQ ID NO: 104)	The signal peptide from the protein listed in the respective left column is shown below in CAPITAL LETTERS. The N-terminus of FimH is depicted in lower case. MELLILKANAIITILTAVTFCFASGfacktangtaipiggg-sanvy vnlapvvvnvgqnlvvdls (SEQ ID NO: 105) <table><tr><th>#</th><th>Measure</th><th>Position</th><th>Value</th><th>Cutoff</th><th>signal peptide?</th></tr><tr><td>max. C</td><td>28</td><td>0.188</td><td></td><td></td><td></td></tr><tr><td>max. Y</td><td>28</td><td>0.263</td><td></td><td></td><td></td></tr><tr><td>max. S</td><td>11</td><td>0.478</td><td></td><td></td><td></td></tr><tr><td>mean S</td><td>1-27</td><td>0.387</td><td></td><td></td><td></td></tr><tr><td>D</td><td>1-27</td><td>0.312</td><td>0.500</td><td>NO</td><td></td></tr></table> Name=Sequence SP='NO' D=0.312 D-cutoff=0.500 Networks=SignalP-TM	#	Measure	Position	Value	Cutoff	signal peptide?	max. C	28	0.188				max. Y	28	0.263				max. S	11	0.478				mean S	1-27	0.387				D	1-27	0.312	0.500	NO	
#	Measure	Position	Value	Cutoff	signal peptide?																																
max. C	28	0.188																																			
max. Y	28	0.263																																			
max. S	11	0.478																																			
mean S	1-27	0.387																																			
D	1-27	0.312	0.500	NO																																	
>tr Q6FGW4 Q6FGW4_HUMAN IL10 protein OS=Homo sapiens OX=9606 GN=IL10 PE=2 SV=1 MHSSALLCCLVLLTGVRASPGQGTQSENSC THFPGNLPNMLRDLRDAFSRVKTFQMKDQ	The signal peptide from the protein listed in the respective left column is shown below in CAPITAL LETTERS. The N-terminus of FimH is depicted in lower case. MHSSALLCCLVLLTGVRafacktangtaipiggg-sanvyvnlapv nvgqnlvvdls (SEQ ID NO: 107) <table><tr><th>#</th><th>Measure</th><th>Position</th><th>Value</th><th>Cutoff</th><th>signal peptide?</th></tr><tr><td>max. C</td><td>19</td><td>0.726</td><td></td><td></td><td></td></tr></table>	#	Measure	Position	Value	Cutoff	signal peptide?	max. C	19	0.726																											
#	Measure	Position	Value	Cutoff	signal peptide?																																
max. C	19	0.726																																			

LDNLLLKESLLEDFKGYLGCCQALSEMIQFYLE EVMPQAENQDPDIKAHVNSLGENLKTLR LRLRRCHRFLPCENKSKAVEQVKNAFNKLQE KGIYKAMSEFDIFINYIEAYMTMKIRN (SEQ ID NO: 106)	max. Y190.829 max. S40.973 mean S1-180.947 D1-180.8930.450YES Name=SequenceSP='YES' Cleavage site between pos. 18 and 19: VRA-FA D=0.893 D-cutoff=0.450 Networks=SignalP-noTM
>sp P03451 HEMA_I57A0 Hemagglutinin OS=Influenza A virus (strain A/Japan/305/1957 H2N2) OX=387161 GN=HA PE=1 SV=1 MAIYLILLFTAVRGDQICIGYHANNSTEKVDTNLERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVILPKDRWTQHTTTGGSRACAVSGNPSFFRNMVWLTKEGSDYPVAKGSYNNTSGEQMLIIWGVHHPIDETEQRITLYQNVGTYSVSGTSTLNKRSTPEIATRPKVNGQGGRMEFSWTLTDMWDTINFESTGNLIAPEYGFKISKRSSGIMKTEGTLENCETKCQTPLGAINTTLPFHNHPLTIGECPKYVKSEKLVLATGLRNVPQIESRGLFGAIAGFIEGGWQGMVDGWYGYHHSNDQGSYAADKESTQKAFDGITNKVNSVIEKMNTQFEAVGKEFGNLERRLENLNKRMEGGLDVWTYNAELLVLMENERTLDFHDSNVKNLYDKVRMQLRDNVKELGNGCFEFYHKCDDECMNSVKNGTYDYPKYEEESKLNRNEIKGVKLSSMGVYQILAIYATVAGSLSLA	The signal peptide from the protein listed in the respective left column is shown below in CAPITAL LETTERS. The N-terminus of FimH is depicted in lower case. MAIYLILLFTAVRGfacktangtaipigggsanvyvnlapvvngqnlvvdls (SEQ ID NO: 109) #MeasurePositionValueCutoffsignal peptide? max. C180.524 max. Y180.690 max. S10.951 mean S1-170.895 D1-170.8000.450YES Name=SequenceSP='YES' Cleavage site between pos. 17 and 18: GFA-CK D=0.800 D-cutoff=0.450 Networks=SignalP-noTM

IMMAGISFWMCSNGSLQCRICI (SEQ ID NO: 108)	
---	--

EXAMPLE 4: Mammalian Expression of Donor Strand Complement Fusion of FimH with the FimG peptide

Several linker lengths were tested. Recombinant expression with these linkers fusing the FimH to the N-terminal FimG peptide in both the wild type FimH and the mlgK signal peptide fused to F22 of FimH were prepared.

The FimH donor strand complement FimG constructs have also been shown to have robust expression in EXPI293 cells.

The preferred N-terminal processing (i.e., processing at F22 of SEQ ID NO: 1) was not shown with the native FimH leader peptide.

For the donor strand complement constructs, oligonucleotides were designed to produce base constructs in pcDNA3.1(+) that contained the various linkers and FimG peptide. A unique BstEII site was incorporated at G294 V295 T296 residues, according to the numbering of SEQ ID NO: 1 of FimH. The same BstEII site was incorporated in the linkers to produce base constructs.

The base constructs for pSB01882-01895 were constructed. Primers were used to PCR amplify pcDNA3.1(+) with ACCUPRIME PFX DNA Polymerase (Thermo Fisher), digest the PCR products with NdeI (in CMV promoter) and BamHI and cloned into pcDNA3.1(+) that was digested with NdeI and BamHI and gel isolated to remove the fragment.

Another transient transfection was performed with pSB01877, 01878, 01879, 01880, 01885, and 01892 alongside EXPI293 cells as control.

Constructs pSB01882 through pSB01895 were used in transient transfection expression tests in EXPI293 cells from Thermo Fisher as per the manufacturer's protocol. See FIG. 3, which shows the results following expression in 20 mL EXPI293 cells, 72 hours, 10 ul of conditioned media loaded; high levels of expression observed; the FimH/FimC complex present following expression from pSB01879 & pSB01880 constructs; 20 ml conditioned media batch bound to Nickel Excel, 40 CV wash, elution in Imdidazole.

Additional FimH-donor strand complement constructs were prepared. See, for example, pSB02198, pSB02199, pSB02200, pSB02304, pSB02305, pSB02306, pSB02307, pSB02308 constructs. The expression of pSB2198 FimH dscG lock mutant construct is shown in FIG. 7. The pSB2198 FimH dscG Lock Mutant yielded 12 mg/L from transient expression.

According to Vi-CELL XR 2.04 (Beckman Coulter, Inc.), the following were observed (actual cell type used for expression was HEK cells):

Table 8

Sample	Cell type parameter entered	Viability (%)	Total cells/ml ($\times 10^6$)	Viable cells/ml ($\times 10^6$)	Avg. diam. (microns)
EXPI P13	CHO	97.8	3.56	3.48	19.33
pSB01882	CHO	90.9	4.98	4.53	17.39
pSB01889	CHO	89.2	5.23	4.67	17.14
cells	CHO	88.9	6.66	5.92	16.91
Expi Start	CHO	93.7	3.35	3.14	18.72
Samples at harvest ~ 85-86 hours after transfection:					
1877	SF-9	57.3	4.32	2.48	16.00
pSB01878	SF-9	57.6	3.88	2.24	15.49
pSB01879	SF-9	59.1	5.24	3.10	15.32
pSB01880	SF-9	56.8	5.97	3.39	15.10
pSB01885	SF-9	63.1	6.95	4.39	16.08
pSB01892	SF-9	56.2	4.89	2.75	15.91
187772	SF-9	79.5	5.14	4.09	18.36
187872	SF-9	72.6	5.26	3.81	17.35
expicont	SF-9	75.5	4.95	3.74	18.62

5 EXAMPLE 5: Molecular weight fragments are with processed signal peptide

Table 9

pSB01877 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Fragment 15-189	Entire Protein
	Length	175 aa	189 aa
	Molecular Weight	18948.34	20522.36 m.w.
	1 microgram =	52.775 pMoles	48.727 pMoles
	Molecular Extinction Coefficient	35800	35800
	1 A(280) corr. to:	0.53 mg/ml	0.57 mg/ml
	A[280] of 1 mg/ml	1.89 AU	1.74 AU

	Isoelectric Point	6.81	8
	Charge at pH7	-0.48	1.52
pSB01878 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Fragment 21-188	Entire Protein
	Length	168 aa	188 aa
	Molecular Weight	18117.48	20344.08 m.w.
	1 microgram =	55.195 pMoles	49.154 pMoles
	Molecular Extinction Coefficient	24420	35800
	1 A(280) corr. to:	0.74 mg/ml	0.57 mg/ml
	A[280] of 1 mg/ml	1.35 AU	1.76 AU
	Isoelectric Point	6.81	6.29
	Charge at pH7	-0.48	-2.47
pSB01885 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Fragment 20-331	Entire Protein
	Length	312 aa	331 aa
	Molecular Weight	32406.19	34537.79 m.w.
	1 microgram =	30.858 pMoles	28.954 pMoles
	Molecular Extinction Coefficient	38030	43720
	1 A(280) corr. to:	0.85 mg/ml	0.79 mg/ml
	A[280] of 1 mg/ml	1.17 AU	1.27 AU
	Isoelectric Point	7.25	8.32
	Charge at pH7	0.5	2.5
pSB01892 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Fragment 21-330	Entire Protein
	Length	310 aa	330 aa
	Molecular Weight	32132.91	34359.51 m.w.
	1 microgram =	31.121 pMoles	29.104 pMoles
	Molecular Extinction Coefficient	32340	43720
	1 A(280) corr. to:	0.99 mg/ml	0.79 mg/ml

	A[280] of 1 mg/ml	1.01 AU	1.27 AU
	Isoelectric Point	7.25	6.51
	Charge at pH7	0.5	-1.49
pSB01893 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Entire Protein	
	Length	331 aa	
	Molecular Weight	34416.56	
	1 microgram =	29.056 pMoles	
	Molecular Extinction Coefficient	43720	
	1 A(280) corr. to:	0.79 mg/ml	
	A[280] of 1 mg/ml	1.27 AU	
	Isoelectric Point	6.51	
	Charge at pH7	-1.49	
pSB01894 FimH J96 ELL41155.1 [E. coli J96]			
Analysis			
	Analysis	Fragment 21-332	Entire Protein
	Length	312 aa	332 aa
	Molecular Weight	32247.01	34473.61 m.w.
	1 microgram =	31.011 pMoles	29.008 pMoles
	Molecular Extinction Coefficient	32340	43720
	1 A(280) corr. to:	1.00 mg/ml	0.79 mg/ml
	A[280] of 1 mg/ml	1.00 AU	1.27 AU
	Isoelectric Point	7.25	6.51
	Charge at pH7	0.5	-1.49

pSB02083

Analysis	Fragment 21-188	Entire Protein
Length	168 aa	188 aa
Molecular Weight	18063.42	20290.02 m.w.
1 microgram =	55.361 pMoles	49.285 pMoles
Molar Extinction coefficient	24420	35800

1 A(280) corr. to:	0.74 mg/ml	0.57 mg/ml
A[280] of 1 mg/ml	1.35 AU	1.76 AU
Isoelectric Point	6.81	6.29
Charge at pH 7	-0.48	-2.47

pSB02198

Sample	FimH PSB 2198 1.45 mg/ml 5 ml 20190918 SS
Volume (mls)	25
Conc. (mg/ml)	1.45
Total Amount (mgs)	36.25
Aliquots	5ml x5
Yield	12mg/L
Buffer : 50 mM TrisCl pH8.0, 300 mM NaCl	

pSB02307

Sample Name	Fim H 2307 0.48 mg/ml 5 ml 20190918 SS
Volume (mls)	22.5 mls
Conc. (mg/ml)	0.48 mg/ml
Total Amount (mgs)	10.8mg
Yield	3.6mg/L
Buffer : 50 mM TrisCl pH8.0, 300 mM NaCl	

5

EXAMPLE 6: The N-terminal α -amino group of Phe1 (according to the numbering of SEQ ID NO: 2) in the FimH mature protein provides critical polar recognition for D-mannose

Without being bound by theory or mechanism, it is suggested that the correct signal peptide cleavage just ahead of Phe1 (according to the numbering of SEQ ID NO: 2) of the FimH mature protein is important to express functional FimH protein. Changes at the N-terminal α -

10

amino group, such as by adding an amino acid at the N-terminus ahead of Phe1 of the FimH protein can abolish the hydrogen bond interactions with O2-, O5- and O6-atoms of the D-mannose and introduce steric repulsion with D-mannose, thereby blocking mannose binding. This is confirmed with our experimental observation that adding an
 5 extra Gly residue ahead of the Phe1 of SEQ ID NO: 2 leads to no detection of mannose binding.

Following an analysis of the crystal structure of FimH bound to D-mannose, the following were observed: The N-Terminal α -amino group of Phe1 along with sidechains of Asp54 of the FimH according to the numbering of SEQ ID NO: 2 and Gln133 of the
 10 FimH according to the numbering of SEQ ID NO: 2 provide critical polar recognition motifs for D-mannose, and mutations and changes of these polar interactions lead to no mannose binding.

**EXAMPLE 7: The sidechain of Phe1 in FimH does not interact directly with D-mannose but is rather buried inside of FimH, suggesting that Phe1 can be replaced by other
 15 residues, e.g. aliphatic hydrophobic residues (Ile, Leu, or Val)**

Analysis of crystal structures of FimH in complex with *D*-mannose and its analogs (e.g. PDB ID: 1QUN) shows that the sidechain of Phe1 (according to the numbering of SEQ ID NO: 2) does not interact directly with D-mannose but rather stabilizes the binding pocket by stacking its aromatic rings with the sidechains of Val56,
 20 Tyr95, Gln133 and Phe144 (according to the numbering of SEQ ID NO: 2).

Alternative N-terminal residue instead of Phe may stabilize the FimH protein, accommodate mannose binding, and allow correct signal peptide cleavage. Such residues may be identified by suitable method known in the art, such as by visual inspection of a crystal structure of FimH, or more quantitative selection using
 25 computational protein design software, such as BioLuminate™ [BioLuminate, Schrodinger LLC, New York, 2017], Discovery Studio™ [Discovery Studio Modeling Environment, Dassault Systèmes, San Diego, 2017], MOE™ [Molecular Operating Environment, Chemical Computing Group Inc., Montreal, 2017], and Rosetta™ [Rosetta, University of Washington, Seattle, 2017]. An illustrative example is shown **FIG. 9A-9C**.
 30 The replacement amino acids can be aliphatic hydrophobic amino acids (e.g. Ile, Leu and Val). **FIG. 11** depicts computational mutagenesis scanning of Phe1 with other amino acids having aliphatic hydrophobic sidechains, e.g. Ile, Leu and Val, which may stabilize the FimH protein and accommodate mannose binding.

EXAMPLE 8: Mutations of Asn7 according to the numbering of SEQ ID NO: 2 in a FimH protein can remove the putative N-glycosylation site and prevent deamidation, without impacting mannose, mAb21, or mAb475 binding.

Over-expression of secreted *E. coli* FimH from mammalian cell lines may lead to N-linked glycosylation at residue Asn7, according to the numbering of SEQ ID NO: 2. In addition, residue Asn7 is solvent exposed and followed with a Gly residue, making it very prone to deamidation.

Analysis of crystal structures of FimH in complex with *D*-mannose and its analogs (e.g. PDB ID: 1QUN) indicates that Asn7 is more than 20 Å away from the mannose binding site and a mutation at the site should not impact mannose binding. Thus, mutations of Asn7 to other amino acids (e.g. Ser, Asp and Gln) can effectively remove the putative N-glycosylation site and prevent deamidation.

EXAMPLE 9: *E. coli* and *S. enterica* strains

Clinical strains and derivatives are listed in **Table 10**. Additional reference strains included: O25K5H1, a clinical O25a serotype strain; and *S. enterica* serovar Typhimurium strain LT2.

Gene knockouts in *E. coli* strains removing the targeted open-reading frame but leaving a short scar sequence were constructed.

The hydrolyzed O-antigen chain and core sugars are indicated subsequently as O-Polysaccharide (OPS) for simplicity.

Table 10: *E. coli* Strains

Strain	Strain Alias	Genotype	Serotype
GAR2401	PFEEC0100	wt (blood isolate)	O25b
'2401ΔwzzB	--	ΔwzzB	O25b
'2401ΔAraAΔ(OPS)	--	ΔAraA Δ(rflB-wzzB)	OPS-
O25K5H1	PFEEC0101	wt	O25a
O25K5H1ΔwzzB		ΔwzzB	O25a
BD559	--	W3110 ΔAraA ΔfhuA ΔrecA	OPS-
BD559ΔwzzB	--	W3110ΔAraA ΔfhuA ΔrecAΔwzzB	OPS-
BD559Δ(OPS)	--	BD559 Δ(rflB-wzzB)	OPS-
GAR2831	PFEEC0102	wt (blood isolate)	O25b
GAR865	PFEEC0103	wt (blood isolate)	O2
GAR868	PFEEC0104	wt (blood isolate)	O2
GAR869	PFEEC0105	wt (blood isolate)	O15

Strain	Strain Alias	Genotype	Serotype
GAR872	PFEEC0106	wt (blood isolate)	O1
GAR878	PFEEC0107	wt (blood isolate)	O75
GAR896	PFEEC0108	wt (blood isolate)	O15
GAR1902	PFEEC0109	wt (blood isolate)	O6
Atlas187913	PFEEC0068	wt (blood isolate)	O25b
<i>Salmonella enterica</i> serovar Typhimurium strain LT2	--	wt	N/A

EXAMPLE 10: Oligonucleotide primers for *wzzB*, *fepE* and O-antigen gene cluster cloning

Table 11: Oligonucleotide Primers

Name	Primer Sequence	Comments
LT2wzzB_S	GAAGCAAACCGTACGCGTAAAG (SEQ ID NO: 40)	based on Genbank GCA_000006945.2
LT2wzzB_AS	CGACCAGCTCTTACACGGCG (SEQ ID NO: 41)	Salmonella enterica serovar Typhimurium strain LT2
O25bFepE_S	GAAATAGGACCACTAATAATACACAAATTAAT AAC (SEQ ID NO: 42)	Based on Genbank GCA_000285655.3
O25bFepE_A	ATAATTGACGATCCGGTTGCC (SEQ ID NO: 43)	O25b EC958 strain ST131 assembly and O25b GAR2401 WGS data
wzzB P1_S	GCTATTTACGCCCTGATTGTCTTTTGT (SEQ ID NO: 44)	based on <i>E. coli</i> K-12 strain sequence, Genbank MG1655
wzzB P2_AS	ATTGAGAACCTGCGTAAACGGC (SEQ ID NO: 45)	NC_000913.3 or W3110 assembly
wzzB P3_S	TGAAGAGCGGTTTCAGATAACTTCC (SEQ ID NO: 46) (UDP-glucose-6-dehydrogenase)	GCA_000010245.1
wzzB P4_AS	CGATCCGGAAACCTCCTACAC (SEQ ID NO: 47) (Phosphoribosyl-AMP cyclohydrolase/ Phosphoribosyl-ATP pyrophosphohydrolase)	
O157 FepE_S	GATTATTCGCGCAACGCTAAACAGAT (SEQ ID NO: 48)	<i>E. coli</i> O157 fepE (based on Genbank EDL933 strain GCA_000732965.1)
O157 FepE_AS	TGATCATTGACGATCCGGTAGCC (SEQ ID NO: 49)	
pBAD33_adaptor_S	CGGTAGCTGTAAAGCCAGGGGCGGTAGCGTG GTTTAAACCCAAGCAACAGATCGGCGTCGTCG GTATGGA (SEQ ID NO: 50)	Adaptor has central <i>PmeI</i> site and homology to conserved 5' OAg operon promoter and 3' <i>gnd</i> gene sequences
pBAD33_adaptor_AS	AGCTTCCATACCGACGACGCCGATCTGTTGCT TGGGTTTAAACCACGCTACCGCCCCTGGCTTT ACAGCTACCGAGCT (SEQ ID NO: 51)	
JUMPSTART_r	GGTAGCTGTAAAGCCAGGGGCGGTAGCGTG (SEQ ID NO: 52)	Universal Jumpstart (OAg operon promoter)
gnd_f	CCATACCGACGACGCCGATCTGTTGCTTGG (SEQ ID NO: 53)	Universal 3' OAg (<i>gnd</i>) operon antisense primer

EXAMPLE 11: Plasmids

- 5 Plasmid vectors and subclones are listed in **Table 12**. PCR fragments harboring various *E. coli* and *Salmonella wzzB* and *fepE* genes were amplified from purified genomic DNA and

subcloned into the high copy number plasmid provided in the Invitrogen PCR®Blunt cloning kit **FIG. 12A-12B**. This plasmid is based on the pUC replicon. Primers P3 and P4 were used to amplify *E. coli* *wzzB* genes with their native promoter, and are designed to bind to regions in proximal and distal genes encoding UDP-glucose-6-dehydrogenase and

5 phosphoribosyladenine nucleotide hydrolase respectively (annotated in Genbank *MG1655* NC_000913.3). A PCR fragment containing *Salmonella* *fepE* gene and promoter were amplified using primers previously described. Analogous *E. coli* *fepE* primers were designed based on available Genbank genome sequences or whole genome data generated internally (in case of GAR2401 and O25K5H1). Low copy number plasmid pBAD33 was used to express O-antigen

10 biosynthetic genes under control of the arabinose promoter. The plasmid was first modified to facilitate cloning (via Gibson method) of long PCR fragments amplified using universal primers homologous to the 5' promoter and 3' 6-phosphogluconate dehydrogenase (*gnd*) gene **Table 12**. The pBAD33 subclone containing the O25b biosynthetic operon is illustrated in **FIG. 12A-12B**.

Table 12

Plasmids

Name	Replicon	Resistance marker	Comments
PCR®Blunt II TOPO	pUC	KanR	Invitrogen PCR cloning vector
pBAD33	P15a	CamR	Arabinose inducible vector
pBAD33-OAg	P15a	CamR	OAg operon Gibson cloning vector
pBAD33-O25b	P15a	CamR	O25b OAg expression plasmid
pBAD33-O21	P15a	CamR	O21 OAg expression plasmid
pBAD33-O16	P15a	CamR	O16 OAg expression plasmid
pBAD33-O75	P15a	CamR	O75 OAg expression plasmid
pBAD33-O1	P15a	CamR	O1 OAg expression plasmid
pBAD33-O2	P15a	CamR	O2 OAg expression plasmid
pTOPO-O25b 2401 wzzB	pUC	KanR	GAR 2401 gDNA template
pTOPO-O25b 2401 fepE	pUC	KanR	
pTOPO-K12 wzzB	pUC	KanR	<i>E. coli</i> K-12 strain gDNA template
pTOPO-O25a wzzB	pUC	KanR	<i>E. coli</i> O25a strain O25K5H1
pTOPO-O25a fepE	pUC	KanR	gDNA template
pTOPO-Salmonella LT2 wzzB	pUC	KanR	Salmonella enterica serovar Typhimurium strain LT2 gDNA
pTOPO-Salmonella LT2 fepE	pUC	KanR	template
pTOPO-O25a ETEC wzzB	pUC	KanR	O25a ETEC strain gDNA
pTOPO-O25a ETEC fepE	pUC	KanR	purchased from ATCC ("NR-5" E2539-C1)
pTOPO-O157fepE	pUC	KanR	O157:H7:K- Shigella toxin strain gDNA purchased from ATCC (EDL933 #43895D-5)

EXAMPLE 12: O-antigen Purification

The fermentation broth was treated with acetic acid to a final concentration of 1 – 2% (final pH of 4.1). The extraction of OAg and delipidation were achieved by heating the acid treated broth to 100°C for 2 hours. At the end of the acid hydrolysis, the batch was cooled to ambient temperature and 14% NH₄OH was added to a final pH of 6.1. The neutralized broth was centrifuged and the centrate was collected. To the centrate was added CaCl₂ in sodium phosphate and the resulting slurry was incubated for 30 mins at room temperature. The solids were removed by centrifugation and the centrate was concentrated 12-fold using a 10kDa membrane, followed by two diafiltrations against water. The retentate which contained OAg was then purified using a carbon filter. The carbon filtrate was diluted 1:1 (v/v) with 4.0M ammonium sulfate. The final ammonium sulfate concentration was 2M. The ammonium sulfate treated carbon filtrate was further purified using a membrane with 2M ammonium sulfate as the running buffer. The OAg was collected in the flow through. For the long OAg the HIC filtrate was concentrated and then buffer exchanged against water (20 diavolumes) using a 5kDa membrane. For the short (native) OAg polysaccharide, the MWCO was further reduced to enhance yield.

EXAMPLE 13: Conjugation of O25b long O-antigen to CRM₁₉₇

The first set of long chain O25b polysaccharide-CRM₁₉₇ conjugates were produced using periodate oxidation followed by conjugation using reductive amination chemistry (RAC) (Table 14). Conjugate variants with three activation levels (low, medium and high) by varying the oxidation levels. Conjugates were produced by reacting the lyophilized activated polysaccharides with lyophilized CRM₁₉₇, reconstituted in DMSO medium, using sodium cyanoborohydride as the reducing agent. Conjugation reactions were carried out at 23 °C for 24 hrs, followed by capping using sodium borohydride for 3 hrs. Following the conjugation quenching step, conjugates were purified by ultrafiltration/diafiltration with 100K MWCO regenerated cellulose membrane, using 5mM Succinate/0.9% NaCl, pH 6.0. Final filtration of the conjugates were performed using a 0.22 µm membrane.

Unless expressly stated otherwise, the conjugates disclosed throughout the following Examples include a core saccharide moiety.

Long O-antigen expression conferred by heterologous polymerase chain length regulators.

Initial *E. coli* strain construction focused on the O25 serotype. Goal was to overexpress heterologous *wzzB* or *fepE* genes to see if they confer longer chain length in O25 *wzzB* knockout strains. First, blood isolates were screened by PCR to identify strains of the O25a and

O25b subtype. Next, strains were screened for sensitivity to ampicillin. A single ampicillin-sensitive O25b isolate GAR2401 was identified into which a *wzzB* deletion was introduced. Similarly, a *wzzB* deletion was made in O25a strain O25K5H1. For genetic complementation of these mutations, *wzzB* genes from GAR 2401 and O25K5H1 were subcloned into the high copy
 5 PCR-Blunt II cloning vector and introduced into both strains by electroporation. Additional *wzzB* genes from *E. coli* K-12 and *S. enterica* serovar Typhimurium LT2 were similarly cloned and transferred; likewise *fepE* genes from *E. coli* O25K5H1, GAR 2401, O25a ETEC NR-5, O157:H7:K- and *S. enterica* serovar Typhimurium LT2.

Bacteria were grown overnight in LB medium and LPS was extracted with phenol,
 10 resolved by SDS PAGE (4-12% acrylamide) and stained. Each well of the gel was loaded with LPS extracted from the same number of bacterial cells (approximately 2 OD₆₀₀ units). Size of LPS was estimated from an internal native *E. coli* LPS standard and by counting the ladder discernable from a subset of samples showing a broad distribution of chain lengths (differing by one repeat unit). On the left side of **FIG. 13A**, LPS profiles of plasmid transformants of O25a
 15 O25K5H1Δ*wzzB* are shown; and on the right, analogous profiles of O25b GAR 2401Δ*wzzB* transformants. An immunoblot of a replicate gel probed with O25-specific sera is shown in **FIG. 13B**.

Results from this experiment show that introduction of the homologous *wzzB* gene into the *E. coli* O25aΔ*wzzB* host restores expression of short O25 LPS (10-20x), as does the
 20 *Salmonella* LT2 *wzzB*. Introduction of the O25b *wzzB* gene from GAR2401 does not, suggesting the WzzB enzyme from this strain is defective. A comparison of *E. coli* WzzB amino acid sequences suggests that A210E and P253S substitutions may be responsible. Significantly, *Salmonella* LT2 *fepE* and *E. coli* *fepE* from O25a O25K5H1 conferred the ability to express very long (VL) OAg LPS, with the *Salmonella* LT2 *fepE* resulting in OAg exceeding in
 25 size that conferred by *E. coli* *fepE*.

A similar pattern of expression was observed with GAR2401Δ*wzzB* transformants: *E. coli* O25a or K12 strain *wzzB* restored ability to produce short LPS. The *Salmonella* LT2 *fepE* generated the longest LPS, the *E. coli* *fepE* a slightly shorter LPS, while the *Salmonella* LT2 *wzzB* yielded an intermediate sized long LPS (L). The ability of other *E. coli* *fepE* genes to
 30 produce very long LPS was assessed in a separate experiment with transformants of *E. coli* O25aΔ*wzzB*. The *fepE* genes from GAR2401, an O25a ETEC strain and an O157 *Shigella* toxin producing strain also conferred the ability to produce very long LPS, but not as long as the LPS generated with the *Salmonella* LT2 *fepE* (**FIG. 14**).

Having established in serotype O25a and O25b strains that *Salmonella* LT2 *fepE*
 35 generates the longest LPS of the polymerase regulators evaluated, we next sought to determine whether it would also produce very long LPS in other *E. coli* serotypes. Wild-type

bacteremia isolates of serotype O1, O2, O6, O15 and O75 were transformed with the *Salmonella* *fepE* plasmid and LPS extracted. The results shown in **FIG. 15** confirm that *Salmonella* *fepE* can confer the ability to make very long LPS in other prevalent serotypes associated with blood-infections. Results also show that plasmid-based expression of

5 *Salmonella* *fepE* appears to override the control of chain length normally exerted by endogenous *wzzB* in these strains.

Plasmid-based expression of O-antigens in a common *E. coli* host strain.

From the perspective of bioprocess development, the ability to produce O-antigens of different serotypes in a common *E. coli* host instead of multiple strains would greatly simplify the

10 manufacturing of individual antigens. To this end, O-antigen gene clusters from different serotypes were amplified by PCR and cloned into a low-copy number plasmid (pBAD33) under control of an arabinose regulated promoter. This plasmid is compatible (can coexist) with the *Salmonella* *LT2 fepE* plasmid in *E. coli* as it harbors a different (p15a) replicon and different selectable marker (chloramphenicol vs kanamycin). In a first experiment, a pBAD33 O25b

15 operon plasmid subclone was cotransfected with the *Salmonella* *LT2 fepE* plasmid into GAR2401Δ*wzzB* and transformants grown in the presence or absence of 0.2% arabinose. Results shown in **FIG. 16A-16B** demonstrated that very long O-antigen LPS was produced in an arabinose-dependent manner.

O-antigen gene clusters cloned from other serotypes were similarly evaluated and the

20 results shown in **FIG. 17**. Co-expression of *Salmonella* *LT2 fepE* and pBAD33-OAg plasmids resulted in detectable long chain LPS corresponding to O1, O2 (for two out of four clones), O16, O21 and O75 serotypes. For unknown reasons, the pBAD33-O6 plasmid failed to yield detectable LPS in all four isolates tested. Although expression level was variable, results show that expression of long chain O-antigens in a common host is feasible. However, in some cases

25 further optimization to improve expression may be required, for example by modification of plasmid promoter sequences.

The profiles of LPS from different serotype O25 *E. coli* strains with or without the *Salmonella* *LT2 fepE* plasmid are shown in **FIG. 18**. Two strains were studied for fermentation, extraction and purification of O-antigens: GAR2831, for the production of native short O25b

30 OAg; and GAR2401Δ*wzzB/fepE*, for the production of long O25b OAg. The corresponding short and long form LPSs shown in the **FIG. 18** SDS-PAGE gel are highlighted in red. Polysaccharides were extracted directly from fermented bacteria with acetic acid and purified. Size exclusion chromatography profiles of purified short and long or very long O25b polysaccharides are shown in **FIG. 19A-19B**. The properties of two lots of short polysaccharide

35 (from GAR2831) are compared with a single very long polysaccharide preparation (from strain

GAR2401 Δ wzzB/*fepE*). The molecular mass of the long O-antigen is 3.3-fold greater than that of the short O-antigen, and the number of repeat units was estimated to be ~65 (very long) vs ~20. See **Table 13**.

Table 13

Poly Lot #	Native	Native	Modified (long chain)
Poly Lot #	709766-24A	709722-24B	709766-25A
Poly MW (kDa)	17.3	16.3	55.3
# Repeat Units	20	19	64

5

The very long O25b O-antigen polysaccharide was conjugated to diphtheria toxoid CRM₁₉₇ using a conventional reductive amination process. Three different lots of glycoconjugate were prepared with varying degree of periodate activation: medium (5.5%), low (4.4%) and high (8.3%). The resulting preparations and unconjugated polysaccharide were shown to be free of endotoxin contamination (**Table 14**).

Groups of four rabbits (New Zealand White females) were each vaccinated with 10 mcg of glycoconjugate and 20 mcg of QS21 adjuvant and serum sampled (VAC-2017-PRL-EC-0723) according to the schedule shown in **FIG. 20A**. It is worth noting that a 10 mcg dose is at the low end of the range customarily given to rabbits in the evaluation of bacterial glycoconjugates (20-50 mcg is more typical). A group of rabbits was also vaccinated in a separate study (VAC-2017-PRL-GB-0698) with unconjugated polysaccharide using the same dose (10 mcg polysaccharide + 20mcg QS21 adjuvant) and identical administration schedule.

Rabbit antibody responses to the three O25b glycoconjugate preparations were evaluated in a LUMINEX assay in which carboxy beads were coated with methylated human serum albumin prebound with unconjugated O25b long polysaccharide. The presence of O25b-specific IgG antibodies in serum samples was detected with a phycoerythrin(PE)-labelled anti-IgG secondary antibody. The profiles of immune responses observed in sera sampled at week 0 (pre-immune), week 6 (post-dose 2, PD2), week 8 (post-dose 3, PD3) and week 12 (post-dose 4, PD4) in best-responding rabbits (one from each group of four) are shown in **FIG. 21A-21C**. No significant pre-immune serum IgG titers were detected in any of the 12 rabbits. In contrast, O25b antigen-specific antibody responses were detected in post-vaccination sera from rabbits in all three groups, with the low-activation glycoconjugate group responses trending slightly higher than the medium or high activation glycoconjugate groups. Maximal responses were observed by the post-dose 3 timepoint. One rabbit in the low activation group and one rabbit from the high activation group failed to respond to vaccination (non-responders).

To assess the impact of CRM₁₉₇ carrier protein conjugation on immunogenicity of the long O25b OAg polysaccharide, the presence of antibodies in sera from rabbits vaccinated with unconjugated polysaccharide was compared with sera from rabbits vaccinated with the low activation CRM₁₉₇ glycoconjugate **FIG. 22A-22F**. Remarkably, the free polysaccharide was not immunogenic, eliciting virtually no IgG responses in immune vs preimmune sera (**FIG. 22A**). In contrast, O25b OAg-specific IgG mean fluorescence intensity values (MFIs) of approximately ten-fold above pre-immune serum levels were observed in PD4 sera from three out of four rabbits vaccinated with O25b OAg- CRM₁₉₇, across a range of serum dilutions (from 1:100 to 1:6400). These results demonstrate the necessity of carrier protein conjugation to generate IgG antibodies to the O25b OAg polysaccharide at the 10 mcg dose level.

Bacteria grown on TSA plates were suspended in PBS, adjusted to OD₆₀₀ of 2.0 and fixed in 4% paraformaldehyde in PBS. After blocking in 4% BSA/PBS for 1h, bacteria were incubated with serial dilutions of pre-immune and PD3 immune sera in 2% BSA/PBS, and bound IgG detected with PE-labeled secondary F(ab) antibody.

Specificity of the O25b antibodies elicited by the O25b OAg-CRM₁₉₇ was demonstrated in flow cytometry experiments with intact bacteria. Binding of IgG to whole cells was detected with PE-conjugated F(ab')₂ fragment goat anti-rabbit IgG in an Accuri flow cytometer.

As shown in **FIG. 23A-23C**, pre-immune rabbit antibodies failed to bind to wild-type serotype O25b isolates GAR2831 and GAR2401 or to a K-12 *E. coli* strain, whereas matched PD3 antibodies stained the O25b bacteria in a concentration dependent manner. Negative control K-12 strain which lacks the ability to express OAg showed only very weak binding of PD3 antibodies, most likely due to the presence of exposed inner core oligosaccharide epitopes on its surface. Introduction of the *Salmonella* *fepE* plasmid into the wild-type O25b isolates resulted in significantly enhanced staining, consistent with the higher density of immunogenic epitopes provided by the longer OAg polysaccharide.

Conclusion: The results described show that not only is *Salmonella* *fepE* the determinant of very long O-antigen polysaccharides in *Salmonella* species, but that it also can confer on *E. coli* strains of different O-antigen serotypes the ability to make very long OAg. This property can be exploited to produce O-antigen vaccine polysaccharides with improved properties for bioprocess development, by facilitating purification and chemical conjugation to appropriate carrier proteins, and by potentially enhancing immunogenicity through the formation of higher molecular weight complexes.

EXAMPLE 14: Initial rabbit studies generated first polyclonal antibody reagents and IgG responses to RAC O25b OAg-CRM₁₉₇

Long chain O25b polysaccharide-CRM₁₉₇ conjugates were produced using periodate oxidation followed by conjugation using reductive amination chemistry (RAC) (Table 14). See also Table 24.

Table 14

CRM ₁₉₇ conjugate	132242-28 Medium 5.5% activation	132242-27 Low 4.5% activation	132242-29 High 8.3% activation	709766-29 Free O25b polysaccharide
Polysaccharide concentration (mg/mL)	0.7	0.6	0.67	1
Endotoxin (EU/ug)	0.02	0.02	0.02	<0.6 EU
Matrix	5 um Succinate buffer/saline, pH 6.0			

In Rabbit Study 1 (VAC-2017-PRL-EC-0723) (also described above in **Example 13**) – five (5) rabbits/group, with 10ug L-, M- or H-activation RAC (+QS21) received a composition according to the schedule shown in **FIG. 20A**. Unconjugated free O25b polysaccharide was observed not to be immunogenic in a follow-up rabbit Study (VAC-2017-PRL-GB-0698) (see **FIG. 25**).

In Rabbit Study 2 (VAC-2018-PRL-EC-077) – 2 rabbits/group, with L-RAC (AlOH₃, QS21, or no adjuvant) received a composition according to the schedule shown in **FIG. 20B**.

Rabbits 4-1, 4-2, 5-1, 5-2, 6-1, and 6-2 received the very long unconjugated O25b polysaccharide described in **Example 13**, and week 18 sera were tested.

More specifically, a composition including 50 ug unconjugated O25b, 100 ug AlOH₃ adjuvant was administered to Rabbit 4-1. A composition including 50 ug unconjugated O25b, 100 ug AlOH₃ adjuvant was administered to Rabbit 4-2. A composition including 50 ug unconjugated O25b, 50 ug QS-21 adjuvant was administered to Rabbit 5-1. A composition including 50 ug unconjugated O25b, 50 ug QS-21 adjuvant was administered to Rabbit 5-2. A composition including 50 ug unconjugated O25b, no adjuvant was administered to Rabbit 6-1. A composition including 50 ug unconjugated O25b, no adjuvant was administered to Rabbit 6-2.

EXAMPLE 15: Rabbit studies with O25b RAC conjugate: dLIA serum dilution titers

Rabbit Study 2 (VAC-2018-PRL-EC-077) O25b dLIA serum dilution titers vs best responding rabbit from study 1 (VAC-2017-PRL-EC-0723). For these experiments a modified direct binding Luminex assay was implemented in which a polylysine conjugate of O25b long O-antigen was passively adsorbed onto the Luminex carboxy beads instead of the methylated serum albumin long O-antigen mixture described previously. The use of the polylysine-O25b conjugate improved the sensitivity of the assay and the quality of IgG concentration dependent responses, permitting determination of serum dilution titers through use of curve-fitting (four parameter non-linear equation). O25b IgG titers in sera from highest titer rabbit from first study is compared with sera from second study rabbits in **Table 15**.

Table 15

	O25b-CRM Low Activation Conjugate with Alum Adjuvant (EC ₅₀ as serum dilution)		O25b-CRM Low Activation Conjugate with QS21 Adjuvant (EC ₅₀ as serum dilution)		O25b-CRM Low Activation Conjugate without Adjuvant (EC ₅₀ as serum dilution)	
	Rabbit 1-1	Rabbit 1-2	Rabbit 2-1	Rabbit 2-2	Rabbit 2-3	Rabbit 2-4
Week 3 Antisera (3 wks after primary)	~1:200	~1:200	<1:100	<1:100	~1:200	~1:200
Week 7 Antisera (1 wk after boost 1)	1:1600	1:4000	1:250	1:500	1:250	1:1500
Week 10 Antisera (1 wk after boost 2)	1:1100	1:1900	1:250	1:500	1:800	1:1200
Week 18 Antisera (1 wk after boost 4)	1:1600	1:4000	1:1300	1:1200	1:1400	1:1600
	Average of 6 replicates of best antisera from rabbit 2-3 (assay standard from first study) EC₅₀ = 1:1700					

Higher doses in second rabbit study (50/20ug vs 10ug) did not improve IgG titers.

Two month rest boosts IgG responses (not observed with shorter intervals).

Alum appears to enhance IgG response in rabbits compared with QS21 or no adjuvant.

An opsonophagocytic assay (OPA) with baby rabbit complement (BRC) and HL60 cells as source of neutrophils was established to measure the functional immunogenicity of O-antigen glycoconjugates. Pre-frozen bacterial stocks of *E. coli* GAR2831 were grown in Luria broth (LB) media at 37°C. Cells were pelleted and suspended to a concentration of 1 OD₆₀₀ unit per ml in PBS supplemented with 20% glycerol and frozen. Pre-titered thawed bacteria were diluted to 0.5X 10⁵ CFU/ml in HBSS (Hank's Balanced Salt Solution) with 1% Gelatin) and 10 µL (10³ CFU) combined with 20 µL of serially diluted sera in a U-bottomed tissue culture microplate and the mixture shaken at 700 rpm BELLCO Shaker) for 30 min at 37°C in a 5% CO₂ incubator. 10 µL of 2.5% complement (Baby Rabbit Serum, PEL-FREEZ 31061–3, prediluted in HBG) and 20 µL of HL-60 cells (0.75X 10⁷ /ml) and 40 µL of HBG added to the U-bottomed tissue culture microplate and the mixture shaken at 700 rpm BELLCO Shaker) for 45min at 37°C in a 5% CO₂ incubator. Subsequently, 10 µL of each 100µL reaction was transferred into the corresponding wells of a pre-wetted MILLIPORE MULTISCREENHTS HV filter plate prepared by applying 100 µL water, filter vacuumed, and applying 150 µL of 50% LB. The filter plate was vacuum filtered and incubated overnight at 37°C in a 5% CO₂ incubator. The next day the colonies were enumerated after fixing, staining, and destaining with COOMASSIE dye and Destain solutions, using an IMMUNOSPOT® analyzer and IMMUNOCAPTURE software. To establish the specificity of OPA activity, immune sera were preincubated with 100 µg/mL purified long O25b O-antigen prior to combining with the other assay components in the OPA reaction. The OPA assay includes control reactions without HL60 cells or complement, to demonstrate dependence of any observed killing on these components.

Matched pre-immune and post-vaccination serum samples from representative rabbits from both rabbit studies were evaluated in the assay and serum dilution titers determined (Table 16, FIG. 26A-26B). Preincubation with unconjugated O25b long O-antigen polysaccharide blocked bactericidal activity demonstrating specificity of the OPA (FIG. 19C).

Table 16 OPA titers

Rabbit 2-3 was dosed as follows: Rabbit 2-3 dosing: 10/10/10/10ug RAC conjugate + QS21, post-dose (PD) 4 bleed. Rabbit 1-2 was dosed as follows: 50/20/20/20ug RAC conjugate + Al(OH)₃, PD4 bleed.

Table 16	
Sample	Titer
Rabbit 2-3 Pre-immune serum	537
Rabbit 2-3 wk13 serum (terminal bleed)	13686
Rabbit 1-2 Pre-immune serum	<200

Rabbit 1-2 wk19 serum (terminal bleed)	22768
--	-------

EXAMPLE 16: O-antigen O25b IgG levels elicited by unconjugated O25b long O-antigen polysaccharide and derived O25b RAC/DMSO long O-antigen glycoconjugate.

Groups of ten CD-1 mice were dosed by sub-cutaneous injection with 0.2 or 2.0
 5 µg/animal of O25b RAC/DMSO long O-antigen glycoconjugate at weeks 0, 5 and 13, with
 bleeds taken at week 3 (post-dose 1, PD1), week 6 (post-dose 2, PD2) and week 13 (post-dose
 3, PD3) timepoints for immunogenicity testing. Levels of antigen-specific IgG were determined
 by quantitative Luminex assay (see details in **Example 15**) with O25b-specific mouse mAb as
 internal standard. Baseline IgG levels (dotted line) were determined in serum pooled from 20x
 10 randomly selected unvaccinated mice. The free unconjugated O25b long O-antigen
 polysaccharide immunogen did not induce IgG above baseline levels at any timepoint. In
 contrast, IgG responses were observed after two doses of O25b-CRM197 RAC long conjugate
 glycoconjugate: robust uniform IgG responses were observed by PD3, with intermediate and
 more variable IgG levels at PD2. GMT IgG values (ng/ml) are indicated with 95% CI error bars.
 15 See **FIG. 27A-27C**.

EXAMPLE 17: Specificity of the O25b baby rabbit complement (BRC) OPA.

A-B) O25b RAC/DMSO long O-antigen post-immune serum from rabbits 2-3 and 1-2
 (but not matched pre-immune control serum) shows bactericidal OPA activity. C) OPA activity of
 immune serum from rabbit 1-2 was blocked by pre-incubation with 100µg/mL long O-antigen
 20 O25b polysaccharide. Strain GAR2831 bacteria were incubated with HL60s, 2.5% BRC and
 serial dilutions of serum for 1h at 37°C and surviving bacteria enumerated by counting
 microcolonies (CFUs) on filter plates. See **FIG. 26A-26C**.

EXAMPLE 18: RAC and eTEC O25b long glycoconjugates are more immunogenic than single end glycoconjugates.

BRC OPA assay with carbapenem-resistant fluoroquinolone-resistant MDR strain
 Atlas187913. Groups of 20 CD-1 mice were vaccinated with 2µg of glycoconjugate according to
 the same schedule as shown in **FIG. 28A-28B** and OPA responses determined at post-dose 2
 (PD2) (**FIG. 28A**) and post-dose 3 (PD3) (**FIG. 28B**) timepoints. Bars indicate GMTs with 95%
 CI. Responder rates above unvaccinated baseline are indicated. Log transformed data from
 30 different groups were evaluated to assess if differences were statistically significant using
 unpaired t-test with Welch's correction (Graphpad Prism). Results are summarized in the **Table**
17. See **FIG. 28A-28B**. In mice that were vaccinated with 2 µg of eTEC O1a long

glycoconjugates, OPA titers against O1a, PD2 and PD3 (data not shown), were observed to be greater than the OPA titers against O25b, PD2 and PD3, respectively, shown in **Table 17**.

Table 17

DESCRIPTION	% Responders (n/N)*	GEOMEAN TITER PD2	% Responders (n/N)*	GEOMEAN TITER PD3
Single end short, 2 µg	45 (9/20)	1,552	85 (17/20)	17,070
Single end long, 2µg	30 (6/20)	763	85 (17/20)	10,838
RAC/DMSO long, 2µg	65 (13/20)	8,297	95 (19/20)	163,210
eTEC (10%) long, 2µg	90 (18/20)	27,368	100 (19/19)	161,526

5 EXAMPLE 19: OPA immunogenicity of eTEC chemistry may be improved by modifying levels of polysaccharide activation.

BRC OPA assay with carbapenem-resistant fluoroquinolone-resistant MDR strain Atlas187913. Groups of 20 CD-1 mice were vaccinated with 0.2µg or 2µg of the indicated long O25b eTEC glycoconjugate and OPA responses determined at PD2 timepoint. Aggregated log transformed data from 4% activation vs 17% activation groups were evaluated to confirm that differences in OPA responses were statistically significant using unpaired t-test with Welch's correction (Graphpad Prism). GMTs and responder rates for individual groups are summarized in **Table 18**. See **FIG. 29**.

Table 18

Description	% Responders (n/N)	GeoMean Titer
eTEC long 4% activation (0.2 µg)	35 (7/20)	628
eTEC long 4% activation (0.2 µg)	65 (13/20)	8,185
eTEC long 10% activation (0.2 µg)	45 (9/20)	1,085
eTEC long 10% activation (0.2 µg)	90 (18/20)	27,368
eTEC long 17% activation (0.2 µg)	70 (14/20)	3,734
eTEC long 17% activation (0.2 µg)	80 (16/20)	25,461

15

EXAMPLE 20: Challenge study indicates long *E. coli* O25b eTEC conjugates elicit protection after three doses.

Groups of 20x CD-1 mice immunized with a 2µg dose according to the indicated schedule were challenged IP with 1×10^9 bacteria of strain GAR2831. Subsequent survival was monitored for six days. Groups of mice vaccinated with eTEC glycoconjugates activated at 4%,

20

10% or 17% levels were protected from lethal infection, whereas unvaccinated control mice or mice vaccinated with 2 μ g unconjugated O25b long polysaccharide were not. See FIG. 30A-30B.

EXAMPLE 21: Process for Preparation of eTEC Linked Glycoconjugates

5 **Activation of Saccharide and Thiolation with Cystamine dihydrochloride.** The saccharide is reconstituted in anhydrous dimethylsulfoxide (DMSO). Moisture content of the solution is determined by Karl Fischer (KF) analysis and adjusted to reach a moisture content of 0.1 and 1.0%, typically 0.5%.

10 To initiate the activation, a solution of 1,1'-carbonyl-di-1,2,4-triazole (CDT) or 1,1'-carbonyldiimidazole (CDI) is freshly prepared at a concentration of 100 mg/mL in DMSO. The saccharide is activated with various amounts of CDT/CDI (1 – 10 molar equivalents) and the reaction is allowed to proceed for 1 - 5 hours at rt or 35 °C. Water was added to quench any residual CDI/CDT in the activation reaction solution. Calculations are performed to determine the added amount of water and to allow the final moisture content to be 2 - 3% of total aqueous. The
15 reaction was allowed to proceed for 0.5 hour at rt. Cystamine dihydrochloride is freshly prepared in anhydrous DMSO at a concentration of 50 mg/mL. The activated saccharide is reacted with 1 - 2 mol. eq. of cystamine dihydrochloride. Alternatively, the activated saccharide is reacted with 1 - 2 mol. eq. of cysteamine hydrochloride. The thiolation reaction is allowed to proceed for 5 - 20 hours at rt, to produce a thiolated saccharide. The thiolation level is determined by the added
20 amount of CDT/CDI.

Reduction and Purification of Activated Thiolated Saccharide. To the thiolated saccharide reaction mixture a solution of tris(2-carboxyethyl)phosphine (TCEP), 3 - 6 mol. eq., is added and allowed to proceed for 3 - 5 hours at rt. The reaction mixture is then diluted 5 - 10-fold by addition to pre-chilled 10 mM sodium phosphate monobasic, and filtered through a 5 μ m
25 filter. Dialfiltration of thiolated saccharide is performed against 30 - 40-fold diavolume of pre-chilled 10 mM sodium phosphate monobasic. An aliquot of activated thiolated saccharide retentate is pulled to determine the saccharide concentration and thiol content (Ellman) assays.

Activation and Purification of Bromoacetylated Carrier Protein. Free amino groups of the carrier protein are bromoacetylated by reaction with a bromoacetylating agent, such as
30 bromoacetic acid N-hydroxysuccinimide ester (BAANS), bromoacetyl bromide, or another suitable reagent.

 The carrier protein (in 0.1M Sodium Phosphate, pH 8.0 $\pm$ 0.2) is first kept at 8 $\pm$ 3 °C, at about pH 7 prior to activation. To the protein solution, the N-hydroxysuccinimide ester of bromoacetic acid (BAANS) as a stock dimethylsulfoxide (DMSO) solution (20 mg/mL) is added
35 in a ratio of 0.25 – 0.5 BAANS: protein (w/w). The reaction is gently mixed at 5 $\pm$ 3 °C for 30 –

60 minutes. The resulting bromoacetylated (activated) protein is purified, e.g., by ultrafiltration/diafiltration using 10 kDa MWCO membrane using 10 mM phosphate (pH 7.0) buffer. Following purification, the protein concentration of the bromoacetylated carrier protein is estimated by Lowry protein assay.

- 5 The extent of activation is determined by total bromide assay by ion-exchange liquid chromatography coupled with suppressed conductivity detection (ion chromatography). The bound bromide on the activated bromoacetylated protein is cleaved from the protein in the assay sample preparation and quantitated along with any free bromide that may be present. Any remaining covalently bound bromine on the protein is released by conversion to ionic
- 10 bromide by heating the sample in alkaline 2-mercaptoethanol.

- Activation and Purification of Bromoacetylated CRM₁₉₇.** CRM₁₉₇ was diluted to 5 mg/mL with 10 mM phosphate buffered 0.9% NaCl pH 7 (PBS) and then made 0.1 M NaHCO₃ pH 7.0 using 1 M stock solution. BAANS was added at a CRM₁₉₇ : BAANS ratio 1 : 0.35 (w:w) using a BAANS stock solution of 20 mg/mL DMSO. The reaction mixture was incubated at
- 15 between 3 °C and 11 °C for 30 mins-1 hour then purified by ultrafiltration/diafiltration using a 10K MWCO membrane and 10mM Sodium Phosphate/0.9% NaCl, pH 7.0. The purified activated CRM₁₉₇ was assayed by the Lowry assay to determine the protein concentration and then diluted with PBS to 5 mg/mL. Sucrose was added to 5% wt/vol as a cryoprotectant and the activated protein was frozen and stored at -25 °C until needed for conjugation.
- 20 Bromoacetylation of lysine residues of CRM₁₉₇ was very consistent, resulting in the activation of 15 to 25 lysines from 39 lysines available. The reaction produced high yields of activated protein.

- Conjugation of Activated Thiolated Saccharide to Bromoacetylated Carrier Protein.** Bromoacetylated carrier protein and activated thiolated saccharide are subsequently
- 25 added. The saccharide/protein input ratio is 0.8 ± 0.2. The reaction pH is adjusted to 9.0 ± 0.1 with 1 M NaOH solution. The conjugation reaction is allowed to proceed at 5 °C for 20 ± 4 hours.

- Capping of Residual Reactive Functional Groups.** The unreacted bromoacetylated residues on the carrier protein are quenched by reacting with 2 mol. eq. of N-acetyl-L-cysteine
- 30 as a capping reagent for 3 - 5 hours at 5 °C. Residual free sulfhydryl groups are capped with 4 mol. eq. of iodoacetamide (IAA) for 20 - 24 hours at 5 °C.

- Purification of eTEC-linked Glycoconjugate.** The conjugation reaction (post-IAA-capped) mixture is filtered through 0.45 µm filter. Ultrafiltration/diafiltration of the glycoconjugate is performed against 5 mM succinate-0.9% saline, pH 6.0. The glycoconjugate
- 35 retentate is then filtered through 0.2 µm filter. An aliquot of glycoconjugate is pulled for assays.

The remaining glycoconjugate is stored at 5 °C. See **Table 21**, **Table 22**, **Table 23**, **Tale 24**, and **Table 25**.

EXAMPLE 22: PREPARATION OF *E. COLI*-O25B ETEC CONJUGATES

Activation Process - Activation of *E. coli*-O25b Lipopolysaccharide.

5 lyophilized *E. coli*-O25b polysaccharide was reconstituted in anhydrous dimethylsulfoxide (DMSO). Moisture content of the lyophilized O25b/DMSO solution was determined by Karl Fischer (KF) analysis. The moisture content was adjusted by adding WFI to the O25b/DMSO solution to reach a moisture content of 0.5%.

To initiate the activation, 1,1'-carbonyldiimidazole (CDI) was freshly prepared as 100
10 mg/mL in DMSO solution. *E. coli*-O25b polysaccharide was activated with various amounts of CDI prior to the thiolation step. The CDI activation was carried out at rt or 35°C for 1 - 3 hours. Water was added to quench any residual CDI in the activation reaction solution. Calculations are performed to determine the added amount of water and to allow the final moisture content to be 2 - 3% of total aqueous. The reaction was allowed to proceed for 0.5 hour at rt.

15 **Thiolation of Activated *E. coli*-O25b Polysaccharide.** Cystamine-dihydrochloride was freshly prepared in anhydrous DMSO and 1 - 2 mol. eq. of cystamine dihydrochloride was added to the activated polysaccharide reaction solution. The reaction was allowed to proceed for 20 ± 4 hours at rt.

Reduction and Purification of Activated Thiolated *E. coli*-O25b Polysaccharide. To
20 the thiolated saccharide reaction mixture a solution of tris(2-carboxyethyl)phosphine (TCEP), 3 - 6 mol. eq., was added and allowed to proceed for 3 - 5 hours at rt. The reaction mixture was then diluted 5 - 10-fold by addition to pre-chilled 10 mM sodium phosphate monobasic and filtered through a 5µm filter. Dialfiltration of thiolated saccharide was performed against 40-fold diavolume of pre-chilled 10 mM sodium phosphate monobasic with 5K MWCO ultrafilter
25 membrane cassettes. The thiolated O25b polysaccharide retentate was pulled for both saccharide concentration and thiol (Ellman) assays. A flow diagram of the activation process is provided in **FIG. 32A**).

Conjugation Process - Conjugation of Thiolated *E. coli*-O25b Polysaccharide to Bromoacetylated CRM₁₉₇. The CRM₁₉₇ carrier protein was activated separately by
30 bromoacetylation, as described in **Example 21**, and then reacted with the activated *E. coli*-O25b polysaccharide for the conjugation reaction. Bromoacetylated CRM₁₉₇ and thiolated O25b polysaccharide were mixed together in a reaction vessel. The saccharide/protein input ratio was 0.8 ± 0.2. The reaction pH was adjusted to 8.0 – 10.0. The conjugation reaction was allowed to proceed at 5 °C for 20 ± 4 hours.

Capping of Reactive Groups on Bromoacetylated CRM₁₉₇ and Thiolated *E. coli*-O25b Polysaccharide. The unreacted bromoacetylated residues on CRM₁₉₇ proteins were capped by reacting with 2 mol. eq. of N-acetyl-L-cysteine for 3 - 5 hours at 5 °C, followed by capping any residual free sulfhydryl groups of the thiolated O25b-polysaccharide with 4 mol. eq. of iodoacetamide (IAA) for 20 - 24 hours at 5 °C.

Purification of eTEC-linked *E. coli*-O25b Glycoconjugate. The conjugation solution was filtered through a 0.45 µm or 5µm filter. Diafiltration of the O25b glycoconjugate was carried out with 100K MWCO ultrafilter membrane cassettes. Diafiltration was performed against 5 mM succinate-0.9% saline, pH 6.0. The *E. coli*-O25b glycoconjugate 100K retentate was then filtered through a 0.22 µm filter and stored at 5 °C.

A flow diagram of the conjugation process is provided in **FIG. 32B**.

Results

The reaction parameters and characterization data for several batches of *E. coli*-O25b eTEC glycoconjugates are shown in **Table 19**. The CDI activation-thiolation with cystamine dihydrochloride generated glycoconjugates having from 41 to 92% saccharide yields and <5 to 14% free saccharides. See also **Table 21**, **Table 22**, **Table 23**, **Table 24** and **Table 25**.

Table 19. Experimental Parameters and Characterization Data of *E. coli*-O25b eTEC Conjugates

Conjugate Batch	O25b-1A	O25b-2B	O25b-3C	O25b-4D	O25b-5E	O25b-6F
Activation level (mol of thiol/mol of polysaccharide), %	10	20	22	17	25	24
Input Sacc/Prot Ratio	0.8	0.8	0.8	0.8	0.8	0.8
Saccharide yield (%)	56	57	79	92	41	59
Output Sacc/Prot Ratio	0.88	1	1.18	1.32	2.9	1.4
Free Saccharide, %	8	<5	6	5	14	5
Free Protein, %	< 1	< 1	< 1	< 1	< 1	< 1
Conjugate Mw, kDa	1057	4124	2259	2306	1825	1537

Total CMCA	3	na	na	7.2	na	na
------------	---	----	----	-----	----	----

EXAMPLE 23: Procedure for the preparation of *E. coli* O-antigen polysaccharide-CRM197 eTEC conjugates (applied to O-antigens from *E. coli* serotypes O25b, O1a, O2, and O6)

Activation of polysaccharide. The *E. coli* O-antigen polysaccharide is reconstituted in anhydrous dimethylsulfoxide (DMSO). To initiate the activation, various amounts of 1,1'-carbonyldiimidazole (CDI) (1 – 10 molar equivalents) is added to the polysaccharide solution and the reaction is allowed to proceed for 1 - 5 hours at rt or 35 °C. Then, water (2 - 3%, v/v) was added to quench any residual CDI in the activation reaction solution. After the reaction was allowed to proceed for 0.5 hour at rt, 1 - 2 mol. eq. of cystamine dihydrochloride is added. The reaction is allowed to proceed for 5 - 20 hours at rt, and then treated with 3 - 6 mol. eq of tris(2-carboxyethyl)phosphine (TCEP) to produce a thiolated saccharide. The thiolation level is determined by the added amount of CDI.

The reaction mixture is then diluted 5 - 10-fold by addition to pre-chilled 10 mM sodium phosphate monobasic, and filtered through a 5µm filter. Dialfiltration of thiolated saccharide is performed against 30 - 40-fold diavolume of pre-chilled 10 mM sodium phosphate monobasic. An aliquot of activated thiolated saccharide retentate is pulled to determine the saccharide concentration and thiol content (Ellman) assays.

Activation of Carrier Protein (CRM₁₉₇). The CRM₁₉₇ (in 0.1M Sodium Phosphate, pH 8.0 ± 0.2) is first kept at 8 ± 3 °C, at about pH 8 prior to activation. To the protein solution, the N-hydroxysuccinimide ester of bromoacetic acid (BAANS) as a stock dimethylsulfoxide (DMSO) solution (20 mg/mL) is added in a ratio of 0.25 – 0.5 BAANS: protein (w/w). The reaction is gently mixed at 5 ± 3 °C for 30 – 60 minutes. The resulting bromoacetylated (activated) protein is purified, e.g., by ultrafiltration/diafiltration using 10 kDa MWCO membrane using 10 mM phosphate (pH 7.0) buffer. Following purification, the protein concentration of the bromoacetylated carrier protein is estimated by Lowry protein assay.

Conjugation. Activated CRM₁₉₇ and activated *E. coli* O-antigen polysaccharide are subsequently added to a reactor and mixed. The saccharide/protein input ratio is 1 ± 0.2. The reaction pH is adjusted to 9.0 ± 0.1 with 1 M NaOH solution. The conjugation reaction is allowed to proceed at 5 °C for 20 ± 4 hours. The unreacted bromoacetylated residues on the carrier protein are quenched by reacting with 2 mol. eq. of N-acetyl-L-cysteine as a capping reagent for 3 - 5 hours at 5 °C. Residual free sulfhydryl groups are capped with 4 mol. eq. of iodoacetamide (IAA) for 20 - 24 hours at 5 °C. Then, the reaction mixture is purified using ultrafiltration/diafiltration performed against 5 mM succinate-0.9% saline, pH 6.0. The purified

conjugate is then filtered through 0.2 µm filter. See Table 21, Table 22, Table 23, Table 24 and Table 25.

EXAMPLE 24: General procedure- Conjugation of O-antigen (from *E. coli* serotypes O1, O2, O6, 25b) Polysaccharide by Reductive mination Chemistry (RAC)
Conjugation in dimethylsulfoxide (RAC/DMSO)

Activating Polysaccharide. Polysaccharide oxidation was carried out in 100 mM sodium phosphate buffer (pH 6.0 ± 0.2) by sequential addition of calculated amount of 500 mM sodium phosphate buffer (pH 6.0) and water for injection (WFI) to give final polysaccharide concentration of 2.0 g/L. If required, the reaction pH was adjusted to pH 6.0, approximately. After pH adjustment, the reaction temperature was cooled to 4 °C. Oxidation was initiated by the addition of approximately 0.09 - 0.13 molar equivalents of sodium periodate. The oxidation reaction was performed at 5 ± 3 °C for 20 ± 4 hrs, approximately.

Concentration and diafiltration of the activated polysaccharide was carried out using 5K MWCO ultrafiltration cassettes. Diafiltration was performed against 20-fold diavolumes of WFI. The purified activated polysaccharide was then stored at 5 ± 3°C. The purified activated saccharide is characterized, *inter alia*, by (i) saccharide concentration by colorimetric assay; (ii) aldehyde concentration by colorimetric assay; (iii) degree of oxidation; and (iv) molecular weight by SEC-MALLS.

Compounding Activated Polysaccharide with Sucrose Excipient, and Lyophilizing.

The activated polysaccharide was compounded with sucrose to a ratio of 25 grams of sucrose per gram of activated polysaccharide. The bottle of compounded mixture was then lyophilized. Following lyophilization, bottles containing lyophilized activated polysaccharide were stored at -20 ± 5°C. Calculated amount of CRM₁₉₇ protein was shell-frozen and lyophilized separately. Lyophilized CRM₁₉₇ was stored at -20 ± 5°C.

Reconstituting Lyophilized Activated Polysaccharide and Carrier Protein.

Lyophilized activated polysaccharide was reconstituted in anhydrous dimethyl sulfoxide (DMSO). Upon complete dissolution of polysaccharide, an equal amount of anhydrous DMSO was added to lyophilized CRM₁₉₇ for reconstitution.

Conjugating and Capping. Reconstituted activated polysaccharide was combined with reconstituted CRM₁₉₇ in the reaction vessel, followed by mixing thoroughly to obtain a clear solution before initiating the conjugation with sodium cyanoborohydride. The final

polysaccharide concentration in reaction solution was approximately 1 g/L. Conjugation was initiated by adding 0.5 – 2.0 MEq of sodium cyanoborohydride to the reaction mixture and incubating at 23 ± 2 °C for 20-48 hrs. The conjugation reaction was terminated by adding 2 MEq of sodium borohydride (NaBH_4) to cap unreacted aldehydes. This capping reaction

5 continued at 23 ± 2 °C for 3 ± 1 hrs.

Purifying the Conjugate. The conjugate solution was diluted 1:10 with chilled 5 mM succinate-0.9% saline (pH 6.0) in preparation for purification by tangential flow filtration using 100-300K MWCO membranes. The diluted conjugate solution was passed through a 5 μm filter, and diafiltration was performed using 5 mM succinate / 0.9% saline (pH 6.0) as the medium.

10 After the diafiltration was completed, the conjugate retentate was transferred through a 0.22 μm filter. The conjugate was diluted further with 5 mM succinate / 0.9% saline (pH 6), to a target saccharide concentration of approximately 0.5 mg/mL. Alternatively, the conjugate is purified using 20 mM Histidine-0.9% saline (pH 6.5) by tangential flow filtration using 100-300K MWCO membranes. Final 0.22 μm filtration step was completed to obtain the immunogenic conjugate.

15 See **Table 21, Table 22, Table 23, Table 24 and Table 25.**

EXAMPLE 25: Conjugation in aqueous buffer (RAC/Aqueous), as applied to from *E. coli* serotypes O25B, O1A, O2, and O6

Polysaccharides activation and diafiltration was performed in the same manner as the one for DMSO based conjugation.

20 The filtered activated saccharide was compounded with CRM₁₉₇ at a polysaccharide to protein mass ratio ranging from 0.4 to 2 w/w depending on the serotype. This input ratio was selected to control the polysaccharide to CRM₁₉₇ ratio in the resulting conjugate.

The compounded mixture was then lyophilized. Upon conjugation, the polysaccharide and protein mixture was dissolved in 0.1M sodium phosphate buffer at the polysaccharide

25 concentration ranging from 5 to 25 g/L depending on the serotype, pH was adjusted between 6.0 to 8.0 depending on the serotype. Conjugation was initiated by adding 0.5 – 2.0 MEq of sodium cyanoborohydride to the reaction mixture and incubating at 23 ± 2 °C for 20-48 hrs. The conjugation reaction was terminated by adding 1-2 MEq of sodium borohydride (NaBH_4) to cap unreacted aldehydes.

30 Alternatively, the filtered activated saccharide and calculated amount of CRM₁₉₇ protein was shell-frozen and lyophilized separately, and then combined upon dissolving in 0.1M sodium phosphate buffer, subsequent conjugation can then be proceeded as described above.

Table 20. Summary of results from both conjugations prepared in DMSO and aqueous buffer

	RAC/DMSO	RAC/Aqueous
Poly MW (kDa)	48K	46K
Degree of Oxidation (DO)	12	12
Saccharide/Protein Ratio	0.8	1.0
% Free Saccharide	<5%	32%
Conjugate MW by SEC-MALLS, kDa	7950	260

EXAMPLE 26: Procedure for the preparation of *E. coli* O-antigen polysaccharide-CRM₁₉₇**5 single-ended conjugates**

Lipopolysaccharides (LPS), which are common components of the outer membrane of Gram-negative bacteria, comprise lipid A, the core region, and the O-antigen (also refer to as the O-specific polysaccharide or O-polysaccharide). Different serotype of O-antigen repeating units differ in their composition, structure and serological features. The O-antigen used in this invention is attached to the core domain which contains a sugar unit called 2-Keto-3-deoxyoctanoic acid (KDO) at its chain terminus. Unlike some conjugation methods based on random activation of the polysaccharide chain (e.g. activation with sodium periodate, or carbodiimide). This invention discloses a conjugation process involving selective activation of KDO with a disulfide amine linker, upon unmasking of thiol functional group, it is then conjugated to bromo activated CRM₁₉₇ protein as depicted in **FIG. 31** (Preparation of Single-Ended Conjugates).

Conjugation based on cystamine linker (A1). O-antigen polysaccharide and cystamine (50-250mol. eq of KDO) were mixed in phosphate buffer, adjust pH to 6.0-7.0. To the mixture, sodium cyanoborohydride (NaCNBH₃) (5-30 mol. eq of KDO) was added and the mixture was stirred at 37°C for 48-72hrs. Upon cooling to room temperature and diluted with equal volume of phosphate buffer, the mixture was treated with tris(2-carboxyethyl)phosphine (TCEP) (1.2 mol, eq of cystamine added). The mixture was then purified through diafiltration using 5 KDa MWCO membrane against 10 mM sodium phosphate monobasic solution, to furnish thiol containing O-antigen polysaccharide. The thiol content can be determined by Ellman assays.

The conjugation was then proceeded by mixing above thiol activated O-antigen polysaccharide with bromo activated CRM₁₉₇ protein at a ratio of 0.5-2.0. The pH of the reaction mixture is adjusted to 8.0 -10.0 with 1 M NaOH solution. The conjugation reaction was proceeded at 5 °C for 24 ± 4 hours. The unreacted bromo residues on the

carrier protein were quenched by reacting with 2 mol. eq. of N-acetyl-L-cysteine for 3 - 5 hours at 5 °C. The addition of 3 mol. eq. of iodoacetamide (related to N-acetyl-L-Cysteine added) was then followed to cap the residual free sulfhydryl groups. This capping reaction was proceeded for another 3-5 hours at 5 °C, and pH of both capping steps was maintained at 8.0-10.0 by addition of 1M NaOH. The resulting conjugate was obtained after ultrafiltration/diafiltration using 30 KDa MWCO membrane against 5 mM succinate-0.9% saline, pH 6.0. See **Table 21, Table 22, Table 23, Table 24 and Table 25.**

EXAMPLE 27: Conjugation based on 3,3'-dithio bis(propanoic dihydrazide) linker (A4)

O-antigen polysaccharide and 3,3'-dithio bis(propanoic dihydrazide) (5-50mol. eq of KDO) were mixed in acetate buffer, adjust pH to 4.5-5.5. To the mixture, sodium cyanoborohydride (NaCNBH₃) (5-30 mol. eq of KDO) was added and the mixture was stirred at 23-37°C for 24-72 hrs. The mixture was then treated with tris(2-carboxyethyl)phosphine (TCEP) (1.2 mol, eq of 3,3'-dithio bis(propanoic dihydrazide) linker added). The mixture was then purified through diafiltration using 5 KDa MWCO membrane against 10 mM sodium phosphate monobasic solution, to furnish thiol containing O-antigen polysaccharide. The thiol content can be determined by Ellman assays.

The conjugation was then proceeded by mixing above thiol activated O-antigen polysaccharide with bromo activated CRM₁₉₇ protein at a ratio of 0.5-2.0. The pH of the reaction mixture is adjusted to 8.0 -10.0 with 1 M NaOH solution. The conjugation reaction was proceeded at 5 °C for 24 ± 4 hours. The unreacted bromo residues on the carrier protein were quenched by reacting with 2 mol. eq. of N-acetyl-L-cysteine for 3 - 5 hours at 5 °C. The addition of 3 mol. eq. of iodoacetamide (related to N-acetyl-L-Cysteine added) was then followed to cap the residual free sulfhydryl groups. This capping reaction was proceeded for another 3-5 hours at 5 °C, and pH of both capping steps was maintained at 8.0-10.0 by addition of 1M NaOH. The resulting conjugate was obtained after ultrafiltration/diafiltration using 30 KDa MWCO membrane against 5 mM succinate-0.9% saline, pH 6.0.

EXAMPLE 28: Conjugation based on 2,2'-dithio-N,N'-bis(ethane-2,1-diyl)bis(2-(aminooxy)acetamide) linker (A6)

O-antigen polysaccharide and 2,2'-dithio-N,N'-bis(ethane-2,1-diyl)bis(2-(aminooxy)acetamide) (5-50mol. eq of KDO) were mixed in acetate buffer, adjust pH to 4.5-5.5. The mixture was then stirred at 23-37°C for 24-72 hrs, followed by the addition of sodium cyanoborohydride (NaCNBH₃) (5-30 mol. eq of KDO) and the mixture was stirred for another 3-24 hrs. The mixture was then treated with tris(2-carboxyethyl)phosphine (TCEP) (1.2 mol, eq of linker added). The mixture was then purified through diafiltration using 5 KDa MWCO

membrane against 10 mM sodium phosphate monobasic solution, to furnish thiol containing O-antigen polysaccharide. The thiol content can be determined by Ellman assays.

The conjugation was then proceeded by mixing above thiol activated O-antigen polysaccharide with bromo activated CRM₁₉₇ protein at a ratio of 0.5-2.0. The pH of the reaction mixture is adjusted to 8.0 -10.0 with 1 M NaOH solution. The conjugation reaction was proceeded at 5 °C for 24 ± 4 hours. The unreacted bromo residues on the carrier protein were quenched by reacting with 2 mol. eq. of N-acetyl-L-cysteine for 3 - 5 hours at 5 °C. The addition of 3 mol. eq. of iodoacetamide (related to N-acetyl-L-Cysteine added) was then followed to cap the residual free sulfhydryl groups. This capping reaction was proceeded for another 3-5 hours at 5 °C, and pH of both capping steps was maintained at 8.0-10.0 by addition of 1M NaOH. The resulting conjugate was obtained after ultrafiltration/diafiltration using 30 KDa MWCO membrane against 5 mM succinate-0.9% saline, pH 6.0.

EXAMPLE 29A: Preparation of Bromo activated CRM₁₉₇

The CRM₁₉₇ was prepared in 0.1M Sodium Phosphate, pH 8.0 ± 0.2 solution, and was cooled to 5 ± 3 °C. To the protein solution, the N-hydroxysuccinimide ester of bromoacetic acid (BAANS) as a stock dimethylsulfoxide (DMSO) solution (20 mg/mL) is added in a ratio of 0.25 – 0.5 BAANS: protein (w/w). The reaction is gently mixed at 5 ± 3 °C for 30 – 60 minutes. The resulting bromoacetylated (activated) protein is purified, e.g., by ultrafiltration/diafiltration using 10 kDa MWCO membrane using 10 mM phosphate (pH 7.0) buffer. Following purification, the protein concentration of the bromoacetylated carrier protein is estimated by Lowry protein assay.

Table 21: O1a Conjugates

Conjugate Lot#	132240-112-2	132242-106	132242-124	132242-127	132242-130
Poly Lot#	709756-160	709756-160	709756-160	710958-116	710958-116
Poly Type	Long Chain			Short Chain	
Poly MW (kDa)	33	33	33	11	11
Variant	eTEC	Single-End	RAC/DMSO	Single-End	RAC/DMSO
Activation	8% SH	2.1% SH	DO: 13	6.4% SH	DO: 16
Conjugate Data					
Yield (%)	30	26	77	45	35
SPRatio	0.6	0.5	1.0	0.7	0.6
Free Sacc (%)	9	9	20	5	6
MW (kDa)	1035	331	1284	280	2266

Sacc Conc (mg.mL)	0.31	0.37	0.58	0.59	0.37
Endotoxin (EU/ug)	0.03	0.02	0.01	0.01	0.01
Buffer	5 mM Succ/Saline, pH 6.0				

Table 22: O2 Conjugates

Conjugate Lot#	00709749-0003-1	132242-161	132242-152	132242-159	132242-157
Poly Lot#	709766-33	709766-65		710958-141-2	
Poly Type	Long Chain			Short Chain	
Poly MW (kDa)	36	39		14	
Variant	eTEC	Single-End	RAC/DMSO	Single-End	RAC/DMSO
Activation	6.8% SH	1.6% SH	DO: 17	6.3% SH	DO: 19
Conjugate Data					
Yield (%)	26	33	50	38	36
SPRatio	1.5	0.8	0.8	1.0	0.6
Free Sacc (%)	11	24%	<5	<5	6
MW (kDa)	1161	422	3082	234	1120
Endotoxin (EU/ug)	0.025	0.02	0.01	0.01	0.01
Buffer	5 mM Succ/Saline, pH 6.0				

Table 23: O6 Conjugates

Conjugate Lot#	132240-117-1	132242-134	132242-137	132242-146	132242-145
Poly Lot#	710958-121-1	710958-143-3			
Poly Type	Long Chain			Short Chain	
Poly MW (kDa)	44	15			
Variant	eTEC	Single-End	RAC/DMSO	Single-End	RAC/DMSO
Activation	18% SH	2.2% SH	DO: 16.5	6.1% SH	DO: 22
Conjugate Data					
Yield (%)	27	23	58	48	30
SPRatio	0.78	0.6	0.82	0.7	0.6
Free Sacc (%)	9	4	4	<5	8
MW (kDa)	1050	340	1910	256	2058
Sacc Conc (mg.mL)	0.39	0.45	0.59	0.88	0.41

Endotoxin (EU/ug)	0.03	0.02	0.01	0.004	0.005
Buffer	5 mM Succ/Saline, pH 6.0				

Table 24: O25b Conjugates

Conjugate Lot#	132242-28	132242-98	132240-73-1-1	132240-62-1	132240-81	132242-116	132242-121	132242-27	132242-29
Poly Lot#	709766-28	709766-29	709766-30	709766-30	709766-30	710958-117/118	710958-117/118	709766-28	709766-28
Poly type	Long Chain					Short Chain		Long Chain	Long Chain
Poly MW (kDa)	51	48	48	48	48	14	14	51	51
Variant	RAC/DMSO	Single-End	eTEC	eTEC	eTEC	Single-End	RAC/DMSO	RAC/DMSO	RAC/DMSO
Activation	DO: 18	2.4% SH	10% SH	4% SH	17% SH	6.6% SH	DO: 17	21	12
Conjugate Data									
Yield (%)	82	26	56	32	92	28	18	71	80
SPRatio	0.9	0.82	0.88	0.64	1.32	0.7	0.36	0.81	0.84
Free Sacc (%)	7.2	5	< 5	11	< 5	< 5	< 5	8.3	<5
Conjugate MW (kDa)	4415	840	1057	1029	2306	380	9114	3303	7953
Sacc Conc (mg.mL)	0.7	0.4	0.43	0.36	0.9	0.45	0.19	0.6	0.67
Endotoxin (EU/ug)	0.01	0.02	0.08	0.08	0.01	0.01	0.01	0.02	0.22
Conjugate (DS) Buffer matrix	5 mM Succ/Saline, pH 6.0								

Table 25: O25b K-12 Conjugates

Conjugate Lot#	709749-015-2	709744-0016
Poly Lot#	710958-137	
Poly Type	Long Chain(K12)	
Poly MW (kDa)	44	
Variant	eTEC	RAC/DMSO
Activation	SH: 24%	DO: 19
Conjugate Data		
Yield (%)	59%	33%
SPRatio	1.4	0.83
Free Sacc (%)	5%	5.2%
MW (kDa)	1537	4775
Sacc Conc (mg/mL)	0.91	0.29
Endotoxin (EU/ug)	0.08	0.01
Buffer	5 mM Succ/Saline, pH 6.0	

EXAMPLE 29B: Preparation of *E. coli* O-Ag-TT conjugates

E. coli serotype O25b long polysaccharide, Lot# 709766-30 (about 6.92 mg/mL, MW: about 39kDa), 50 mg, lyophilized was used for Tetanus Toxoid (TT) conjugation.

- 5 *E. coli* serotype O1a long polysaccharide 710958-142-3 (about 6.3 mg/mL, MW: about 44.3 kDa) (50mg, 7.94 mL) was lyophilized.

E. coli serotype O6 long polysaccharide, 710758-121-1 (about 16.8 mg/mL, MW: about 44 kDa) (50mg, 2.98 mL) was lyophilized.

- 10 Each of the lyophilized polysaccharides listed above was dissolved in WFI to make at approx 5-10 mg/mL to it, 0.5 mL (100 mg (1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP) solution in 1 mL acetonitrile) was added and stirred at RT. Triethylamine (TEA) 0.2M (2mL) was added and stirred at RT.

- 15 Preparation of Tetanus toxoid (TT): TT (100 mg, 47 ml) was concentrated to approximately 20 mL and washed twice with saline (2x50mL) using filtration tubes. After that it was diluted with HEPES and saline to make final HEPES conc as about 0.25M. TT was prepared as described above and pH of the reaction was adjusted to about 9.1-9.2. The reaction mixture was stirred at RT.

After 20-24 hrs the reaction was quenched with Glycine (0.5 mL). After that it was concentrated to using MWCO regenerated cellulose membranes and diafiltration was performed against saline. Filtered and analyzed. See **Table 26**.

Table 26

5 Exemplary embodiments:

<i>E. coli</i> serotype O25b–TT conjugate	<i>E. coli</i> serotype O6-TT conjugate
Volume: 41 mL Sacc Conc (Anthrone): 1.122 mg/mL (92% yield) Protein Conc (Lowry): 1.133 mg/mL SPRatio: 0.99 Free Sacc (DOC): 74.7% The product obtained was concentrated to 15 mL using MWCO regenerated cellulose membranes and diafiltration was performed against saline (40X diavolumes). Filtered through 0.22 um filter and analyzed. Volume: 27 mL Sacc Conc (Anthrone): 1.041 mg/mL (56% yield) Protein Conc (Lowry): 1.012 mg/mL SPRatio: 1.03 Free Sacc (DOC): 60.6% (poly recovery 100%)	Volume: 42 mL Sacc Conc (Anthrone): 0.790 mg/mL (66% yield) Protein Conc (Lowry): 1.895 mg/mL SPRatio: 0.42 Free Sacc (DOC): <5% MW (kDa): 1192 Endotoxin (EU/ug): 0.022

EXAMPLE 30: Additional results from O-antigen Fermentation, Purification, and Conjugation

10 The exemplary processes described below is generally applicable to all *E. coli* serotypes. The production of each polysaccharide included a batch production fermentation followed by chemical inactivation prior to downstream purification.

15 **Strains and storage.** Strains employed for biosynthesis of short chain O-antigen were clinical wild type strains of *E. coli*. Long chain O-antigen was produced with derivatives of the short chain-producers that had been engineered by the Wanner-Datsenko method to possess a deletion of the native *wzzb* gene and were complemented by the “long-chain” extender function *fepE* from Salmonella. The *fepE* function was expressed from its native promoter on either a high copy *colE1*-based

“topo” vector or a low copy derivative of the colE1-based vector pET30a, from which the T7 promoter region had been deleted.

Cell banks were prepared by growing cells in either animal free LB or minimal medium to an OD₆₀₀ of at least 3.0. The broth was then diluted in fresh medium and
5 combined with 80% glycerol to obtain a 20% glycerol final concentration with 2.0 OD₆₀₀/mL.

Media used for seed culture and fermentation. The seed and fermentation medium employed share the following formulation: KH₂PO₄, K₂HPO₄, (NH₄)₂SO₄, sodium citrate, Na₂SO₄, aspartic acid, glucose, MgSO₄, FeSO₄·7H₂O, Na₂MoO₄·2H₂O, H₃BO₃, CoCl₂·6H₂O,
10 CuCl₂·2H₂O, MnCl₂·4H₂O, ZnCl₂ and CaCl₂·2H₂O.

Seed and fermentation conditions. Seeds were inoculated at 0.1% from a single seed vial. The seed flask was incubated at 37°C for 16-18 hours and typically achieved 10-20 OD₆₀₀/mL.

Fermentation was performed in a 10L stainless steel, steam in place fermentor.

15 Inoculation of the fermentor was typically 1:1000 from a 10 OD₆₀₀ seed. The batch phase, which is the period during which growth proceeds on the 10 g/L batched glucose, typically lasts 8 hours. Upon glucose exhaustion, there was a sudden rise in dissolved oxygen, at which point glucose was fed to the fermentation. The fermentation typically then proceeds for 16-18 hours with harvest giving > 120 OD₆₀₀/mL.

20 **Initial evaluation of short/long chain O-antigen production for serotypes O1a, O2, O6 and O25b.** Wild type strains for O1a, O2, O6 and O25b were fermented in a supplemented minimal medium in batch mode to an OD₆₀₀ = 15-20. Upon glucose exhaustion, which results in a sudden decrease in oxygen consumption, a growth limiting glucose feed was applied from a glucose solution for 16-18 hours. Cell densities of 124-145 OD₆₀₀ units/mL were reached. The
25 pH of the harvest broths was subsequently adjusted to about 3.8 and heated to 95°C for 2 hours. The hydrolyzed broth was then cooled to 25°C, brought to pH 6.0 and centrifuged to remove solids. The resulting supernatant was then applied to a SEC-HPLC column for quantitation of the O-antigen. Productivities in the range of 2240-4180 mg/L were obtained. The molecular weight of purified short-chain O-antigen from these batches was found to range from
30 10-15 kDa. It was also noted that SEC chromatography of the O2 and O6 hydrolysates revealed a distinct and separable contaminating polysaccharide that was not evident in the O1a and O25b hydrolysates.

Long chain versions of the O1a, O2, O6 and O25b O-antigens were accessed through fermentation of a *wzzb* deletion version of each strain which carried a heterologous,
35 complementing *fepE* gene on a high-copy, kanamycin-selectable topo plasmid. Fermentation was performed as for the short chain, albeit with kanamycin selection. The final cell densities

observed at 124-177 OD₆₀₀/mL were associated with O-antigen productivities of 3500-9850 mg/L. The complementation-based synthesis of long chain O-antigen was at least as productive as in the parental short chain strain and in some cases more so. The molecular weights of purified O-antigen polysaccharide were 33-49 kDa or about 3 times the size of the corresponding short chain.

It was noted that the long chain hydrolysates for O2 and O6 showed evidence of a contaminating polysaccharide peak that, in the case of long chain antigen, was observed as a shoulder on the main O-antigen peak; O1 and O25b showed no evidence of production of a contaminating polysaccharide, as was seen earlier with the short chain parent.

Growth rate suppression was found to be associated with the presence of the topo replicon absent the *fepE*. Additionally, the $\Delta wzzb$ mutation itself had not adverse effect on growth rate, indicating that the disturbed growth rates were conveyed by the plasmid vector.

Evaluation of strains for production of O11, O13, O16, O21 and O75 O-antigen. Multiple wild-type strains of serotypes O11, O13, O16, O21 and O75 were evaluated for their propensity to produce unwanted polysaccharide in fermentation by SEC-HPLC. Strains for O11, O13, O16, O21 and O75 were selected as absent contaminating polysaccharide, as well as for their ability to produce > 1000 mg/L O-antigen and for the display of an antibiotic sensitivity profile that allowed Wanner - Datsenko recombineering for introduction of the $\Delta wzzb$ trait.

Chloramphenicol-selectable versions of topo-*fepE* and pET-*fepE* were constructed that allowed for the introduction of *fepE* into the O11, O13, O16, O21 and O75 $\Delta wzzb$ strains that in general were found to be kanamycin-resistant. The resulting topo-*fepE* and pET-*fepE* bearing strains were fermented with chloramphenicol selection and the supernatant from acid-hydrolyzed broth was evaluated by SEC-HPLC. Both the high (topo) and low copy (pET) *fepE* constructs directed the synthesis of O-antigen with productivities for each that were equivalent to the parental wild-type. Expression of potentially interfering polysaccharides was not observed.

An evaluation of growth rates for wzzb plasmid-bearing strains showed that the O11, O13 and O21 were retarded by the presence of topo-*fepE* but not by pET-*fepE*; strains O16 and O75 strains showed acceptable growth rates irrespective of replicon choice.

Table 27

O- antigen type	IHMA type	short (SC) or long chain (LC)	fep E plasmid type	marker	final cell density OD ₆₀₀	final Oag productivity (mg/L)	MW - kDa	SEC impurity
O1a	wt	SC	None	None	125	2550	11	N
O1a	Δ wzzb/fepE	LC	topo	Kana	130	5530	33	N
O1a	Δ wzzb/fepE	LC	pET	Kana	Not done (ND)	ND	ND	ND
O2	wt	SC	None	None	127	2240	13	Y
O2	Δ wzzb/fepE	LC	topo	Kana	177	3750	49	Y
O2	x	LC	pET	x	NA	NA	NA	NA
O6	wt	SC	None	None	145	4180	16	Y
O6	Δ wzzb/fepE	LC	topo	Kana	124	9850	44	Y
O6	Δ wzzb/fepE	LC	pET	Kana	ND	ND	ND	ND
O11	wt	SC	None	None	194	4720	x	N
O11	Δ wzzb/fepE	LC	topo	Kana	142	7220	x	N
O11	x	LC	pET	x	NA	NA	NA	NA
O13	wt	SC	None	x	113	4770	x	N
O13	Δ wzzb/fepE	LC	topo	cam	101	4680	x	N
O13	Δ wzzb/fepE	LC	pET	cam	108	4600	x	N
O16	wt	SC	None	x	154	1870	x	N
O16	Δ wzzb/fepE	LC	topo	cam	129	1180	x	N
O16	Δ wzzb/fepE	LC	pET	cam	137	1280	x	N
O21	wt	SC	None	x	140	1180	x	N
O21	Δ wzzb/fepE	LC	topo	cam	ND	ND	x	N
O21	Δ wzzb/fepE	LC	pET	cam	131	820	x	N
O25b	2831	SC	None	None	126	3550	10	N
O25b	Δ wzzb/fepE	LC	topo	Kana	152	3500	49	N
O25b	x	LC	pET	x	NA	NA	NA	NA
O75	wt	SC	None	x	149	1690	x	N
O75	Δ wzzb/fepE	LC	topo	cam	132	1500	x	N
O75	Δ wzzb/fepE	LC	pET	cam	138	1520	x	N

The purification process for the polysaccharides included acid hydrolysis to release the O-antigens. A crude suspension of serotype specific *E. coli* culture in fermentation reactor was

5 directly treated with acetic acid to the final pH of 3.5±0.5 and the acidified broth was heated to

- the temperature of $95 \pm 5^\circ\text{C}$ for at least one hour. This treatment cleaves the labile linkage between KDO, at the proximal end of the oligosaccharide and the lipid A, thus releasing the O-Ag chain. The acidified broth that contains the released O-Ag was cooled to $20 \pm 10^\circ\text{C}$ before being neutralized to $\text{pH } 7 \pm 1.0$ using NH_4OH . The process
- 5 further included several centrifugation, filtration, and concentration/ diafiltration operations steps.

Table 28

Serotype (core)	Description	Expected Poly size	Titer (g/L)	Purified Poly M.W. (kDa)	Number of Repeat Units	Increase in M.W. (kDa) over short	NMR	Purified Conjugate M.W. (kDa)	Conjugation Lot #
O25b (R1)	$\Delta\text{wzzB} + \text{LT2FepE}$	Long	5.3	47	55	34	✓	5365	132242-28 (RAC/DMSO)
								1423	132242-98 (Single-end)
								1258	132240-73-1-1 (eTEC)
	$\Delta\text{wzzB} + \text{O25a wzzB}$	Short	2.3	13/14	15	NA	✓	380	132242-116 (Single-end)
								9114	132242-121 (RAC/DMSO)
O25b (K12)	$\Delta\text{wzzB} + \text{LT2FepE}$	Long	3.5	44	51	27	✓	1537	709749-015-2 (eTEC)
								4775	709744-0016 (RAC/DMSO)
	wt	Short	3.5	17	17	NA	✓		
O1a (R1)	$\Delta\text{wzzB} + \text{LT2FepE}$	Long	5.5	33	39	22	✓	1035	132240-112-2 (eTEC)
								331	132242-106 (Single-end)
								1284	132242-124 (RAC/DMSO)

	wt	Short	2.5	11	13	NA	✓	280	132242-127 (Single-end)
								2266	132242-130 (RAC/DMSO)
O2 (R1)	Δ wzzB + LT2FepE	Long	4.9	36	43	22	✓	1161	00707947- 0003-1 (eTEC)
				39	47	25		422	132242-161 (single-end)
								3082	132242-152 (RAC/DMSO)
	wt	Short	2.8	14	17	NA	✓	234	132242-159 (single-end)
								1120	1322421-157 (RAC/DMSO)
O2 (R4)	Δ wzzB + LT2FepE	Long	5.1	NA	NA	NA	NA		
	wt	Short	2.1	14.7	18	NA	✓		
O6 (R1)	Δ wzzB + LT2FepE	Long	6.9	37.2	42	22.2	✓		
	wt	Short	3.5	15	17	NA	✓	256	132242-146 (Single-end_
								2058	123342-145 (RAC/DMSO)
O6 (R1)	Δ wzzB + LT2FepE	Long	8.4	44.4	50	28.2	✓	1050	132240-117- 1 (eTEC)
								340	132242-134 (Single-end)
								1910	132242-137 (RAC/DMSO)
	wt	Short	3.6	16.2	18	NA	✓		

EXAMPLE 31: Conjugation towards O-antigen (O4, O11, O21, O75) studied (RAC/DMSO)**Table 29: O4 conjugates**

Conjugate Lot#	709744-70	709744-73	709744-72
Poly Lot#	709740-168		
Poly MW (kDa)	52		
DO	26	19	15
Act poly Mw (kDa)	51		
Conjugation			
Input SP	1.0	1.0	1.0
SPRatio	0.85	1.0	1.0
Free Sacc (%)	<5%	<5%	<5%
MW (kDa)	4764	4758	3423
Yield (%)	72	80	82
Endotoxin (EU/ug)	0.003	0.001	0.005

Table 30: O11 conjugates

Conjugate Lot#	709744-64	709744-66	709744-65	709744-67
Poly Lot#	709740-162			
Poly MW (kDa)	39			
DO	21		14	
Act poly Mw (kDa)	40			
Conjugation				
Input SP	1.0	1.3	1.0	1.3
SPRatio	0.5	0.64	0.65	0.75
Free Sacc (%)	<5%	<5%	<5%	<5%

MW (kDa)	10520	7580	4814	4338
Yield (%)	30	30	44	38
Endotoxin (EU/ug)	0.005	0.005	0.005	0.005

Table 31: O21 conjugates

Conjugate Lot#	709749-113	709749-111	709749-112	709749-115	709749-116
Poly Lot#	709740-165				
Poly MW (kDa)	40				
DO	25	18		15	
Act poly Mw (KDa)	40	41		40	
Conjugation					
Input SP	1.0	1.0	0.8	1.0	1.25
SPRatio	0.6	0.6	0.5	0.9	1.1
Free Sacc (%)	6%	5%	<5%	12%	7%
MW (kDa)	6920	5961	9729	2403	1960
Yield (%)	31	36	37	52	54
Endotoxin (EU/ug)	0.02	0.02	0.03	0.01	0.009

Table 32: O75 conjugates

Conjugate Lot#	709749-101	709749-102	709749-103
Poly Lot#	709766-080B		
Poly MW (kDa)	48		
DO	18		25
Act poly Mw (kDa)	43		44
Conjugation			

Input SP	1.0	0.8	1.0
SPRatio	0.94	0.76	0.78
Free Sacc (%)	<5%	6%	6%
MW (kDa)	2304	2427	5229
Yield (%)	62	65	45
Endotoxin (EU/ug)	0.02	0.01	0.01

EXAMPLE 32: PLL conjugates prepared**Table 33**

Serotype	O11	O75	O21	O4
Conjugate Lot#	00707779-0413	00707779-0414	00707779-0415	00707779-0416
Poly Lot#	709740-162	709766-080B	709740-165	709740-168
Poly MW (kDa)	39	48	40	52
Conjugate Data				
SPRatio	13.5	16.8	18.1	21.2
Free Sacc (%)	9.8%	<5%	<5%	6.9%
Sacc Conc	789 µg/mL	676 µg/mL	978 µg/mL	837 µg/mL
PLL Conc	58.3 µg/mL	40.3 µg/mL	54.0 µg/mL	39.4 µg/mL
Endotoxin (EU/ug)	0.002	0.002	0.005	0.004
Conjugate (DS) Matrix	1X PBS, 1M NaCl			

5 EXAMPLE 33: Stable mammalian cell expression of E. coli polypeptides

Stable CHO clones expressing FimH GSD or FimH LD were generated using a SSI (Site Specific Integration) stable expression system. The host CHO cell is an engineered cell line from a CHOK1SV GS-KO background (see, for example, United States Patent Application 20200002727, for a description of the CHOK1SV GS-KO host cell line). Briefly a landing pad

with green fluorescent protein (GFP) gene surrounded by two FRT sites were targeted into a transcription hot spot in the genome of the host cell. The GFP gene can be exchanged with GS gene and the gene of interest which are also surrounded by FRT sites from the LVEC vector co-expressed with flippase recombinase (FLPe). This system not only has growth and productivity profiles that compare favorably with random integration but also displays genotypic and phenotypic stability to at least 100 generations.

As referred to herein, the term "FRT site" refers to a nucleotide sequence at which the product of the flippase (FLP) gene of the yeast 2 μ m plasmid, FLP recombinase, can catalyze a site-specific recombination. A variety of non-identical FRT sites are known to the art. The sequences of the various FRT sites are similar in that they all contain identical 13-base pair inverted repeats flanking an 8-base pair asymmetric core region in which the recombination occurs. It is the asymmetric core region that is responsible for the directionality of the site and for the variation among the different FRT sites. Illustrative (non-limiting) examples of these include the naturally occurring FRT (F), and several mutant or variant FRT sites such as FRT F1 and FRT F2.

As referred to herein, the term "landing pad" refers to a nucleic acid sequence comprising a first recombination target site chromosomally-integrated into a host cell. In some embodiments, a landing site comprises two or more recombination target sites chromosomally-integrated into a host cell. In some embodiments, the cell comprises 1, 2, 3, 4, 5, 6, 7, or 8 landing pads. In some embodiments, the cell comprises 1, 2, or 3 landing pads. In some embodiments, the cell comprises 4 landing pads. In some embodiments, landing pads are integrated at up to 1, 2, 3, 4, 5, 6, 7, or 8 distinct chromosomal loci. In some embodiments, landing pads are integrated at up to 1, 2, or 3 distinct chromosomal loci. In some embodiments, landing pads are integrated at 4 distinct chromosomal loci.

The LVEC expression vector for FimH GSD or FimH LD and the FLPe expression vector were co-transfected into a SSI host cell by electroporation either with BioRad Gene Pulser Xcell or Amaxa 4D-Nucleofector. Then cells were cultured in media without glutamine to select cells that has GS gene integrated at the landing pad site. Usually cells recover in 2-3 weeks. Then single cell cloning were carried out in 96 well plates either by FACS or limiting dilution. Titers from wells with cells were ranked to narrow down to top 48 clones. A second round of fed batch screening in 24 deep-well plates was conducted to narrow down the clones to top 12. A third round of fed batch screening in Ambr15 was executed to narrow down the clones to top 3. Ambr250 experiments were used to identify the best clone. Master cell bank and working cell bank were generated for the top clone after its identification.

EXAMPLE 34: Cell line development and production reactor expression of FimH-DSG WT and FimH_{LD} WT proteins

The example described herein, describes an exemplary production of both FimH-DSG WT and FimH_{LD} WT proteins from stable CHO cell lines, where the coding sequences for each protein has been stably intergraded into the CHO genome.

In a production bioreactor setting, the stable CHO cell lines selected were able to produce the target protein at around 1gram per liter of culture for FimH-DSG WT, and 250 miligrams per liter of culture for FimH_{LD} WT. The seed train for the production reactor was continuously scaled up from vial thaw of a working cell bank and expanded in shake flasks using an inoculation viable cell density of 0.3×10^6 cells/ml through three passage cycles in shake flasks to provide enough cells for the production reactor. The cells were grown at 36.5 deg C, at 5% CO₂ for three-four days.

The production reactor was seeded from the final shake flask, targeting an inoculation cell density of 1×10^6 cells/ml. The production reactor was grown at 36.5 deg C for seven days, using a pH of 7.05 (+/- 0.15), and targeting a CO₂ saturation of 5-10%. pH is controlled by sodium/potassium bicarbonate for base control, and CO₂ sparge for acid control. Dissolved oxygen is controlled at a setpoint of 40% using pure oxygen through the sparge. The temperature was adjusted to 31 deg C on day seven. The reactor was fed on day 1 using a feed strategy that adds feed in correlation to the viable cell density, this is achieved by using a feed factor of 0.75 in order to ensure feed components do not run out during the run. The feed is then added continuously to provide the desired volume of feed over the course of the day.

The production reactor was harvested on day 13, and the harvest culture was centrifuged and 0.22 µm filtered, prior to downstream processing.

EXAMPLE 35: Immunogenicity of FimH Antigens and O-antigen Combinations in Mice

E. coli FimH lectin binding domain (FimH_{LD}) and full-length (FimH-DSG or FimCH) variants were evaluated in mouse immunogenicity studies that assessed the ability of elicited antibodies to neutralize the binding of fimbriated *E. coli* to mannosylated ligands. Wild-type FimH antigens that were secreted and purified at high yield from mammalian cells were similarly immunogenic in mice as analogous native antigens purified at low yield from the *E. coli* periplasm. By comparison, full-length FimH antigens containing both the lectin and pillin domains were significantly more immunogenic than FimH_{LD}. Use of an adjuvant formulation containing QS21 was required to generate robust functional immune responses to these fimbrial antigens. A follow-up study investigated the immunogenicity of FimH in combination with a four-valent O-antigen glycoconjugate mixture and different adjuvants. After a second dose, mice

vaccinated with a liposomal QS21 formulation induced consistently higher functional responses to the combined antigens in both FimH neutralization and O-antigen-specific OPA assays compared with groups dosed with no adjuvant or liposomal monophosphoryl lipid A (MPLA).

Animal Immunogenicity. 6-8 week old CD-1 mice were obtained from Charles River and groups of 10 or 20 animals were dosed subcutaneously (SC) with 0.1 mL of test antigen or buffer control. Polyclonal anti-FimH rabbit control serum was prepared by immunizing rabbits with three doses of *E. coli* FimH_{LD} antigen with 100 µg complete Freund's adjuvant (at wk 0) followed by antigen in 100 µg incomplete Freund's adjuvant (at wks 4 and 8) (Covance). Each vaccination administered 0.5 mLs containing 50 µg of antigen at two sites (sub-cutaneous-

Dorsal).

Adjuvant Formulations. Diluent for unadjuvanted antigens was 10 mM phosphate buffer, pH 6.1. For AlPO₄, 50 µg (100 µL of a 0.5 mg/mL suspension) was given per mouse dose. The default dose of Quillaja Saponaria-21 (QS21) used in preclinical studies was 20 µg per mouse from a stock containing 5.1 mg/mL QS-21, 5 mM Succinate, 60mM NaCl, 0.1% PS80, pH 5.6. Liposomal formulations of monophosphoryl Lipid A (MPLA, Synthetic, PHAD®, Avanti) and QS21 were prepared with 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and cholesterol (Avanti polar lipids). Liposomal MPLA was used at 5 µg per dose from stock comprising 15 mM phosphate buffer, pH 6.1, 4 mg/mL DOPC, 1 mg/mL Cholesterol, 0.2 mg/mL MPLA (Lot 00714551-0018-2XLipoMPL), having a liposomal particle size of 71 nm determined by dynamic light scattering. A liposomal MPLA/QS21 formulation was used at 5 µg each per dose from stock comprising 15 mM phosphate buffer, pH 6.1, 4 mg/mL DOPC, 1 mg/mL Cholesterol, 0.2 mg/mL MPLA, and 0.2 mg/mL QS-21 (Lot 00714551-0018-2XLipoMQ), having particle size of 75 nm for MPLA-QS21 liposomes determined by dynamic light scattering.

Expression Plasmids and Derived FimH Fimbrial Antigens used in Immunogenicity Studies. The following expression plasmids were used to generate recombinant protein for mouse experiments:

>pSB02083 -- FimH_{LD} (mIgK signal pept, N28S, N91S), processed protein sequence:
 FACKTASGTAIPIGGGSANVYVNLAPCVNVGQNCVVDLSTQIFCHNDYPETITDYVTLQRGSAY
 GGVLSFSFGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAV
 LILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLDPYPGSVPIPLTVYCAKSQ
 NLGYLLSGTTADAGNSIFTNTASFSPAQGVGVQLTRQGTIIPANNTVSLGAVGTSASVSLGLTAN
 YARTGGQVTAGNVQSIIGVTFVYQGGSSGGGADVITVNGKVAKGGHHHHHHHHH (SEQ ID
 NO:110);

>pSB02158 -- FimH_{LD}-LM (mIgK signal pept, N28S N91S V48C L55C), processed protein sequence:

FACKTASGTAIPIGGGSANVYVNLAPCVNVGQNCVVDLSTQIFCHNDYPETITDYVTLQRGSAY
GGVLSSFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAV
5 LILRQTNNYNSDDFQFVWNIYANNDVVVPTGGHHHHHHHHH (SEQ ID NO:111);

>pSB02307 – FimH-DSG (mIgK signal pept, N28S N91S N249Q 7aa linker FimG A1..K14), processed protein sequence:

FACKTASGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAY
10 GGVLSSFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAV
LILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTLPDYPGSPVPIPLTVYCAKSQ
NLGYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRQGTIIPANNTVSLGAVGTSVAVSLGLTAN
YARTGGQVTAGNVQSIIGVTFVYQGGSSGGGADVITVNGKVVAKGHHHHHHHHH (SEQ ID
NO:112);

15

>pSB02198 – FimH-DSG-LM (mIgK signal pept, N28S N91S 249Q V48C L55C 7aa linker FimG A1..K14), processed protein sequence:

FACKTASGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAY
GGVLSSFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLPVSSAGGVAIKAGSLIAV
20 LILRQTNNYNSDDFQFVWNIYANNDVVVPTGGHHHHHHHHH (SEQ ID NO:113).

Amino acid numbering is based on the full-length amino acid sequence of the J96 strain FimH, whereas academic publications use numbering based on the N-terminal Phenylalanine (in position 21) that is exposed after signal peptide processing. Other mutations remove
25 glycosylation site asparagine residues in the lectin domain (N28S and N91S), in the pilin domain of the full-length FimH-DSG construct (N249Q). The donor-strand G peptide that is added to the C-terminus stabilizes FimH. The 7 amino acid linker length was optimized for mammalian expression as described in the above Examples.

FimH Whole Cell Neutralization Assays. FimH neutralization assays were developed
30 to measure the inhibition by serum antibodies of the binding of fimbriated *E. coli* to mannoside ligands. Ligands were either microplate-immobilized yeast mannan or bladder 5637 cells that express the mannosylated uroplakin receptor UPIa.

For the yeast mannan assay black microtiter 96-well plates (Maxisorb, Nunc) were coated with 20 µg/ml of yeast mannan (Sigma-Aldrich) in PBS buffer. The wells were blocked
35 with 1% bovine serum albumin (BSA, Sigma-Aldrich) in PBS for 20 min. The human bladder epithelial cell line 5637 was obtained from ATCC (ATCC HTB-9). Cells were grown on black

tissue-culture microplates (Greiner) in RPMI 1640 (Sigma, St Louis, MO) supplemented with 10% fetal bovine serum (FBS, Sigma), 2.0 g/l sodium bicarbonate (Sigma) and 0.3 g/l L-glutamine, grown at 37°C with 5% CO₂ and utilized between passages 10 and 24. Surface expression of Uroplakin 1a receptor was confirmed by immunofluorescence staining with polyclonal antibody (Novus #NBP214694). *E. coli* serotype O25b UTI strain PFEEC0547 was serially passaged in 10 mL static LB cultures at 37°C to induce FimH expression. Expression of FimH on the bacterial surface was confirmed by flow cytometry with rabbit immune serum to FimH_{LD} antigen. Specificity of bacterial binding to mannan or bladder cells was established by the inclusion of negative control compound methyl α -D-mannopyranoside (Sigma) which reduced binding by >95% at 50 mM levels. Eight-step two-fold serial dilutions of test sera starting at 1:100 (in PBS, 0.1% BSA) were co-incubated with 1×10^7 *E. coli* for 1 h at 37°C before adding to immobilized yeast mannan or 5637 cell monolayers. Serially diluted anti-FimH_{LD} rabbit serum was used as an internal standard on each plate. Plates were incubated for 1 h at 37°C before washing away unbound bacteria. Bound *E. coli* were stained for 45 min at RT with 3 μ g/ml of an O25b-specific mAb conjugated to Alexafluor 488. The mAb was reconstructed from variable light and heavy chain sequences of O25b antibody 3E9-11 (Nagy E, Nagy G, Szijarto V, et al. Antibodies to multi-drug resistant Escherichia coli: Arsanis Biosciences GmbH, Austria. 2014:76pp). The human pTT5 IgG expression plasmid was used as cloning vector. The fluorescence intensity of individual wells was read on a ClarioStar Plus instrument. IC₅₀ inhibition values were interpolated using sigmoidal dose response variable-slope curve fitting (Graphpad Prism). Titers are the reciprocal of the serum dilution at which half-maximal inhibition is observed. A vaccine antigen responder was defined as a neutralizing titer that exceeds 80% inhibition at the starting serum dilution of 1:100. In addition, the serum dilution titration of binding activity must satisfy variable slope sigmoidal curve fitting parameters ($R^2 > 0.95$, with interpolated Log IC₅₀ value or trigger an experimental repeat with a broader dilution until resolved). The statistical significance (*p*-value) of differences in responses between groups was determined using an unpaired t-test with Welch's correction applied to log-transformed data.

For experiments involving mice vaccinated with multivalent O-antigen conjugates of different serotypes the yeast mannan assay was adapted for use with the Bactiter Glo detection reagent (Promega), to avoid potential interference by elicited anti-O-Ag IgG with bacterial detection. In this case detection is based on photon emission catalyzed by firefly luciferase following release of ATP from permeabilized viable bacteria. Use of this reagent simplified detection by eliminating the wash step and facilitated implementation of a higher throughput 384-well assay format with liquid handling automation (Bravo, Agilent). The miniaturized assay reduced reaction volumes from 100 μ L to 30 μ L and the required number of bacteria from 1×10^7

per well to 1×10^6 . After diluting 1:1 in PBS, Bactiter Glo was added to all wells and after mixing, read on the ClarioStar reader in luminometer mode. The serotype O6:K2 *E. coli* UTI strain CFT073 (ATCC® 700928™) was selected for use in the semi-automated 384-well assay due to its enhanced sensitivity compared with O25b UTI strain PFEEC0547. Both strains yielded

5 similar results in bridging studies comparing individual serum titers in a subset of mouse vaccination groups (data not shown).

***E. coli* O-Ag Mouse IgG Direct Luminex Immunoassay (dLIA).** Long chain *E. coli* O-Ag polysaccharides of serotypes O25b, O1a, O2 and O6 produced in-house by eBPD (Pfizer) were conjugated to poly-L-lysine and then to Luminex bead microspheres with EDC/NHS. Use

10 of beads with distinct spectral addresses for each O-antigen enabled four-fold multiplexing. Beads were incubated with serially diluted individual mouse sera or control mAbs with shaking at 4°C for 18h. After washing, bound serotype-specific IgG was detected with a PE-conjugated goat anti-mouse IgG mouse secondary antibody (90 min RT incubation). Microplates were read

15 on a FlexMap 3D instrument (Biorad). Serotype specific mouse IgG mAbs with similar binding properties (generated in-house) were used as internal standards to quantify IgG levels. Plots of standard curves for each mAb yielded overlapping linear slope profiles across 10^3 serum dilutions (log luminescence vs log serum dilution). A lower limit of quantitation (LLOQ) for the 4-plex IgG dLIA of 0.15259 ug/mL was calculated from standard curve bias, which was the same for each of the four antigens.

20 ***E. coli* Opsonophagocytic Assays (OPA).** Clinical *E. coli* invasive blood isolates were obtained from the Pfizer-sponsored Antimicrobial Testing Leadership and Surveillance (ATLAS) database which is maintained by the International Health Management associates (IHMA) clinical lab. Strains were genotypically characterized by WGS using the Illumina Miseq platform, including *in-silico* serotyping for prediction of O-antigen and K-capsule types. Serotype O1a,

25 O2, O6 and O25b strains that express both O-antigen and K-antigens when grown in Dulbecco's Modified Eagle's medium (DMEM) under cell-banking conditions were used as test strains, as well as multidrug-resistant unencapsulated strains representing O6 and O25b serotypes. O-antigen expression was confirmed by flow cytometry with antigen-specific monoclonal or polyclonal antibodies generated in-house. K1-antigen expression was confirmed

30 using K1-capsule mAb mAb13D9-151 (obtained from NRC, Canada). K2-antigen expression was detected with K2-typing serum (Statens Serum Institute). Expression of the K5 heparosan capsule was confirmed by increased surface exposure of underlying O-antigen upon treatment with heparinase. The following *E. coli* clinical strains were developed for OPA assays:

O1a:K1 (PFEEC0435, ST-95),

35 O2:K1 (PFEEC0146, ST-95, Ticarcillin/Clavulanic Acid^R),

O6:K- (PFEEC0412, ST-127, Ticarcillin/Clavulanic Acid^R, ESBL),

O6:K2 (PFEEC0150, ST-127),
 O25b:K- (PFEEC0068, ST-131, Imipenem^R, fluoroquinolone^R),
 O25b:K5 (PFEEC0066, ST-131, fluoroquinolone^R).

Pre-frozen bacterial stocks were grown in DMEM to an OD₆₀₀ of between 0.5 and 1.0 and glycerol added to a final concentration 20% prior to freezing. Pre-titrated thawed bacteria were diluted to 1 X 10⁵ CFU/ml in OPA buffer (Hanks Balanced Salt Solution (Life Technologies), 0.1% gelatin, 1mM MgCl₂, 2.5mM CaCl₂) and 20 µL (10³ CFU) of the bacterial suspension was opsonized with 10 µL of serially diluted sera for 30min at RT in a 384-well tissue culture microplate. Conditions, such as the need for the opsonization step, the level of baby rabbit complement, the length of assay incubation, and HL60 to bacteria ratio, varied slightly and was optimized according to the assay strain used. Subsequently, 10 µL of 2-6% complement (Baby Rabbit Serum, Pel-Freez) and 10 µL of HL-60 cells (at 100-200:1 ratio) were added to each well and the mixture shaken at 2000 rpm for 45-60 min at 37°C in a 5% CO₂ incubator. Ten µL of each 50 µL reaction was transferred into the corresponding wells of a prewetted 384-well Millipore MultiScreen HTS HV filter plate containing 50 µL water. After vacuum filtering the liquid, 50 µL of 50% DMEM was applied and filtered and plate incubated overnight at 37°C in a sealed zip-lock bag. The next day the colonies were enumerated after staining with Coomassie dye using an ImmunoSpot® analyzer and ImmunoCapture software. To establish the specificity of OPA activity, immune sera were preincubated with 20 µg/mL of the homologous serotype purified O-antigen prior to the opsonization step. The OPA assay includes control reactions without HL60 cells or complement, to demonstrate dependence of any observed killing on these components. Individual serum OPA titers were calculated using variable slope curve fitting (Excel). Combined data were plotted using GraphPad Prism to generate GMTs and associated p values for significance (unpaired t- test with Welch's correction from log transformed data).

Results

Mouse Immunogenicity Studies. A first study was designed to evaluate the immunogenicity of recombinant FimH_{LD} antigens expressed and purified from the *E. coli* periplasm at different dose levels and with different adjuvants. Purification of FimH_{LD} was based on a previously described method (Schembri MA, Hasman H, Klemm P. Expression and purification of the mannose recognition domain of the FimH adhesin. FEMS microbiology letters 2000; 188:147-51). A second study compared the activity of *E. coli* FimH_{LD} with mammalian expressed variants of FimH_{LD} and full-length FimH-DsG. Finally, a third study investigated immunogenicity of combinations of FimH-DsG with a 4-valent long O-antigen glycoconjugate mixture and different adjuvant formulations.

Immunogenicity of FimH Antigens. Induction of functional anti-FimH antibodies induced by the various FimH constructs was assessed using neutralization assays similar to those described previously based on inhibition of binding by *E. coli* to surface immobilized yeast mannan (Kisiela DI, Avagyan H, Friend D, et al. Inhibition and Reversal of Microbial Attachment by an Antibody with Parasteric Activity against the FimH Adhesin of Uropathogenic *E. coli*. PLOS Pathogens 2015; 11:e1004857) or bladder epithelial cells (Langermann S, Palaszynski S, Barnhart M, et al. Prevention of mucosal Escherichia coli infection by FimH-adhesin-based systemic vaccination. Science 1997; 276:607-11; Starks CM, Miller MM, Broglie PM, et al. Optimization and Qualification of an Assay that Demonstrates that a FimH Vaccine Induces Functional Antibody Responses in Women with Histories of Urinary Tract Infections. Human Vaccines & Immunotherapeutics 2020:1-10). The functional potency of antigen variants was quantified by determining serum neutralization titers that prevent the binding of fimbriated bacteria to the yeast mannoside ligands. The titer is defined as the reciprocal of the serum dilution at which 50% of bacteria are bound to the assay microplate in an eight-point two-fold titration of vaccinated animal sera. Results of an initial study with FimH_{LD} antigens purified from *E. coli* are shown in **FIG. 36A** and **FIG. 36B** and **Table 34**. The vaccination schedule is shown in **FIG. 36A** and individual mouse responses plotted in **FIG. 36B**.

Mouse Immunogenicity Experiment 1 – Comparison of Different *E. coli*

Periplasmic Produced FimH Constructs. After vaccination with three 3µg doses of the FimH_{LD} antigen, only 25% of mice in the unadjuvanted group or group adjuvanted with 50 µg of AlPO₄ yielded neutralizing antibody titers in the yeast mannan assay. In contrast, 61% of mice in the group adjuvanted with 20 µg QS21/PS80, yielded neutralizing titers. In this case the geometric mean titer (GMT) was 5-fold higher (*p* value of <0.05) than the other groups. At the 30 µg FimH_{LD} dose level with QS21/PS80, a further threefold increase in GMT and a 78% seroresponder rate was observed. In this adjuvant context, the presence of the disulfide mutation, designed and predicted to enhance functional immunogenicity by locking the FimH_{LD} antigen in an open conformation (Rodriguez VB, Kidd BA, Interlandi G, et al. Allosteric coupling in the bacterial adhesive protein FimH. J Biol Chem 2013; 288:24128-39; Kisiela DI, Rodriguez VB, Tchesnokova V, et al. Conformational inactivation induces immunogenicity of the receptor-binding pocket of a bacterial adhesin. Proceedings of the National Academy of Sciences of the United States of America 2013; 110:19089-94) did not improve immunogenicity compared with the wild-type FimH_{LD} construct. An approximately three-fold lower GMT was observed for the lock-mutant compared with wild-type that bordered on statistical significance (*p* = 0.09). A key observation from this study is that the *E. coli* FimH_{LD} antigen requires a potent QS21 adjuvant formulation to elicit robust levels of neutralizing antibodies in mice.

Table 34. VAC-2019-PRL-EC-1369 Neutralizing Titer GMTs and Responder Rates

Vaccination Group	% Responders	# of Responders	Geometric mean (IC ₅₀) titers
3ug FimH _{LD} no adjuvant	25	3/12	88
3ug FimH _{LD} AlPO ₄	25	5/20	93
3ug FimH _{LD} QS21/PS80	61	11/18	495
30ug FimH _{LD} QS21/PS80	78	14/18	1486
30ug FimH _{LD} LM ^a - QS21	58	11/19	459

a. LM is the disulfide lock mutation V48C L55C

Immunogenicity of Periplasmic Compared to Mammalian Produced FimH

Constructs. In the second study the immunogenicity of FimH_{LD} and full-length FimH proteins expressed in *E. coli* were compared with analogous variants expressed in Expi293 mammalian cells. *E. coli*-expressed full-length FimH complexed with its periplasmic chaperone FimC was used as a benchmark. Neutralization data comparing the relative level of functional antibody induction by *E. coli* vs mammalian expressed antigens with or without the lock-mutation after three vaccinations are shown in **FIG. 37A** and **FIG. 37B** and **Table 35**. Antigens were dosed at 10 µg each with 20 µg QS21/PS80 according to the schedule shown in **FIG. 37B**. The *E. coli*-expressed FimH_{LD} group of mice yielded a neutralization assay GMT of 300 and 55% responder rate, which was not statistically different from the analogous mammalian expressed FimH_{LD} group (with GMT of 194 and 45 % responder rate). The mammalian full-length FimH-DSG was significantly more immunogenic than its FimH_{LD} counterpart, showing a two-fold higher GMT (529 vs 194) and improved seroresponder rate (75% vs 45%). As with the first study, the presence of the lock mutation offered no benefit in terms of generating a functional neutralizing response, either in the context of FimH_{LD} (*E. coli* or mammalian expressed) or full-length FimH-DSG. Mammalian expressed FimH-DSG (with or without the lock mutation) also presented as more immunogenic than the *E. coli* FimCH antigen, although the differences were not statistically significant. Taken together these results identified full-length FimH-DSG, as the antigen of choice for subsequent investigation. We speculate that the presence of the FimH pilin domain in this construct, in complex with the C-terminal FimG donor strand peptide, may provide additional functional epitopes and/ or enhanced stability.

Table 35. VAC-2019-PRL-EC-1438 Neutralizing Titer GMTs and Responder Rates

Vaccination Group	% Responders	# of responders	Geometric mean (IC ₅₀) titers
Periplasmic- FimH _{LD}	55	11/20	300

Periplasmic- FimH _{LD} -LM	40	8/20	254
Mammalian- FimH _{LD}	45	9/20	194
Mammalian- FimH _{LD} -LM	10	2/20	73
Mammalian FimH-DSG	75	15/20	529
Mammalian FimH-DSG-LM	80	16/20	579
Periplasmic- FimCH	67	12/18	354

Immunogenicity of Coformulated Mammalian-generated FimH-DSG and 4-valent O-antigen Glycoconjugate in Mice. The third mouse study was designed to assess the impact of combined O-antigen and adjuvant formulations on FimH-DSG immunogenicity. Long O-antigen glycoconjugates of CRM₁₉₇ representing four most prevalent serotypes O25b, O1a, O2 and O6 were mixed and administered at a dose of 2 µg each as a baseline control group (FIG. 38). Other vaccine groups evaluated the impact on immunogenicity of three adjuvants with 10 µg of FimH-DSG, either individually or in combination with the four-valent O-antigens. Adjuvants were the default 20 µg QS21/PS80 per dose formulation (used in FimH studies 1 and 2); a liposomal synthetic MPLA formulation (5 µg per dose); and a liposomal synthetic MPLA /QS21 formulation (at 5 µg each per dose). Sera were tested in functional FimH neutralization and O-antigen specific opsonophagocytic (OPA) assays at post-dose 2 (PD2) and post-dose 3 (PD3) timepoints.

Results of yeast mannan FimH binding neutralization assays at PD2 and PD3 timepoints are shown in FIG. 39A and FIG. 39B with GMTs and responder rates summarized in Table 36.

Table 36. VAC-2020-PRL-EC-1679 Combination Study Yeast Mannan Binding Neutralization Assay Responder Rates and GMTs

Vaccination Group	% Responders		Geomean of IC50s		# of responders (out of 10)	
	PD2	PD3	PD2	PD3	PD2	PD3
4plex O-antigen combo - No adjuvant	0	0	50	50	0	0
FimH-DSG (10µg)- QS21 (20µg)	100	100	3938	1457	10	10
FimH-DSG (10µg)- MPLA (5µg)	90	90	1243	559	9	9
FimH-DSG (10µg)- MPLA/QS21 (5µg/5µg)	70	90	1075	990	7	9

Table 36. VAC-2020-PRL-EC-1679 Combination Study Yeast Mannan Binding Neutralization Assay Responder Rates and GMTs

Vaccination Group	% Responders		Geomean of IC50s		# of responders (out of 10)	
	PD2	PD3	PD2	PD3	PD2	PD3
FimH-DSG (10µg) + 4plex O-antigen combo - No adjuvant	40	100	169	581	4	10
FimH-DSG (10µg) + 4plex O-antigen combo – MPLA (5µg)	40	70	176	397	4	8
FimH-DSG (10µg) + 4plex O-antigen combo - MPLA/QS21 (5µg/5µg)	90	90	1332	1223	9	9

As antibody responses to individual FimH and monovalent serotype O25b antigens were sub-maximal after two doses in previous experiments, it was anticipated that the PD2 timepoint would best differentiate effects of vaccine antigen and adjuvant composition. Indeed, by PD3 responder rates and FimH neutralizing GMTs were not significantly different between any of the groups ($p > 0.05$). At PD2, the liposomal adjuvant formulations had no significant effect on groups with FimH-DSG alone. In contrast, for the groups vaccinated with combined FimH-DSG/O-antigens, significantly lower FimH neutralizing GMTs were observed with the no adjuvant or liposomal MPLA groups, compared with the group of mice vaccinated with liposomal MPLA/QS21 or any of the FimH-DSG alone groups. PD1 sera were also tested but none of the mice showed neutralizing titers.

The bladder cell binding neutralization assay was used to test sera pooled from each vaccine group at PD2 and PD3 timepoints. Bladder 5637 cells constitutively express the mannosylated uroplakin receptor UPIa which is a natural ligand for FimH (Thumbikat P, Berry RE, Zhou G, et al. Bacteria-induced uroplakin signaling mediates bladder response to infection. PLoS pathogens 2009; 5:e1000415-e; Kątnik-Prastowska I, Lis J, Matejuk A. Glycosylation of uroplakins. Implications for bladder physiopathology. Glycoconjugate journal 2014; 31:623-36). Results shown in **FIG. 40A** and **FIG. 40B** and **Table 37** were consistent with the findings of the yeast mannan neutralization assay run with individual sera. Again, combined FimH-DSG/O-antigens formulated with liposomal MPLA or no adjuvant generated significantly weaker neutralizing responses than the liposomal MPLA/QS21 group, which responded similarly to the groups of mice vaccinated with QS21/PS80 or liposomal MPLA/QS21 and FimH-DSG without

O-antigens. A minor difference between the mannan and bladder neutralization assay data was a trend in the latter toward weaker responses with the liposomal MPLA formulation for groups vaccinated with either FimH-DSG only or combined FimH-DSG/O-antigens.

Table 37. VAC-2020-PRL-EC-1679 Bladder Cell Neutralization Titers with Mouse Serum Pools

Vaccination Group	PD2 IC50	PD3 IC50
4V O-Ag	Inactive	Inactive
FimH-DSG QS21	3656	1140
FimH-DSG MPLA	775	750-1000 ^a
FimH-DSG MPLA/QS21	2458	1198
FimH-DSG + 4V O-Ag no adj	337	1177
FimH-DSG + 4V O-Ag MPLA	Inactive	434
FimH-DSG + 4V O-Ag MPLA/QS21	3447	931
Rabbit serum control	3266	3809

a. IC₅₀ titer estimate due to poor fit of biphasic curve with variable slope equation.

5

Levels of O-antigen specific total IgGs generated with the 4-plex dLIA after vaccination of four groups of mice with the 4-valent O-antigens alone or in combination with FimH-DSG and liposomal adjuvants are shown in **FIG. 41** and **Table 38**. Serotype O25b-specific IgG responses after two doses were markedly lower than those elicited by serotype O1a, O2 and O6 O-antigens, which is consistent with results of previous conjugate chemistry studies with monovalent O-antigens (not shown) and reported preclinical studies with tetravalent bioconjugate O-antigens (van den Dobbelsteen G, Faé KC, Serroyen J, et al. Immunogenicity and safety of a tetravalent *E. coli* O-antigen bioconjugate vaccine in animal models. Vaccine 2016; 34:4152-60). The lowest level of anti-O-Ag IgG across the O25b groups at PD2 was observed with the unadjuvanted FimH-DSG and O-antigen combination, which yielded only a single titer exceeding the five-fold over baseline IgG responder threshold; in comparison, the unadjuvanted O-antigen only group generated higher responder rates of 56% (5/9), although the difference in GMTs (0.79 vs 0.24 µg/ mL IgG) between the two groups was not statistically significant ($p > 0.5$). At PD2, the liposomal MPLA/QS21 adjuvant significantly enhanced the immunogenicity of the FimH-DSG and O-antigen combination by inducing a 16-fold increase in GMT (3.84 vs 0.24) and improved responder rate (70% vs 10%); a significant increase in GMT was also observed at PD3 (40.78 vs 3.75). A responder is defined as mouse with serum that yields a complete dilution dependent killing response that meets variable slope curve fitting

10

15

20

- criteria. Despite higher responder rates and titers across the groups than for serotype O25b, serotype O1a IgG titers for the FimH-DSG and O-antigen combination were also significantly improved at PD2 and PD3 by liposomal MPLA/QS21 by 5.9 and 4.6-fold, respectively. The impact of the MPLA/QS21 on O2 and O6 O-antigens was less pronounced than for O1a and O25b due to smaller differences overall in GMTs between the vaccination groups.

Table 38. VAC-2020-PRL-EC-1679: Influence of Adjuvant and FimH-DSG on O-antigen Specific Serum IgG GMTs

Vaccination Group	O1a		O2		O6		O25b	
	PD2	PD3	PD2	PD3	PD2	PD3	PD2	PD3
4v O-Ag	13.86	56.72	21.96	145.99	9.66	74.25	0.79	13.73
FimH-DSG + 4v O-Ag no adjuvant	13.42	39.05	20.56	86.81	19.19	101.48	0.24	3.75
FimH-DSG + 4v O-Ag in liposomal MPLA	54.48	61.59	89.69	138.18	13.86	35.15	0.68	19.05
FimH-DSG + 4v O-Ag in liposomal MPLA/QS21	79.59	181.73	52.82	230.35	39.64	94.34	3.84	40.78

- Functional bactericidal antibodies induced by individual O-antigens of the 4-valent composition were assessed in OPA assays with invasive clinical strains. Selection of OPA strains was based on a combination of factors: expression of both O-antigen and K-capsule under in vitro growth conditions and/or multidrug-resistance; and the ability to identify compatible baby rabbit serum complement capable of facilitating bactericidal killing in the presence of neutrophil-like HL60 cells and O-antigen specific antibodies. While *E. coli* O-antigens are considered to be the primary determinant for preventing non-specific killing by serum complement, K1, K2 and K5-capsular antigens have also been shown to play a role in resisting phagocytosis in whole blood (Sarkar S, Ulett GC, Totsika M, et al. Role of Capsule and O Antigen in the Virulence of Uropathogenic Escherichia coli. PLoS ONE 2014; 9:e94786; Burns SM, Hull SI. Loss of resistance to ingestion and phagocytic killing by O(-) and K(-) mutants of a uropathogenic Escherichia coli O75:K5 strain. Infection and immunity 1999; 67:3757-62; Buckles EL, Wang X, Lane MC, et al. Role of the K2 Capsule in Escherichia coli Urinary Tract Infection and Serum Resistance. The Journal of infectious diseases 2009; 199:10.1086/598524).

Results of OPA assays with four encapsulated O1a, O2, O6 and O25b strains and two multidrug-resistant unencapsulated O6 and O25b strains are shown in **FIG. 42** and **Table 39**.

With the exception of the O6:K- strain assay which exhibited a relatively high baseline titer of 174 for pre-vaccination sera, sera from unvaccinated mice for all other strains generated titers that were at the limit of assay detection (LOD) or 100, the starting serum dilution. Individual OPA responses to the four-valent O-antigens largely parallel the IgG binding antibody titers (FIG. 41, Table 38). For example, the significant positive impact of the liposomal MPLA/QS21 formulation on the immunogenicity of the FimH-DSG 4-valent O-antigen combination was confirmed in OPAs with encapsulated serotype O1a, O2, and O6 strains, which reported responder rates of 70% or greater after two vaccine doses. In contrast, only after a third dose did the O25b:K5 OPA strain report titers reaching this level of functional activity with the MPLA/QS21 formulation. Compared with their encapsulated O6:K1 and O25b:K5 counterparts, unencapsulated O6 and O25b strains generated higher OPA titers and responder rates, especially among non-adjuvanted groups of mice vaccinated with O-antigens alone or the combined FimH-DSG/ 4-valent O-antigens. This greater sensitivity is consistent with the absence of respective K2 or K5 capsules shown previously to confer resistance to phagocytosis (Burns SM, Hull SI. Loss of resistance to ingestion and phagocytic killing by O(-) and K(-) mutants of a uropathogenic Escherichia coli O75:K5 strain. Infection and immunity 1999; 67:3757-62; Buckles EL, Wang X, Lane MC, et al. Role of the K2 Capsule in Escherichia coli Urinary Tract Infection and Serum Resistance. The Journal of infectious diseases 2009; 199:10.1086/598524).

20

Table 39. VAC-2020-PRL-EC-1679: Influence of Adjuvant and FimH-DSG on O-antigen Specific Serum OPA GMTs

Vaccination Group	O1a:K1 OPA		O2:K1 OPA		O6:K2		O6:K-		O25b:K5		O25b:K-	
	PD2	PD3	PD2	PD3	PD2	PD3	PD2	PD3	PD2	PD3	PD2	PD3
4v O-Ag	119	348	191	1108	50	155	383	1931	68	461	1516	14260
FimH-DSG + 4v O-Ag no adjuvant	157	575	161	385	98	1256	760	5375	50	443	223	9000
FimH-DSG + 4v O-Ag in liposomal MPLA	675	979	499	1525	510	388	714	1104	91	457	1382	28204
FimH-DSG + 4v O-Ag in	6978	7714	984	8610	1064	737	1681	1961	110	606	2593	19462

liposomal MPLA/QS21												
------------------------	--	--	--	--	--	--	--	--	--	--	--	--

Conclusion

Taken together, the results of these studies support the use of mammalian-produced FimH-DSG (or derived mutant variants) in combination with O-antigen glycoconjugates and QS21 adjuvant as a vaccine composition to prevent invasive *E. coli* infections. Such a formulation may be advantageous for preventing recurrent UTI infections and sepsis in elderly patients. Results confirm the benefit of QS21 adjuvant for enhancing the functional immune response to FimH-DSG and the long O25b O-antigen, which are individually less immunogenic than analogous serotype O1a, O2, and O6 O-antigen glycoconjugates in mice.

EXAMPLE 36: Antibodies Elicited by CRM197 Conjugates of E.coli serotype O8 and O9 O-antigens show cross-protective bactericidal activity against K.pneumoniae serotype O5 and O3 invasive isolates.

This Example demonstrates that antibodies elicited by short single-end native CRM₁₉₇ polymannan conjugates of *E. coli* serotype O8 and O9 O-antigens are bactericidal and cross-protective against *Klebsiella* O5 and O3 strains that express equivalent or related O-antigens.

E. coli and *K. pneumoniae* share common polymannan O-antigens which are synthesized by enzymes encoded by highly homologous biosynthetic gene clusters. The *E. coli* O8 and O9 O-antigen polysaccharides are linear mannose homopolymers whose repeat units differ in their monosaccharide linkages and in the number of residues. Their counterparts in *K. pneumoniae* are the serotype O5 and O3 O-antigens. Biosynthesis of the *E. coli* and *K. pneumoniae* polymannan O-antigens involves a different mechanism of O-unit translocation and chain synthesis than other *E. coli* O-antigens. In this case chain elongation is regulated by the biosynthetic WbdA–WbdD complex (King JD, Berry S, et al. Proceedings of the National Academy of Sciences 2014; 111:6407-12), which is distinct from the Wzx/Wzy-dependent pathway where chain length is controlled by WzzB or FepE enzymes. As a consequence, the native polymannan O-antigens can only be produced in their short form, which requires different bioprocesses methods for purification and carrier protein conjugation than the engineered long *E. coli* O-antigens. The same mechanism and limitation applies to the predominant *K. pneumoniae* serotype O1 and O2 O-antigens, which are polygalactans comprised of galactose residues.

The structural relationship between these O-antigens and their subtypes is shown in FIG. 33. The *E. coli* O8 and *K. pneumoniae* O5 O-antigens are identical (Vinogradov E, et al. J

Biol Chem 2002; 277:25070-81). *E. coli* O9 and *K. pneumoniae* O3 O-antigens share common tetrameric O9a/O3a and pentameric O9/O3 repeat unit subtypes, while the trimeric O3b subtype is only found in *K. pneumoniae*. These subtypes can be identified serologically and genotypically (Guachalla LM, et al. Scientific Reports 2017; 7:6635). Serotype O3a loci are distinguished by a single point mutation in wbdA (C80R). An analogous point mutation in the *E. coli* O9 wbdA enzyme (C55R) converts the O9 polysaccharide into O9a (Kido N, Kobayashi H. Journal of bacteriology 2000; 182:2567-73). The O3b subtype has sufficient nucleotide divergence in the sequence of the WbdD enzyme to necessitate a separate reference sequence. The Kaptive web algorithm (Wick RR, et al. J Clin Microbiol 2018; 56), implemented into Pfizer's BigSdb whole genome sequencing (WGS) pipeline, designates O3 loci as either O3/O3a (covered by the same reference sequence) or O3b.

Materials and Methods

a. Production of *E. coli* serotype O8 and O9 CRM₁₉₇ immune sera in rabbits

Two groups of four female New Zealand White rabbits each were used for the study run at Covance. Animals received 10µg/animal of serotype O8 or O9 CRM₁₉₇ conjugate per dose with CFA/IFA as adjuvant. Native O8 and O9 O-antigens were conjugated using single-end chemistry. Each 1 mL dose of 10µg of antigen was split across two subcutaneous vaccination sites. Vaccinations were given at weeks 0, 6 and 14, with blood draws at weeks 7 and 15, which correspond to post-dose two (PD2) and post-dose three (PD3) timepoints.

b. Bacterial strains

E. coli and *K. pneumoniae* clinical isolates were obtained from the Pfizer-sponsored Antimicrobial Testing Leadership and Surveillance (ATLAS) collection which is maintained by the International Health Management Associates (IHMA) clinical lab. Strains were genotypically characterized by whole genome sequencing (WGS) using the Miseq platform (Illumina). WGS data was used to generate multi-locus-sequence type (MLST) information using established *E. coli* and *K. pneumoniae* schemes integrated into the BigDdb platform (Wirth T, et al. Molecular microbiology 2006; 60:1136-51; Jolley KA, et al. Wellcome Open Res 2018; 3:124; Diancourt L, et al. Journal of clinical microbiology 2005; 43:4178-82). Embedded in silico serotyping algorithms for *E. coli* and *K. pneumoniae* were used to predict O-antigen serotype (Wick RR, et al. J Clin Microbiol 2018; 56; Joensen KG, et al. J Clin Microbiol 2015; 53:2410-26).

Table 40. Clinical isolates used for O-antigen production or development of bactericidal assays

ID	Species	MLST ST	Serotype (subtype)	Source
EC0130	<i>E. coli</i>	162	O8	Blood
EC0423	<i>E. coli</i>	46	O9a	Blood
EC0305	<i>E. coli</i>	448	O8	Blood
KP0121	<i>K. pneumoniae</i>	279	O5	Blood
EC0611	<i>E. coli</i>	New	O9a	UTI, Kidney
KP0009	<i>K. pneumoniae</i>	37	O3b	UTI, Bladder

c. *E. coli* O8 and O9 CRM₁₉₇ conjugates

- Serotype O8 and O9a O-antigen polysaccharides were extracted and purified from strains EC0130 and EC0423, respectively (**Table 40**). The conjugation process involves selective activation of the Kdo monosaccharide present on the reducing end of the short native
- 5 *E. coli* O8 and O9 O-antigens with a disulfide amine linker. Upon unmasking of thiol functional group, it is then conjugated to bromo activated CRM₁₉₇ protein as described **Example 26** set forth herein.

d. Bactericidal assays

- Pre-frozen *E. coli* and *K. pneumoniae* stocks were prepared by growing strains in DMEM or LB media to an OD₆₀₀ of between 0.5 and 1.0 and glycerol was added to a final concentration of 20% prior to freezing. Specific assay conditions varied according to conditions optimized for each bacterial strain. Pre-titrated thawed bacteria were diluted to 1 X 10⁵ CFU/ml in OPA buffer (Hanks Balanced Salt Solution (Life Technologies) and 0.1% gelatin) and 20 µL (10³ CFU) of the bacterial suspension was opsonized with 20 µL of serially diluted sera for 30
- 15 min at RT in a tissue culture microplate. Subsequently, 10 µl of complement (Baby Rabbit Serum or IgG/IgM depleted human serum, Pel-Freez) and 20 µL of HL-60 cells (at 100-200:1 ratio) were added to each well with OPA buffer to a final volume of 100 µL. The reaction mixture was shaken for 60 min at 37°C in a 5% CO₂ incubator. In some cases, bacteria were directly combined with complement and HL60s without the pre-opsonization step and shaken for 60 min
- 20 at 37°C under 5% CO₂. After the incubation, 10 µL of each reaction was transferred into the corresponding wells of a prewetted Millipore MultiScreen HTS HV filter plate containing 100 µL water. After vacuum filtering the liquid, 100 µL of 50% bacterial growth media was applied and filtered, and the plate incubated overnight at 37°C in a sealed zip-lock bag. The next day

microcolonies were enumerated after staining with Coomassie dye using an ImmunoSpot® analyzer and ImmunoCapture software. In the case of the *E. coli* serotype O9 assay, the OPA was miniaturized for 50 µL reaction reactions volumes in the 384-well format. To establish the specificity of OPA activity, immune sera were preincubated with purified O-antigen polysaccharide prior to the opsonization step. The OPA assay included control reactions without HL60 cells or complement, to demonstrate dependence of any observed killing on these components. For the *Klebsiella* serotype O5 assay where presence of HL60s had no impact, serum bactericidal reactions were run in the absence of effector cells.

Results

a. *E. coli* and *K. pneumoniae* strain selection for bactericidal assays

Bacterial clinical strains were initially selected after confirming O-antigen expression by LPS profiling (by SDS-PAGE), and O-antigen surface accessibility by flow cytometry with O-antigen specific rabbit antisera. Next, empiric screening of serum complement was done to identify individual compatible lots across a range of concentrations that provided a suitable balance of low levels of non-specific killing combined with a high degree of susceptibility in the presence of immune sera. Additional assay optimization parameters included adjusting the ratio of HL60 effector cell to bacteria, shaker speed, presence/ absence of plate sealer and inclusion of an opsonization preincubation step.

b. *E. coli* O8 and *K. pneumoniae* O5 O-antigen immune serum cross-protection and specificity

E. coli O8 strain EC0305 and *K. pneumoniae* O5 strain KP0121 were selected for assay development. Both are blood isolates. EC0305 is resistant to cephalosporins and tetracycline while KP0121 is resistant to ampicillin. An OPA assay was developed for the *E. coli* O8 strain EC0305 with conditions including 3.0% BRC, a 1:100 bacteria to HL60 ratio, and a single step 60 min OPA incubation reaction. As the bactericidal activity of the *K. pneumoniae* O5 strain KP0121 in the presence of immune sera was found to be independent of HL60 effector cells, an SBA was developed. In this case, the SBA reaction required the use of 10% depleted human serum as source of complement. Results of bactericidal assays with these *E. coli* O8 and *K. pneumoniae* O5 strains are shown in **FIG. 34A-34B**. In the *E. coli* serotype O8 OPA, rabbit immune serum generated after two doses of O8-CRM₁₉₇ conjugate showed potent O-antigen specific killing that was blocked by preadsorption of the immune serum with free O8 O-antigen polysaccharide. Complete killing was observed at serum dilutions of less than 1:1000. Matched pre-immune sera from the same rabbit was inactive. The same rabbit sera was evaluated in the

K. pneumoniae O5 SBA, and found to be similarly bactericidal at a 1:2000 serum dilution. Killing was blocked by free O8 O-antigen and absent with the pre-immune serum. In this case a serum matrix prozone masked SBA activity at serum dilutions of less than 1:1000.

c. *E. coli* O9 and *K. pneumoniae* O3 OPA O-antigen immune serum cross-protection and specificity

E. coli O9a strain EC0611 and *K. pneumoniae* O3b strain KP0009 were selected for assay development. EC0611 is resistant to ampicillin, while KP0009 is resistant to cephalosporins, fluoroquinolones, and tetracycline. Both are UTI isolates from kidney and bladder infections, respectively. The O9a O-antigen used to generate the CRM₁₉₇ conjugate and resulting immune serum has the tetrameric polymannan repeat unit structure and is identical to the O9a EC0611 assay strain O-antigen; however, it is structurally heterologous to the *K. pneumoniae* O3b O-antigen KP0009 assay strain, which is predicted to express the shorter trimeric repeat unit based on the sequence of its *wbdD* gene (See FIG. 33). Results of OPAs with the *E. coli* O9a and *K. pneumoniae* O3b strains show that anti-*E. coli* O9a immune serum is potent against both (FIG. 35A-35B). Complete killing in the OPAs was observed at serum dilutions of less than 1:8,000 for the *E. coli* O9a strain and at less than 1:1,600 for the *Klebsiella* O3b strain. Specificity was demonstrated by the lack of activity upon preadsorption of serum with free O9a O-antigen and with the matched-pre-immune serum.

Conclusion

E. coli serotype O8 and O9 polymannan CRM₁₉₇ conjugates elicit functional antibodies that are capable of killing not only homologous *E. coli* clinical strains but also *Klebsiella* serotype O5 and O3 strains in bactericidal assays. Results confirm that these conjugates elicit antibodies that are cross-protective against isolates of both species expressing structurally related polymannan O-antigens.

The following clauses describe additional embodiments of the invention:

C1. A composition comprising a polypeptide derived from FimH or a fragment thereof; and a saccharide comprising a structure selected from any one of Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2; K13; K15 and Formula O6:K54), Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18 (e.g., Formula

O18A, Formula O18ac, Formula O18A1, Formula O18B, and Formula O18B1), Formula O19, Formula O20, Formula O21, Formula O22, Formula O23 (e.g., Formula O23A), Formula O24, Formula O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73 (e.g., Formula O73 (strain 73-1)), Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein *n* is an integer from 1 to 100.

C2. The composition according to clause C1, wherein the saccharide comprises a structure selected from Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2;

- K13; K15 and Formula O6:K54), Formula O7, Formula O10, Formula O16, Formula O17,
 Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and
 Formula O18B1), Formula O21, Formula O23 (e.g., Formula O23A), Formula O24, Formula
 O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O28, Formula O44,
 5 Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O55, Formula O56,
 Formula O58, Formula O64, Formula O69, Formula O73 (e.g., Formula O73 (strain 73-1)),
 Formula O75, Formula O77, Formula O78, Formula O86, Formula O88, Formula O90,
 Formula O98, Formula O104, Formula O111, Formula O113, Formula O114, Formula O119,
 Formula O121, Formula O124, Formula O125, Formula O126, Formula O127, Formula
 10 O128, Formula O136, Formula O138, Formula O141, Formula O142, Formula O143,
 Formula O147, Formula O149, Formula O152, Formula O157, Formula O158, Formula
 O159, Formula O164, Formula O173, Formula 62D₁, Formula O22, Formula O35, Formula
 O65, Formula O66, Formula O83, Formula O91, Formula O105, Formula O116, Formula
 O117, Formula O139, Formula O153, Formula O167, and Formula O172, wherein *n* is an
 15 integer from 20 to 100.
- C3. The composition according to clause C2, wherein the saccharide comprises a structure
 selected from Formula O1 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula
 O2, Formula O3, Formula O4 (e.g., Formula O4:K52 and Formula O4:K6), Formula O5
 (e.g., Formula O5ab and Formula O5ac (strain 180/C3)), Formula O6 (e.g., Formula O6:K2;
 20 K13; K15 and Formula O6:K54), Formula O7, Formula O10, Formula O16, Formula O17,
 Formula O18 (e.g., Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, and
 Formula O18B1), Formula O21, Formula O23 (e.g., Formula O23A), Formula O24, Formula
 O25 (e.g., Formula O25a and Formula O25b), Formula O26, Formula O28, Formula O44,
 Formula O45 (e.g., Formula O45 and Formula O45rel), Formula O55, Formula O56,
 25 Formula O58, Formula O64, Formula O69, Formula O73 (e.g., Formula O73 (strain 73-1)),
 Formula O75, Formula O77, Formula O78, Formula O86, Formula O88, Formula O90,
 Formula O98, Formula O104, Formula O111, Formula O113, Formula O114, Formula O119,
 Formula O121, Formula O124, Formula O125, Formula O126, Formula O127, Formula
 O128, Formula O136, Formula O138, Formula O141, Formula O142, Formula O143,
 30 Formula O147, Formula O149, Formula O152, Formula O157, Formula O158, Formula
 O159, Formula O164, Formula O173, and Formula 62D₁, wherein *n* is an integer from 20 to
 100.
- C4. The composition according to clause C2, comprising a structure selected from Formula O1
 (e.g., Formula O1A, Formula O1B, and Formula O1C), Formula O2, Formula O6 (e.g.,
 35 Formula O6:K2; K13; K15 and Formula O6:K54), Formula O15, Formula O16, Formula
 O21, Formula O25 (e.g., Formula O25a and Formula O25b), and Formula O75.

- C5. The composition according to clause C2, comprising a structure selected from Formula O4, Formula O11, Formula O21, and Formula O75.
- C6. The composition according to clause C1, wherein the saccharide does not comprise a structure selected from Formula O8, Formula O9a, Formula O9, Formula O20ab, Formula O20ac, Formula O52, Formula O97, and Formula O101.
- C7. The composition according to clause C1, wherein the saccharide does not comprise a structure selected from Formula O12.
- C8. The composition according to clause C4, wherein the saccharide is produced by expressing a wzz family protein in a Gram-negative bacterium to generate said saccharide.
- C9. The composition according to clause C8, wherein the wzz family protein is selected from the group consisting of wzzB, wzz, wzz_{SF}, wzz_{ST}, fepE, wzz_{fepE}, wzz1 and wzz2.
- C10. The composition according to clause C8, wherein the wzz family protein is wzzB.
- C11. The composition according to clause C8, wherein the wzz family protein is fepE.
- C12. The composition according to clause C8, wherein the wzz family protein is wzzB and fepE.
- C13. The composition according to clause C8, wherein the wzz family protein is derived from *Salmonella enterica*.
- C14. The composition according to clause C8, wherein the wzz family protein comprises a sequence selected from any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, and SEQ ID NO: 39.
- C15. The composition according to clause C8, wherein the wzz family protein comprises a sequence having at least 90% sequence identity to any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34.
- C16. The composition according to clause C8, wherein the wzz family protein comprises a sequence selected from any one of SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, and SEQ ID NO: 39.
- C17. The composition according to clause C1, wherein the saccharide is synthetically synthesized.
- C18. The composition according to any one of clauses C1 to C17, wherein the saccharide further comprises an *E. coli* R1 moiety.
- C19. The composition according to any one of clauses C1 to C17, wherein the saccharide further comprises an *E. coli* R2 moiety.
- C20. The composition according to any one of clauses C1 to C17, wherein the saccharide further comprises an *E. coli* R3 moiety.

- C21. The composition according to any one of clauses C1 to C17, wherein the saccharide further comprises an *E. coli* R4 moiety.
- C22. The composition according to any one of clauses C1 to C17, wherein the saccharide further comprises an *E. coli* K-12 moiety.
- 5 C23. The composition according to any one of clauses C1 to C22, wherein the saccharide further comprises a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- C24. The composition according to any one of clauses C1 to C17, wherein the saccharide does not further comprise an *E. coli* R1 moiety.
- C25. The composition according to any one of clauses C1 to C17, wherein the saccharide
10 does not further comprise an *E. coli* R2 moiety.
- C26. The composition according to any one of clauses C1 to C17, wherein the saccharide does not further comprise an *E. coli* R3 moiety.
- C27. The composition according to any one of clauses C1 to C17, wherein the saccharide does not further comprise an *E. coli* R4 moiety.
- 15 C28. The composition according to any one of clauses C1 to C17, wherein the saccharide does not further comprise an *E. coli* K-12 moiety.
- C29. The composition according to any one of clauses C1 to C22, wherein the saccharide does not further comprise a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- C30. The composition according to any one of clauses C1 to C23, wherein the saccharide
20 does not comprise a Lipid A.
- C31. The composition according to any one of clauses C1 to C30, wherein the polysaccharide has a molecular weight of between 10 kDa and 2,000 kDa, or between 50 kDa and 2,000 kDa.
- C32. The composition according to any one of clauses C1 to C31, wherein the saccharide
25 has an average molecular weight of 20-40 kDa.
- C33. The composition according to any one of clauses C1 to C32, wherein the saccharide has an average molecular weight of 40,000 to 60,000 kDa.
- C34. The composition according to any one of clauses C1 to C33, wherein n is an integer 31 to 90.
- 30 C35. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a saccharide covalently bound a carrier protein, wherein the saccharide is derived from *E. coli*.
- C36. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a saccharide according to any one of clause C1 to clause C34,
35 covalently bound a carrier protein.

- C37. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate according to any one of clause C35 to clause C36, wherein the carrier protein is selected from any one of poly(L-lysine), CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).
- 10 C38. The composition according to any one of clause C35 to clause C37, wherein the carrier protein is CRM₁₉₇.
- C39. The composition according to any one of clause C35 to clause C37, wherein the carrier protein is tetanus toxoid (TT).
- C40. The composition according to any one of clause C35 to clause C37, wherein the carrier protein is poly(L-lysine).
- 15 C41. The composition according to any one of clause C35 to clause C39, wherein the conjugate is prepared by reductive amination.
- C42. The composition according to any one of clause C35 to clause C39, wherein the conjugate is prepared by CDAP chemistry.
- 20 C43. The composition according to any one of clause C35 to clause C39, wherein the conjugate is a single-end linked conjugated saccharide.
- C44. The composition according to any one of clause C35 to clause C39, wherein the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer.
- 25 C45. The composition according to clause C44, wherein the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer, wherein the saccharide is covalently linked to the eTEC spacer through a carbamate linkage, and wherein the carrier protein is covalently linked to the eTEC spacer through an amide linkage.
- 30 C46. The composition according to any one of clause C44 to clause C45, wherein the CRM₁₉₇ comprises 2 to 20, or 4 to 16, lysine residues covalently linked to the polysaccharide through an eTEC spacer.
- C47. The composition according to any one of clause C35 to clause C46, wherein the saccharide:carrier protein ratio (w/w) is between 0.2 and 4.
- 35 C48. The composition according to any one of clause C35 to clause C46, wherein the ratio of saccharide to protein is at least 0.5 and at most 2.

- C49. The composition according to any one of clause C35 to clause C46, wherein the ratio of saccharide to protein is between 0.4 and 1.7
- C50. The composition according to any one of clause C43 to clause C49, wherein the saccharide is conjugated to the carrier protein through a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) residue.
- 5 C51. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a saccharide covalently bound a carrier protein, wherein the saccharide comprises a structure selected from Formula O8, Formula O9a, Formula O9, Formula O20ab, Formula O20ac, Formula O52, Formula O97, and Formula O101, wherein
- 10 n is an integer from 1 to 10.
- C52. A composition comprising a polypeptide derived from FimH or fragment thereof; and a saccharide according to any one of clause C1 to clause C34, and a pharmaceutically acceptable diluent.
- C53. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate according to any one of clause C35 to clause C51, and a pharmaceutically acceptable diluent.
- 15 C54. The composition according to clause C53, comprising at most about 25% free saccharide as compared to the total amount of saccharide in the composition.
- C55. The composition according to any one of clause C52 to clause C53, further comprising
- 20 an adjuvant.
- C56. The composition according to any one of clause C52 to clause C53, further comprising aluminum.
- C57. The composition according to any one of clause C52 to clause C53, further comprising QS-21.
- 25 C58. The composition according to any one of clause C52 to clause C53, further comprising a CpG oligonucleotide.
- C59. The composition according to any one of clause C52 to clause C53, wherein the composition does not include an adjuvant.
- C60. A composition comprising a polypeptide derived from FimH or fragment thereof; and a saccharide derived from *E. coli*, conjugated to a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer, wherein the polysaccharide is covalently linked to the eTEC spacer through a carbamate linkage, and wherein the carrier protein is covalently linked to the eTEC spacer through an amide linkage.
- 30 C61. The composition according to clause C60, wherein the saccharide is an O-antigen
- 35 derived from *E. coli*.

- C62. The composition according to clause C60, further comprising a pharmaceutically acceptable excipient, carrier or diluent.
- C63. The composition according to clause C60, wherein the saccharide is an O-antigen derived from *E. coli*.
- 5 C64. A composition comprising a polypeptide derived from FimH or fragment thereof; and a saccharide according to any one of clause C1 to clause C17, conjugated to a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer, wherein the polysaccharide is covalently linked to the eTEC spacer through a carbamate linkage, and wherein the carrier protein is covalently linked to the eTEC spacer through an amide linkage.
- 10 C65. A composition comprising a polypeptide derived from FimH or fragment thereof; and (i) a conjugate of an *E. coli* O25B antigen covalently coupled to a carrier protein, (ii) a conjugate of an *E. coli* O1A antigen covalently coupled to a carrier protein, (iii) a conjugate of an *E. coli* O2 antigen covalently coupled to a carrier protein, and (iv) a conjugate of an O6 antigen covalently coupled to a carrier protein, wherein the *E. coli* O25B antigen comprises the
- 15 structure of Formula O25B, wherein n is an integer greater than 30.
- C66. The composition of clause C65, wherein the carrier protein is selected from any one of poly(L-lysine), CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose
- 20 binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).
- C67. A composition comprising a polypeptide derived from FimH or fragment thereof; and (i) a
- 25 conjugate of an *E. coli* O25B antigen covalently coupled to a carrier protein, (ii) a conjugate of an *E. coli* O4 antigen covalently coupled to a carrier protein, (iii) a conjugate of an *E. coli* O11 antigen covalently coupled to a carrier protein, and (iv) a conjugate of an O21 antigen covalently coupled to a carrier protein, wherein the *E. coli* O25B antigen comprises the structure of Formula O75, wherein n is an integer greater than 30.
- 30 C68. The composition of clause C67, wherein the carrier protein is selected from any one of poly(L-lysine), CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose
- 35 binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and

detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).

- 5 C69. A method of making a composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a saccharide conjugated to a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer, comprising the steps of a) reacting a saccharide with 1,1'-carbonyl-di-(1,2,4-triazole) (CDT) or 1,1'-carbonyldiimidazole (CDI), in an organic solvent to produce an activated saccharide; b) reacting the activated saccharide with cystamine or cysteamine or a salt thereof, to produce a thiolated saccharide; c) reacting the thiolated saccharide with a reducing agent to produce an
- 10 activated thiolated saccharide comprising one or more free sulfhydryl residues; d) reacting the activated thiolated saccharide with an activated carrier protein comprising one or more α -haloacetamide groups, to produce a thiolated saccharide-carrier protein conjugate; and e) reacting the thiolated saccharide-carrier protein conjugate with (i) a first capping reagent capable of capping unconjugated α -haloacetamide groups of the activated carrier protein; and/or (ii) a second capping reagent capable of capping unconjugated free sulfhydryl
- 15 residues; whereby an eTEC linked glycoconjugate is produced, wherein the saccharide is derived from *E. coli*; further comprising expressing a polynucleotide encoding a polypeptide derived from FimH or fragment thereof in a recombinant mammalian cell, and isolating said polypeptide or fragment thereof.
- 20 C70. The method according to clause C69, comprising making the composition according to any one of clause C1 to clause C34.
- C71. The method according to any of one clause C69 to clause C70, wherein the capping step e) comprises reacting the thiolated saccharide-carrier protein conjugate with (i) N-acetyl-L-cysteine as a first capping reagent, and/or (ii) iodoacetamide as a second capping
- 25 reagent.
- C72. The method according to any of one clause C69 to clause C71, further comprising a step of compounding the saccharide by reaction with triazole or imidazole to provide a compounded saccharide, wherein the compounded saccharide is shell frozen, lyophilized and reconstituted in an organic solvent prior to step a).
- 30 C73. The method according to any of one clause C69 to clause C72, further comprising purification of the thiolated polysaccharide produced in step c), wherein the purification step comprises diafiltration.
- C74. The method according to any of one clause C69 to clause C73, wherein the method further comprises purification of the eTEC linked glycoconjugate by diafiltration.
- 35 C75. The method according to any of one clause C69 to clause C74, wherein the organic solvent in step a) is a polar aprotic solvent selected from any one of dimethyl sulfoxide

(DMSO), dimethylformamide (DMF), dimethylacetamide (DMA), N-methyl-2-pyrrolidone (NMP), acetonitrile, 1,3-Dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone (DMPU) and hexamethylphosphoramide (HMPA), or a mixture thereof.

- 5 C76. A medium comprising KH_2PO_4 , K_2HPO_4 , $(\text{NH}_4)_2\text{SO}_4$, sodium citrate, Na_2SO_4 , aspartic acid, glucose, MgSO_4 , $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, H_3BO_3 , $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, ZnCl_2 and $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$.
- C77. The medium according to clause C76, wherein the medium is used for culturing *E. coli*.
- 10 C78. A method for producing a saccharide according to any one of clause C1 to clause C34, comprising culturing a recombinant *E. coli* in a medium; producing said saccharide by culturing said cell in said medium; whereby said cell produces said saccharide.
- C79. The method according to clause C78, wherein the medium comprises an element selected from any one of KH_2PO_4 , K_2HPO_4 , $(\text{NH}_4)_2\text{SO}_4$, sodium citrate, Na_2SO_4 , aspartic acid, glucose, MgSO_4 , $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, H_3BO_3 , $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, ZnCl_2 and $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$.
- 15 C80. The method according to clause C78, wherein the medium comprises soy hydrolysate.
- C81. The method according to clause C78, wherein the medium comprises yeast extract.
- C82. The method according to clause C78, wherein the medium does not further comprise soy hydrolysate and yeast extract.
- C83. The method according to clause C78, wherein the *E. coli* cell comprises a heterologous
- 20 wzz family protein selected from any one of wzzB, wzz, wzz_{SF}, wzz_{ST}, fepE, wzz_{fepE}, wzz1 and wzz2.
- C84. The method according to clause C78, wherein the *E. coli* cell comprises a *Salmonella enterica* wzz family protein selected from any one of wzzB, wzz, wzz_{SF}, wzz_{ST}, fepE, wzz_{fepE}, wzz1 and wzz2.
- 25 C85. The method according to clause C84, wherein the wzz family protein comprises a sequence selected from any one of SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, and SEQ ID NO: 19.
- C86. The method according to clause C78, wherein the culturing produces a yield of > 120
- 30 OD₆₀₀/mL.
- C87. The method according to clause C78, further comprising purifying the saccharide.
- C88. The method according to clause C78, wherein the purifying step comprises any one of the following: dialysis, concentration operations, diafiltration operations, tangential flow filtration, precipitation, elution, centrifugation, precipitation, ultra-filtration, depth filtration,
- 35 and column chromatography (ion exchange chromatography, multimodal ion exchange chromatography, DEAE, and hydrophobic interaction chromatography).

- C89. A method for inducing an immune response in a mammal comprising administering to the subject a composition according to any one of clause C1 to clause C68.
- C90. The method according to clause C89, wherein the immune response comprises induction of an anti-*E. coli* O-specific polysaccharide serum antibody.
- 5 C91. The method according to clause C89, wherein the immune response comprises induction of an anti-*E. coli* IgG antibody.
- C92. The method according to clause C89, wherein the immune response comprises induction of bactericidal activity against *E. coli*.
- C93. The method according to clause C89, wherein the immune response comprises
10 induction of opsonophagocytic antibodies against *E. coli*.
- C94. The method according to clause C89, wherein the immune response comprises a geometric mean titer (GMT) level of at least 1,000 to 200,000 after initial dosing.
- C95. The method according to clause C89, wherein the composition comprises a saccharide comprising the Formula O25, wherein n is an integer 40 to 100, wherein the immune
15 response comprises a geometric mean titer (GMT) level of at least 1,000 to 200,000 after initial dosing.
- C96. The method according to clause C89, wherein the mammal is at risk of any one of the conditions selected from urinary tract infection, cholecystitis, cholangitis, diarrhea, hemolytic uremic syndrome, neonatal meningitis, urosepsis, intra-abdominal infection, meningitis,
20 complicated pneumonia, wound infection, post-prostate biopsy-related infection, neonatal/infant sepsis, neutropenic fever, and other blood stream infection; pneumonia, bacteremia, and sepsis.
- C97. The method according to clause C89, wherein the mammal is has any one of the conditions selected from urinary tract infection, cholecystitis, cholangitis, diarrhea, hemolytic uremic syndrome, neonatal meningitis, urosepsis, intra-abdominal infection, meningitis,
25 complicated pneumonia, wound infection, post-prostate biopsy-related infection, neonatal/infant sepsis, neutropenic fever, and other blood stream infection; pneumonia, bacteremia, and sepsis.
- C98. A method for (i) inducing an immune response in a subject against extra-intestinal pathogenic *Escherichia coli*, (ii) inducing an immune response in a subject against extra-intestinal pathogenic *Escherichia coli*, or (iii) inducing the production of opsonophagocytic antibodies in a subject that are specific to extra-intestinal pathogenic *Escherichia coli*,
30 wherein the method comprises administering to the subject an effective amount of the composition according to any one of clause C1 to clause C68.
- C99. The method of clause C98, wherein the subject is at risk of developing a urinary tract infection.
35

- C100. The method of clause C98, wherein the subject is at risk of developing bacteremia.
- C101. The method of clause C98, wherein the subject is at risk of developing sepsis.
- C102. A composition comprising a polypeptide derived from FimH or fragment thereof; and a (i)
5 a conjugate of an *E. coli* O25B antigen covalently coupled to a carrier protein, (ii) a
conjugate of an *E. coli* O1A antigen covalently coupled to a carrier protein, (iii) a conjugate
of an *E. coli* O2 antigen covalently coupled to a carrier protein, and (iv) a conjugate of an O6
antigen covalently coupled to a carrier protein, wherein the *E. coli* O25B antigen comprises
the structure of Formula O25B, wherein n is an integer greater than 30.
- C103. The composition of clause C102, wherein the carrier protein is selected from the group
10 consisting of poly(L-lysine), detoxified Exotoxin A of *P. aeruginosa* (EPA), CRM197, maltose
binding protein (MBP), Diphtheria toxoid, Tetanus toxoid, detoxified hemolysin A of *S.*
aureus, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), cholera toxin,
detoxified variants of cholera toxin, *Streptococcus pneumoniae* Pneumolysin and detoxified
variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a
15 peptidase (SCP).
- C104. A method for (i) inducing an immune response in a subject against extra-intestinal
pathogenic *Escherichia coli*, (ii) inducing an immune response in a subject against extra-
intestinal pathogenic *Escherichia coli*, or (iii) inducing the production of opsonophagocytic
antibodies in a subject that are specific to extra-intestinal pathogenic *Escherichia coli*,
20 wherein the method comprises administering to the subject an effective amount of the
composition of clause C1.
- C105. The method of clause C104, wherein the subject is at risk of developing a urinary tract
infection.
- C106. The method of clause C104, wherein the subject is at risk of developing bacteremia.
- 25 C107. The method of clause C104, wherein the subject is at risk of developing sepsis.
- C108. A composition comprising a polypeptide derived from FimH or fragment thereof; and a
saccharide comprising an increase of at least 5 repeating units, compared to the
corresponding wild-type O-polysaccharide of an *E. coli*.
- C109. The composition according to clause C108, wherein the saccharide comprises Formula
30 O25a and the *E. coli* is an *E. coli* serotype O25a.
- C110. The composition according to clause C108, wherein the saccharide comprises Formula
O25b and the *E. coli* is an *E. coli* serotype O25b.
- C111. The composition according to clause C108, wherein the saccharide comprises Formula
O2 and the *E. coli* is an *E. coli* serotype O2.
- 35 C112. The composition according to clause C108, wherein the saccharide comprises Formula
O6 and the *E. coli* is an *E. coli* serotype O6.

- C113. The composition according to clause C108, wherein the saccharide comprises Formula O1 and the *E. coli* is an *E. coli* serotype O1.
- C114. The composition according to clause C108, wherein the saccharide comprises Formula O17 and the *E. coli* is an *E. coli* serotype O17.
- 5 C115. The composition according to clause C108, wherein the saccharide comprises a structure selected from: Formula O1, Formula O2, Formula O3, Formula O4, Formula O5, Formula O6, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O24,
- 10 Formula O25, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55,
- 15 Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87,
- 20 Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula
- 25 O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, O145, Formula O146,
- 30 Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula
- 35 O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179,

Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 5 to 1000.

- C116. The composition according to clause C108, wherein the *E. coli* is *E. coli* serotype selected from the group consisting of: O1, O2, O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18, O19, O20, O21, O22, O23, O24, O25, O25b, O26, O27, O28, O29, O30, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45, O46, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, O63, O64, O65, O66, O68, O69, O70, O71, O73, O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O95, O96, O97, O98, O99, O100, O101, O102, O103, O104, O105, O106, O107, O108, O109, O110, O111, O112, O113, O114, O115, O116, O117, O118, O119, O120, O121, O123, O124, O125, O126, O127, O128, O129, O130, O131, O132, O133, O134, O135, O136, O137, O138, O139, O140, O141, O142, O143, O144, O145, O146, O147, O148, O149, O150, O151, O152, O153, O154, O155, O156, O157, O158, O159, O160, O161, O162, O163, O164, O165, O166, O167, O168, O169, O170, O171, O172, O173, O174, O175, O176, O177, O178, O179, O180, O181, O182, O183, O184, O185, O186, and O187.
- C117. The composition according to clause C108, wherein the saccharide is produced by increasing repeating units of O-polysaccharides produced by a Gram-negative bacterium in culture comprising overexpressing wzz family proteins in a Gram-negative bacterium to generate said saccharide.
- C118. The composition according to clause C117, wherein the overexpressed wzz family protein is selected from the group consisting of wzzB, wzz, wzz_{sf}, wzz_{st}, fepE, wzz_{fepE}, wzz1 and wzz2.
- C119. The composition according to clause C117, wherein the overexpressed wzz family protein is wzzB.
- C120. The composition according to clause C117, wherein the overexpressed wzz family protein is fepE.
- C121. The composition according to clause C117, wherein the overexpressed wzz family protein is wzzB and fepE.
- C122. The composition according to clause C108, wherein the saccharide is synthetically synthesized.
- C123. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a saccharide according to clause C108, covalently bound to a carrier protein.
- C124. The composition according to clause C123, wherein the carrier protein is CRM₁₉₇.

- C125. The composition according to clause C123, wherein the saccharide comprises a structure selected from: Formula O1, Formula O2, Formula O3, Formula O4, Formula O5, Formula O6, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O24, Formula O25, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 5 to 1000.
- C126. The composition according to clause C123, wherein said saccharide comprises an increase of at least 5 repeating units, compared to the corresponding wild-type O-polysaccharide.

- C127. The composition according to clause C1, further comprising a pharmaceutically acceptable diluent.
- C128. The composition according to clause C127, further comprising an adjuvant.
- C129. The composition according to clause C127, further comprising aluminum.
- 5 C130. The composition according to clause C127, further comprising QS-21.
- C131. The composition according to clause C127, wherein the composition does not include an adjuvant.
- C132. A method for inducing an immune response in a subject comprising administering to the subject a composition according to clause C127.
- 10 C133. The composition according to clause C123, further comprising a pharmaceutically acceptable diluent.
- C134. A method for inducing an immune response in a subject comprising administering to the subject a composition according to clause C133.
- C135. The method according to clauses C132 or C134, wherein the immune response
- 15 comprises induction of an anti-*E. coli* O-specific polysaccharide serum antibody.
- C136. The method according to clause C135, wherein the anti-*E. coli* O-specific polysaccharide serum antibody is an IgG antibody.
- C137. The method according to clause C135, wherein the anti-*E. coli* O-specific polysaccharide serum antibody is an IgG antibody has bactericidal activity against *E. coli*.
- 20 C138. An immunogenic composition comprising a polypeptide derived from FimH or fragment thereof; and a saccharide derived from *E. coli*, conjugated to a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer, wherein the polysaccharide is covalently linked to the eTEC spacer through a carbamate linkage, and wherein the carrier protein is covalently linked to the eTEC spacer through an amide linkage.
- 25 C139. The immunogenic composition according to clause C138, further comprising a pharmaceutically acceptable excipient, carrier or diluent.
- C140. The immunogenic composition according to clause C138, wherein the saccharide is an O-antigen derived from *E. coli*.
- C141. The immunogenic composition according to clause C138, wherein the saccharide
- 30 comprises a structure selected from: Formula O1, Formula O2, Formula O3, Formula O4, Formula O5, Formula O6, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O24, Formula O25, Formula O25b, Formula O26, Formula O27, Formula O28,
- 35 Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41,

Formula O42, Formula O43, Formula O44, Formula O45, Formula O46, Formula O48,
 Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54,
 Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60,
 Formula O61, Formula O62, Formula O63, Formula O64, Formula O65, Formula O66,
 5 Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O74,
 Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80,
 Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86,
 Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92,
 Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99,
 10 Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula
 O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110,
 Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula
 O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121,
 Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula
 15 Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133,
 Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula
 O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, O145,
 Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula
 O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156,
 20 Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula
 O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167,
 Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula
 O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178,
 Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula
 25 O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 5 to
 1000.

C142. The immunogenic composition according to clause C138, wherein the saccharide has a
 degree of O-acetylation between 75-100%.

C143. The immunogenic composition according to clause C138, wherein the carrier protein is
 30 CRM197.

C144. The immunogenic composition according to clause C143, wherein the CRM197
 comprises 2 to 20 lysine residues covalently linked to the polysaccharide through an eTEC
 spacer.

C145. The immunogenic composition according to clause C143, wherein the CRM197
 35 comprises 4 to 16 lysine residues covalently linked to the polysaccharide through an eTEC
 spacer.

- C146. The immunogenic composition according to clause C138, further comprising an additional antigen.
- C147. The immunogenic composition according to clause C138, further comprising an adjuvant.
- 5 C148. The immunogenic composition according to clause C147, wherein the adjuvant is an aluminum-based adjuvant selected from the group consisting of aluminum phosphate, aluminum sulfate and aluminum hydroxide.
- C149. The immunogenic composition according to clause C138, wherein the composition does not comprise an adjuvant.
- 10 C150. An immunogenic composition comprising a polypeptide derived from FimH or fragment thereof; and a glycoconjugate comprising a saccharide derived from *E. coli* conjugated to a carrier protein, wherein the glycoconjugate is prepared using reductive amination.
- C151. The immunogenic composition according to clause C150, further comprising a pharmaceutically acceptable excipient, carrier or diluent.
- 15 C152. The immunogenic composition according to clause C150, wherein the saccharide is an O-antigen derived from *E. coli*.
- C153. The immunogenic composition according to clause C150, wherein the saccharide comprises a structure selected from: Formula O1, Formula O2, Formula O3, Formula O4, Formula O5, Formula O6, Formula O7, Formula O8, Formula O9, Formula O10, Formula
- 20 O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O24, Formula O25, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O46, Formula O48,
- 25 Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80,
- 30 Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula
- 35 O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula

- O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187, wherein n is an integer from 5 to 1000.
- 15 C154. The immunogenic composition according to clause C150, wherein the saccharide has a degree of O-acetylation between 75-100%.
- C155. The immunogenic composition according to clause C150, wherein the carrier protein is CRM197.
- C156. The immunogenic composition according to clause C150, further comprising an additional antigen.
- 20 C157. The immunogenic composition according to clause C150, further comprising an adjuvant.
- C158. The immunogenic composition according to clause C157, wherein the adjuvant is an aluminum-based adjuvant selected from the group consisting of aluminum phosphate, aluminum sulfate and aluminum hydroxide.
- 25 C159. The immunogenic composition according to clause C150, wherein the composition does not comprise an adjuvant.
- C160. A method for inducing an immune response in a subject comprising administering to the subject a composition according to any one of clauses C138-C159.
- 30 C161. The method according to clause C160, wherein the immune response comprises induction of an anti-*E. coli* O-specific polysaccharide serum antibody.
- C162. The method according to clause C135, wherein the anti-*E. coli* O-specific polysaccharide serum antibody is an IgG antibody.
- C163. The method according to clause C135, wherein the anti-*E. coli* O-specific polysaccharide serum antibody is an IgG antibody has bactericidal activity against *E. coli*.
- 35

C164. A composition comprising a polypeptide derived from FimH or fragment thereof; and a saccharide comprising a structure selected from any one of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula

O184, Formula O185, Formula O186, Formula O187, wherein n is greater than the number of repeat units in the corresponding wild-type *E. coli* polysaccharide.

C165. The composition according to clause C164, wherein n is an integer from 31 to 100.

C166. The composition according to clause C164, wherein the saccharide comprises a
5 structure according to any one of Formula O1A, Formula O1B, and Formula O1C, Formula O2, Formula O6, and Formula O25B.

C167. The composition according to clause C164, wherein the saccharide is produced in a recombinant host cell that expresses a wzz family protein having at least 90% sequence identity to any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33,
10 SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, and SEQ ID NO: 39.

C168. The composition according to clause C167, wherein the protein comprises any one of SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34.

C169. The saccharide according to clause C164, wherein the saccharide is synthetically
15 synthesized.

C170. A composition comprising a polypeptide derived from FimH or fragment thereof; and a conjugate comprising a carrier protein covalently bound to a saccharide, said saccharide comprising a structure selected from any one of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6,
20 Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34,
25 Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81,
30 Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93,

- Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, or Formula O187, wherein n is an integer from 1 to 100.
- C171. The composition according to clause C170, wherein the saccharide comprises any one of the following Formula O25b, Formula O1A, Formula O2, and Formula O6.
- C172. The composition according to clause C170, wherein the saccharide further comprises any one of an *E. coli* R1 moiety, *E. coli* R2 moiety, *E. coli* R3 moiety, *E. coli* R4 moiety, and *E. coli* K-12 moiety.
- C173. The composition according to clause C170, wherein the saccharide does not further comprise any one of an *E. coli* R1 moiety, *E. coli* R2 moiety, *E. coli* R3 moiety, *E. coli* R4 moiety, and *E. coli* K-12 moiety. The composition according to clause C170, wherein the saccharide does not further comprise an *E. coli* R2 moiety.
- C174. The composition according to clause C170, wherein the saccharide further comprises a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- C175. The composition according to clause C170, wherein the carrier protein is selected from any one of CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified

variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).

C176. The composition according to clause C170, wherein the carrier protein is CRM₁₉₇.

C177. The composition according to clause C170, wherein the carrier protein is tetanus toxoid.

5 C178. The composition according to clause C170, wherein the ratio of saccharide to protein is at least 0.5 to at most 2.

C179. The composition according to clause C170, wherein the conjugate is prepared via reductive amination.

10 C180. The composition according to clause C170, wherein the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer.

C181. The composition according to clause C170, wherein the saccharide is a single-end linked conjugated saccharide.

C182. The composition according to clause C174, wherein the saccharide is conjugated to the carrier protein through a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) residue.

15 C183. The composition according to clause C170, wherein the conjugate is prepared via CDAP chemistry.

C184. A composition comprising a polypeptide derived from FimH or fragment thereof; and (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein *n* is an integer from 31 to 90, (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein *n* is an integer from 31 to 90, (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein *n* is an integer from 31 to 90, and (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein *n* is an integer from 31 to 90.

25 C185. The composition according to clause C184, further comprising a conjugate comprising a carrier protein covalently bound to a saccharide comprising a structure selected from any one of the following: Formula O15, Formula O16, Formula O17, Formula O18 and Formula O75, wherein *n* is an integer from 31 to 90.

30 C186. The composition according to clause C184, comprising at most 25% free saccharide as compared to the total amount of saccharide in the composition.

C187. A method of eliciting an immune response against *Escherichia coli* in a mammal, comprising administering to the mammal an effective amount of the composition according to any one of clauses C184 to C186.

35 C188. The method according to clause C187, wherein the immune response comprises opsonophagocytic antibodies against *E. coli*.

- C189. The method according to clause C187, wherein the immune response protects the mammal from an *E. coli* infection.
- C190. A mammalian cell comprising (a) a first gene of interest encoding a polypeptide derived from *E. coli* or a fragment thereof, wherein the gene is integrated between at least two recombination target sites (RTS).
- C191. The embodiment of clause C190, wherein the two RTS are chromosomally-integrated within the NL1 locus or the NL2 locus.
- C192. The embodiment of clause C190, wherein the first gene of interest further comprises a reporter gene, a gene encoding a difficult to express protein, an ancillary gene or a combination thereof.
- C193. The embodiment of clause C190, further comprising a second gene of interest that is integrated within a second chromosomal locus distinct from the locus of (a), wherein the second gene of interest comprises a reporter gene, a gene encoding a difficult to express protein, an ancillary gene or a combination thereof.
- C194. A recombinant mammalian cell, comprising a polynucleotide encoding a polypeptide derived from *E. coli* or a fragment thereof.
- C195. The recombinant cell according to C194, wherein the polypeptide is derived from *E. coli* fimbrial H (FimH).
- C196. The recombinant cell according to C195, wherein the polypeptide comprises a phenylalanine residue at the N-terminus of the polypeptide.
- C197. The recombinant cell according to C195, wherein the polypeptide comprises a phenylalanine residue within the first 20 residue positions of the N-terminus.
- C198. The recombinant cell according to C195, wherein the polypeptide comprises a phenylalanine residue at position 1 of the polypeptide.
- C199. The recombinant cell according to C198, wherein the polypeptide does not comprise a glycine residue immediately before the phenylalanine residue at position 1 of the polypeptide.
- C200. The recombinant cell according to C195, wherein the polypeptide does not comprise an N-glycosylation site at position 7 of the polypeptide.
- C201. The recombinant cell according to C199, wherein the polypeptide does not comprise an Asn residue at position 7 of the polypeptide.
- C202. The recombinant cell according to C201, wherein the polypeptide comprises a residue selected from the group consisting of Ser, Asp, Thr, and Gln at position 7.
- C203. The recombinant cell according to C198, wherein the polypeptide does not comprise an N-glycosylation site at position 70 of the polypeptide.

- C204. The recombinant cell according to C203, wherein the polypeptide does not comprise an Asn residue at position 70 of the polypeptide.
- C205. The recombinant cell according to C203, wherein the polypeptide does not comprise a Ser residue at position 70 of the polypeptide.
- 5 C206. The recombinant cell according to C194, wherein the polypeptide comprises a residue substitution selected from the group consisting of Ser, Asp, Thr, and Gln at an N-glycosylation site of the polypeptide.
- C207. The recombinant cell according to C206, wherein the N-glycosylation site comprises position N235 of the polypeptide.
- 10 C208. The recombinant cell according to C206, wherein the N-glycosylation site comprises position N228 of the polypeptide.
- C209. The recombinant cell according to C206, wherein the N-glycosylation site comprises position N235 and position N228 of the polypeptide.
- C210. The recombinant cell according to C195, wherein the polypeptide comprises SEQ ID
15 NO: 3.
- C211. The recombinant cell according to C195, wherein the polypeptide comprises SEQ ID NO: 2.
- C212. The recombinant cell according to C194, wherein the polypeptide comprises an aliphatic hydrophobic amino acid residue at position 1 of the polypeptide.
- 20 C213. The recombinant cell according to C212, wherein the aliphatic hydrophobic amino acid residue is selected from the group consisting of Ile, Leu, and Val.
- C214. The recombinant cell according to C194, wherein the polypeptide comprises a fragment of FimH.
- C215. The recombinant cell according to C214, wherein the polypeptide comprises a lectin
25 domain of FimH.
- C216. The recombinant cell according to C215, wherein the lectin domain comprises a mass of about 17022 Daltons.
- C217. The recombinant cell according to C194, wherein the polypeptide is complexed with a FimC polypeptide or a fragment thereof.
- 30 C218. The recombinant cell according to C217, wherein the FimC polypeptide or a fragment thereof comprises a glycine residue at position 37 of the FimC polypeptide or a fragment thereof.
- C219. The recombinant cell according to C195, wherein the polypeptide is in the low affinity conformation.
- 35 C220. The recombinant cell according to C195, wherein the polypeptide is stabilized by FimG.

- C221. The recombinant cell according to C195, wherein the polypeptide is stabilized by a donor-strand peptide of FimG (DsG).
- C222. The recombinant cell according to C221, wherein the polynucleotide sequence further encodes a linker sequence.
- 5 C223. The recombinant cell according to C222, wherein the linker comprises at least 4 amino acid residues and at most 15 amino acid residues.
- C224. The recombinant cell according to C222, wherein the linker comprises at least 5 amino acid residues and at most 10 amino acid residues.
- C225. The recombinant cell according to C222, wherein the linker comprises 7 amino acid
10 residues.
- C226. The recombinant cell according to C194, wherein the polypeptide does not comprise a signal peptide selected from the group consisting of a native FimH leader peptide, influenza hemagglutinin signal peptide, and a human respiratory syncytial virus A (strain A2) fusion glycoprotein F0 signal peptide.
- 15 C227. The recombinant cell according to C194, wherein the polypeptide comprises a murine IgK signal peptide sequence.
- C228. The recombinant cell according to C194, wherein the polypeptide comprises any one signal peptide sequence selected from human IgG receptor FcRn large subunit p51 signal peptide and a human IL10 protein signal peptide.
- 20 C229. The recombinant cell according to C195, wherein the polypeptide comprises a mutation of arginine to proline at amino acid position 60 (R60P), according to the numbering of SEQ ID NO: 3.
- C230. The recombinant cell according to C194, wherein the expression level of the polypeptide is greater than the expression level of the corresponding wild-type polypeptide expressed in
25 the periplasm of a wild-type *E. coli* cell.
- C231. The recombinant cell according to C194, wherein the expression level of the polypeptide is greater than 10 mg/L.
- C232. The recombinant cell according to C194, wherein the polynucleotide sequence is integrated into the genomic DNA of said mammalian cell.
- 30 C233. The recombinant cell according to C194, wherein the polynucleotide sequence is codon optimized for expression in the cell.
- C234. The recombinant cell according to C194, wherein the cell is a human embryonic kidney cell.
- C235. The recombinant cell according to C234, wherein the human embryonic kidney cell
35 comprises a HEK293 cell.

- C236. The recombinant cell according to C235, wherein the HEK293 cell is selected from any one of HEK293T cells, HEK293TS cells, and HEK293E cells.
- C237. The recombinant cell according to C195, wherein the cell is a CHO cell.
- C238. The recombinant cell according to C237, wherein said CHO cell is a CHO-K1 cell, CHO-DUXB11, CHO-DG44 cell, or CHO-S cell.
- C239. The recombinant cell according to C194, wherein the polypeptide is soluble.
- C240. The recombinant cell according to C194, wherein the polypeptide is secreted from the cell.
- C241. The recombinant cell according to C195, wherein the polypeptide comprises a N28Q substitution, according to the numbering of SEQ ID NO: 1.
- C242. The recombinant cell according to C195, wherein the polypeptide comprises a N28D substitution, according to the numbering of SEQ ID NO: 1.
- C243. The recombinant cell according to C195, wherein the polypeptide comprises a N28S substitution, according to the numbering of SEQ ID NO: 1.
- C244. The recombinant cell according to C195, wherein the polypeptide comprises a substitution selected from any one of N28Q, V48C, and L55C, according to the numbering of SEQ ID NO: 1.
- C245. The recombinant cell according to C195, wherein the polypeptide comprises a substitution N92S according to the numbering of SEQ ID NO: 1.
- C246. The recombinant cell according to C194, wherein the polypeptide derived from FimH or fragment thereof comprises a substitution selected from any one of V48C and L55C, according to the numbering of SEQ ID NO: 1.
- C247. A culture comprising the recombinant cell of C194, wherein said culture is at least 5 liter in size.
- C248. The culture according to C242, wherein the yield of the polypeptide or fragment thereof is at least 0.05 g/L.
- C249. The culture according to C248, wherein the yield of the polypeptide or fragment thereof is at least 0.10 g/L.
- C250. A method for producing a polypeptide derived from *E. coli* or a fragment thereof, comprising culturing a recombinant mammalian cell according to C194 under a suitable condition, thereby expressing the polypeptide or fragment thereof; and harvesting the polypeptide or fragment thereof.
- C251. The method according to C250, further comprising purifying the polypeptide or fragment thereof.
- C252. The method according to C250, wherein the cell comprises a nucleic acid encoding any one of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, and SEQ ID NO: 27.

- C253. The method according to C250, wherein the yield of the polypeptide or fragment thereof is at least 0.05 g/L.
- C254. The method according to C250, wherein the yield of the polypeptide or fragment thereof is at least 0.10 g/L.
- 5 C255. A composition comprising a polypeptide having at least 70% identity to any one of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, and SEQ ID NO: 29.
- C256. The composition according to C255, further comprising a saccharide comprising a structure selected from any one Formula in **Table 1**.
- 10 C257. The composition according to C256, wherein the saccharide is covalently bound a carrier protein.
- C258. The composition according to C257, wherein the carrier protein is selected from any one of poly(L-lysine), CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).
- 15 C259. The composition according to C257, wherein the carrier protein is CRM₁₉₇.
- C260. The composition according to C257, wherein the carrier protein is tetanus toxoid (TT).
- C261. The composition according to C257, wherein the carrier protein is poly(L-lysine).
- C262. The composition according to C257, wherein the saccharide is covalently bound a carrier protein by reductive amination.
- 25 C263. The composition according to C257, wherein the saccharide is covalently bound a carrier protein by CDAP chemistry.
- C264. The composition according to C257, wherein the saccharide is covalently bound a carrier protein by single-end linked conjugation.
- C265. The composition according to C257, wherein the saccharide is covalently bound a carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer.
- 30 C266. A polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, and SEQ ID NO: 27.
- C267. A composition comprising a polypeptide derived from FimH or a fragment thereof; and a saccharide comprising a structure selected from any one of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2;
- 35

K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula
 O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17,
 Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula
 O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula
 5 O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula
 O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34,
 Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40,
 Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45,
 Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51,
 10 Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57,
 Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D1,
 Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69,
 Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75,
 Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81,
 15 Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87,
 Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93,
 Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100,
 Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula
 O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111,
 20 Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula
 O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123,
 Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula
 O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134,
 Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula
 25 O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145,
 Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula
 O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156,
 Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula
 O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167,
 30 Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula
 O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178,
 Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula
 O184, Formula O185, Formula O186, Formula O187.

C268. The composition according to Clause C267, further comprising at least one saccharide
 35 derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2,
 O3, and O5.

- C269. The composition according to Clause C267, further comprising a saccharide derived from *Klebsiella pneumoniae* type O1.
- C270. The composition according to Clause C267, further comprising a saccharide derived from *K. pneumoniae* type O2.
- 5 C271. The composition according to Clause C267, further comprising a saccharide derived from *K. pneumoniae* type O3.
- C272. The composition according to Clause C267, further comprising a saccharide derived from *K. pneumoniae* type O5.
- C273. The composition according to Clause C267, further comprising a saccharide derived from *K. pneumoniae* type O1 and a saccharide derived from *K. pneumoniae* type O2.
- 10 C274. The composition according to Clause C268, wherein the saccharide derived from *K. pneumoniae* is conjugated to a carrier protein; and the saccharide derived from *E. coli* is conjugated to a carrier protein.
- C275. The composition according to Clause C267, further comprising a polypeptide derived from *K. pneumoniae*.
- 15 C276. A composition comprising a polypeptide derived from FimH or a fragment thereof; and at least one saccharide derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2, O3, and O5.
- C277. The composition according to Clause C276, further comprising at least one saccharide comprising a structure selected from any one of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D1, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81,
- 20
- 25
- 30
- 35

- Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, Formula O187.
- 20 C278. The composition according to Clause C277, wherein the saccharide derived from *K. pneumoniae* is conjugated to a carrier protein; and the saccharide derived from *E. coli* is conjugated to a carrier protein.
- C279. The composition according to Clause C277, further comprising a polypeptide derived from *K. pneumoniae*.
- 25 C280. A composition comprising at least one saccharide derived from any one *K. pneumoniae* type selected from the group consisting of O1, O2, O3, and O5; and at least one saccharide comprising a structure selected from any one of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40,

Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45,
 Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51,
 Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57,
 Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula 62D1,
 5 Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69,
 Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75,
 Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81,
 Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87,
 Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93,
 10 Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100,
 Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula
 O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111,
 Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula
 O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123,
 15 Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula
 O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134,
 Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula
 O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145,
 Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula
 20 O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156,
 Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula
 O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167,
 Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula
 O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178,
 25 Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula
 O184, Formula O185, Formula O186, Formula O187.

- C281. The composition according to Clause C280, further comprising a polypeptide derived
 from FimH or a fragment thereof.
 C282. The composition according to Clause C280, wherein the E. coli saccharide comprises
 30 Formula O8.
 C283. The composition according to Clause C280, wherein the E. coli saccharide comprises
 Formula O9.
 C284. The composition according to Clause C280, further comprising a polypeptide derived
 from K. pneumoniae.
 35 C285. The composition according to any of Clauses C267-C284, wherein the saccharide is
 covalently bound to a carrier protein.

- C286. The composition according to Clause C285, wherein the saccharide further comprises a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- C287. The composition according to Clause C285, wherein the saccharide comprises Lipid A.
- C288. The composition according to any one of clauses C285-C287, wherein the saccharide is
5 synthetically synthesized.
- C289. The composition according to Clause C285, wherein the carrier protein is selected from any one of CRM197, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding
10 protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins and Streptococcal C5a peptidase (SCP).
- C290. A method of eliciting an immune response against *Escherichia coli* in a mammal,
15 comprising administering to the mammal an effective amount of the composition according to any one of Clauses C267-C289.
- C291. The method according to Clause C290, wherein the immune response comprises opsonophagocytic antibodies against *E. coli*.
- C292. The method according to Clause C290, wherein the immune response protects the
20 mammal from an *E. coli* infection.
- C293. A method of eliciting an immune response against *Klebsiella pneumoniae* in a mammal, comprising administering to the mammal an effective amount of the composition according to any one of Clauses C267-C289.
- C294. The method according to Clause C293, wherein the immune response comprises
25 opsonophagocytic antibodies against *Klebsiella pneumoniae*.
- C295. The method according to Clause C293, wherein the immune response protects the mammal from a *Klebsiella pneumoniae* infection.
- C296. The compositions and methods according to any one of Clauses C1-C266, further comprising at least one saccharide derived from any one of *K. pneumoniae* type selected
30 from the group consisting of O1, O2, O3, and O5.
- C297. The compositions and methods of Clause C296, wherein the *K. pneumoniae* serotype O1 comprises variant O1V1 or O1V2.
- C298. The compositions and methods of Clause C296, wherein the *K. pneumoniae* serotype O2 comprises variant O2V1 or O2V2.
- C299. Use of the compositions set forth in any one of Clauses C1-C298 as set forth herein.
- C300. A composition comprising

- (i) a polypeptide derived from fimbrial H (FimH) or a fragment thereof; and
- (ii) one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a saccharide, said saccharide comprising a structure selected from the group consisting of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62D1, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula

O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187,

wherein n is an integer from 1 to 100.

- 5 C301. The composition according to clause C300, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, Formula O6, Formula O8 and Formula O9.
- C302. The composition according to clause C301, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, and Formula O6.
- 10 C303. A composition comprising a polypeptide derived from Fimbrial antigen H (FimH) or a fragment thereof; and
- (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,
 - (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising
 - 15 Formula O1A, wherein n is an integer from 31 to 90,
 - (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90,
 - (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90,
 - 20 (e) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O8, wherein n is an integer from 31 to 90, and
 - (f) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O9, wherein n is an integer from 31 to 90.
- C304. The composition of clause C303 comprising a polypeptide derived from Fimbrial antigen
- 25 H (FimH) or a fragment thereof; and
- (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,
 - (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,
 - 30 (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90, and
 - (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90.
- C305. The composition according to clauses C303 or C304, further comprising one or more
- 35 conjugates having a saccharide selected from the group consisting of Formula O15, Formula O16, Formula O17, Formula O18 and Formula O75, wherein n is an integer from

- 31 to 90.7. The composition according to any one of clauses 1-6, wherein the saccharide further comprises any one of an E. coli R1 moiety, E. coli R2 moiety, E. coli R3 moiety, E. coli R4 moiety, and E. coli K-12 moiety.
- 5 C306. The composition according to any one of clauses C300-C304, wherein the saccharide does not further comprise any one of an E. coli R1 moiety, E. coli R2 moiety, E. coli R3 moiety, E. coli R4 moiety, and E. coli K-12 moiety.
- C307. The composition according to clause C306, wherein the saccharide does not further comprise an E. coli R2 moiety.
- 10 C308. The composition according to any one of clauses C300-C307, wherein the saccharide further comprises a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- C309. The composition according to any one of clauses C300-C308, wherein the carrier protein is selected from the group consisting of CRM197, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxin, cholera toxin, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins, and Streptococcal C5a peptidase (SCP) or a variant thereof.
- 15 C310. The composition according to clause C309, wherein the carrier protein is CRM197.
- 20 C311. The composition according to clause C309, wherein the carrier protein is tetanus toxoid.
- C312. The composition according to any one of clauses C300-C311, wherein the ratio of saccharide to protein is at least 0.5 to at most 2.
- C313. The composition according to any one of clauses C300-C312, wherein the conjugate is prepared via reductive amination.
- 25 C314. The composition according to any one of clauses C300-C312, wherein the saccharide is conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer.
- C315. The composition according to any one of clauses C300-C314, wherein the saccharide is a single-end linked conjugated saccharide.
- 30 C316. The composition according to clause C308, wherein the saccharide is conjugated to the carrier protein through a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) residue.
- C317. The composition according to any one of clauses C300-C313, wherein the conjugate is prepared via CDAP chemistry.
- 35 C318. The composition according to any one of clauses C300-C317, wherein the polypeptide is derived from E. coli fimbrial H (FimH).

- C319. The composition according to clause C318, wherein the polypeptide comprises a phenylalanine residue at the N-terminus of the polypeptide.
- C320. The composition according to clause C318, wherein the polypeptide comprises a phenylalanine residue within the first 20 residue positions of the N-terminus.
- 5 C321. The composition according to clause C314, wherein the polypeptide comprises a phenylalanine residue at position 1 of the polypeptide.
- C322. The composition according to clause C317, wherein the polypeptide does not comprise a glycine residue immediately before the phenylalanine residue at position 1 of the polypeptide.
- 10 C323. The composition according to clause C314, wherein the polypeptide does not comprise an N-glycosylation site at position 7 of the polypeptide.
- C324. The composition according to clause C318, wherein the polypeptide does not comprise an Asn residue at position 7 of the polypeptide.
- C325. The composition according to clause C320, wherein the polypeptide comprises a
15 residue selected from the group consisting of Ser, Asp, Thr, and Gln at position 7.
- C326. The composition according to clause C317, wherein the polypeptide does not comprise an N-glycosylation site at position 70 of the polypeptide.
- C327. The composition according to clause C322, wherein the polypeptide does not comprise an Asn residue at position 70 of the polypeptide.
- 20 C328. The composition according to clause C322, wherein the polypeptide does not comprise a Ser residue at position 70 of the polypeptide.
- C329. The composition according to clause C314, wherein the polypeptide comprises a residue substitution selected from the group consisting of Ser, Asp, Thr, and Gln at an N-glycosylation site of the polypeptide.
- 25 C330. The composition according to clause C325, wherein the N-glycosylation site comprises position N235 of the polypeptide.
- C331. The composition according to clause C325, wherein the N-glycosylation site comprises position N228 of the polypeptide.
- C332. The composition according to clause C325, wherein the N-glycosylation site comprises
30 position N235 and position N228 of the polypeptide.
- C333. The composition according to clause C314, wherein the polypeptide comprises SEQ ID NO: 2.
- C334. The composition according to clause C314, wherein the polypeptide comprises SEQ ID NO: 3.
- 35 C335. The composition according to clause C314, wherein the polypeptide comprises amino acids having an amino acid sequence selected from the group consisting of SEQ ID NO: 5,

SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 27, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113.

- 5 C336. The composition of claim C314, wherein the polypeptide comprises amino acids having an amino acid sequence having at least 70% identity to the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112 and SEQ ID NO: 113.
- 10 C337. The composition according to claim C314, wherein the polypeptide comprises an aliphatic hydrophobic amino acid residue at position 1 of the polypeptide.
- C338. The composition according to claim C333, wherein the aliphatic hydrophobic amino acid residue is selected from the group consisting of Ile, Leu, and Val.
- C339. The composition according to claim C314, wherein the polypeptide comprises a fragment of FimH.
- 15 C340. The composition according to claim C335, wherein the polypeptide comprises a lectin domain of FimH.
- C341. The composition according to claim C336, wherein the lectin domain comprises a mass of about 17022 Daltons.
- C342. The composition according to claim C314, wherein the polypeptide is complexed with a FimC polypeptide or a fragment thereof.
- 20 C343. The composition according to claim C338, wherein the FimC polypeptide or a fragment thereof comprises a glycine residue at position 37 of the FimC polypeptide or a fragment thereof.
- C344. The composition according to claim C314, wherein the polypeptide is in the low affinity conformation.
- 25 C345. The composition according to claim C314, wherein the polypeptide is stabilized by FimG.
- C346. The composition according to claim C314, wherein the polypeptide is stabilized by a donor-strand peptide of FimG (DsG).
- C347. The composition according to claim C346, wherein the polypeptide further comprises a linker.
- 30 C348. The composition according to claim C343, wherein the linker comprises at least 4 amino acid residues and at most 15 amino acid residues.
- C349. The composition according to claim C343, wherein the linker comprises at least 5 amino acid residues and at most 10 amino acid residues.
- 35 C350. The composition according to claim C343, wherein the linker comprises 7 amino acid residues.

- C351. The composition according to clause C314, wherein the polypeptide does not comprise a signal peptide selected from the group consisting of a native FimH leader peptide, influenza hemagglutinin signal peptide, and a human respiratory syncytial virus A (strain A2) fusion glycoprotein F0 signal peptide.
- 5 C352. The composition according to clause C314, wherein the polypeptide comprises a murine IgK signal peptide sequence.
- C353. The composition according to clause C314, wherein the polypeptide comprises any one signal peptide sequence selected from human IgG receptor FcRn large subunit p51 signal peptide and a human IL10 protein signal peptide.
- 10 C354. The composition according to clause C314, wherein the polypeptide comprises a mutation of arginine to proline at amino acid position 60 (R60P), according to the numbering of SEQ ID NO: 3.
- C355. The composition according to clauses C300-C354, comprising at most 25% free saccharide as compared to the total amount of saccharide in the composition.
- 15 C356. The composition of any one of clauses C300-C354, further comprising one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a K. pneumoniae O-antigen selected from O1 and O2.
- C357. The composition of clause C356, wherein the K. pneumoniae O-antigen is selected from the group consisting of a) serotype O1 subtype v1 (O1v1), b) serotype O1 subtype v2 (O1v2), c) serotype O2 subtype v1 (O2v1), and d) serotype O2 subtype v2 (O2v2).
- 20 C358. A composition comprising
- (i) one or more conjugates comprising a carrier protein covalently bound to a K. pneumoniae O-antigen selected from the group consisting of serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and
- 25 (ii) one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a saccharide, said saccharide comprising a structure selected from the group consisting of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8,
- 30 Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula
- 35 O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37,

Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, 5 Formula O61, Formula O62, Formula 62D1, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, 10 Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula 15 O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, 20 Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula 25 O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187,

wherein n is an integer from 1 to 100.

30 C359. The composition according to clause C358, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, Formula O6, Formula O8 and Formula O9.

C360. The composition according to clause C359, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, and Formula O6.

35 C361. A composition comprising one or more conjugates comprising a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from the group consisting of

serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and

(a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,

5 (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,

(c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90,

(d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90,

(e) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O8, wherein n is an integer from 31 to 90, and

(f) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O9, wherein n is an integer from 31 to 90.

15 C362. The composition of clause C361 comprising (i) one or more conjugates comprising a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from the group consisting of serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and

(a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising 20 Formula O25b, wherein n is an integer from 31 to 90,

(b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,

(c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90, and

25 (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90.

C363. The composition according to clauses C361 or C362, further comprising one or more conjugates having a saccharide selected from the group consisting of Formula O15, Formula O16, Formula O17, Formula O18 and Formula O75, wherein n is an integer from 30 31 to 90.

C364. The composition according to any one of clauses C358-C363, wherein the saccharide further comprises any one of an *E. coli* R1 moiety, *E. coli* R2 moiety, *E. coli* R3 moiety, *E. coli* R4 moiety, and *E. coli* K-12 moiety.

35 C365. The composition according to any one of clauses C358-C363, wherein the saccharide does not further comprise any one of an *E. coli* R1 moiety, *E. coli* R2 moiety, *E. coli* R3 moiety, *E. coli* R4 moiety, and *E. coli* K-12 moiety.

- C366. The composition according to clause 67, wherein the saccharide does not further comprise an *E. coli* R2 moiety.
- C367. The composition according to any one of clauses C358-C366, wherein the saccharide further comprises a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) moiety.
- 5 C368. The composition according to any one of clauses C358-C367, wherein the carrier protein is selected from the group consisting of CRM197, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*,
10 clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins, and Streptococcal C5a peptidase (SCP) or a variant thereof.
- C369. The composition according to clause C368, wherein the carrier protein is CRM197.
- C370. The composition according to clause C368, wherein the carrier protein is tetanus toxoid.
- 15 C371. The composition according to any one of clauses C358-C370, wherein the ratio of saccharide to protein is at least 0.5 to at most 2.
- C372. The composition according to any one of clauses C358-C371, wherein the conjugate is prepared via reductive amination.
- C373. The composition according to any one of clauses C358-C371, wherein the saccharide is
20 conjugated to the carrier protein through a (2-((2-oxoethyl)thio)ethyl)carbamate (eTEC) spacer.
- C374. The composition according to any one of clauses C358-C373, wherein the saccharide is a single-end linked conjugated saccharide.
- C375. The composition according to clause C367, wherein the saccharide is conjugated to the
25 carrier protein through a 3-deoxy-d-manno-oct-2-ulosonic acid (KDO) residue.
- C376. The composition according to any one of clauses C358-C372, wherein the conjugate is prepared via CDAP chemistry.
- C377. The composition according to clauses C358-C376, comprising at most 25% free saccharide as compared to the total amount of saccharide in the composition.
- 30 C378. The composition of any one of clauses C300-C377 further comprising an adjuvant.
- C379. The composition of clause C378, wherein the adjuvant comprises aluminum.
- C380. The composition of clause C378, wherein the adjuvant comprises QS21.
- C381. The composition of clause C378, wherein the adjuvant comprises MPLA.
- C382. The composition of clause C380 or C381, wherein the adjuvant is a liposomal adjuvant.

- C383. A method of eliciting an immune response against *Escherichia coli* in a mammal, comprising administering to the mammal an effective amount of the composition according to any one of clauses C300-C382.
- 5 C384. The method according to clause C383, wherein the immune response comprises opsonophagocytic antibodies against *E. coli*.
- C385. The method according to clause C383, wherein the immune response protects the mammal from an *E. coli* infection.
- C386. Use of the composition of any one of clauses C300-C382 for inducing an immune response against *E. coli*.
- 10 C387. Use of the composition of any one of clauses C300-C382 in the manufacture of a medicament for inducing an immune response against *E. coli*.
- C388. The use of clause C386 or C387, wherein the immune response comprises opsonophagocytic antibodies against *E. coli*.
- C389. The use of clause C386 or C387, wherein the immune response protects the mammal
15 from an *E. coli* infection.
- C390. A method of eliciting an immune response against *K. pneumoniae* in a mammal, comprising administering to the mammal an effective amount of the composition according to any one of clauses C300-C382.
- C391. The method according to clause C390, wherein the immune response comprises
20 opsonophagocytic antibodies against *K. pneumoniae*.
- C392. The method according to clause C390, wherein the immune response protects the mammal from a *K. pneumoniae* infection.
- C393. Use of the composition of any one of clauses C300 - C382 for inducing an immune response against *K. pneumoniae*.
- 25 C394. Use of the composition of any one of clauses C300 - C382 in the manufacture of a medicament for inducing an immune response against *K. pneumoniae*.
- C395. The use of clause C386 or C387, wherein the immune response comprises opsonophagocytic antibodies against *K. pneumoniae*.
- C396. The use of clause C386 or C387, wherein the immune response protects the mammal
30 from an *K. pneumoniae* infection.
- C397. A nucleic acid comprising nucleotides encoding the polypeptides of any one of clauses C1-C382.
- C398. The nucleic acid of clause C397, wherein the nucleic acid is RNA.
- C399. A nanoparticle comprising the nucleic acid of clause C397 or C398.
- 35 C400. The composition according to any one of clauses 303, 304, 305, 359, 360, 361, 362 or 363, further comprising one or more conjugates having a saccharide selected from the

group consisting of Formula O4, Formula O11, Formula O13, Formula O21 and Formula O86, wherein n is an integer from 1 to 100, preferably from 31 to 90.

CLAIMS

1. A composition comprising
- (i) a polypeptide derived from fimbrial H (FimH) or a fragment thereof; and
- (ii) one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a saccharide, said saccharide comprising a structure selected from the group consisting of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula

- O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181, Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187,
- 5 wherein n is an integer from 1 to 100.
2. A composition comprising a polypeptide derived from Fimbrial antigen H (FimH) or a fragment thereof; and
- (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising
- 10 Formula O25b, wherein n is an integer from 31 to 90,
- (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,
- (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90,
- 15 (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90,
- (e) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O8, wherein n is an integer from 31 to 90, and
- (f) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and
- 20 Formula O9, wherein n is an integer from 31 to 90.
3. The composition of claim 2, comprising a polypeptide derived from Fimbrial antigen H (FimH) or a fragment thereof; and
- (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,
- 25 (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,
- (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90, and
- (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising
- 30 and Formula O6, wherein n is an integer from 31 to 90.
4. The composition according to claims 2 or 3, further comprising one or more conjugates having a saccharide selected from the group consisting of Formula O15, Formula O16, Formula O17, Formula O18 and Formula O75, wherein n is an integer from 31 to 90.
5. The composition according to any one of claims 1-4, wherein the carrier protein is selected
- 35 from the group consisting of CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera

- toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae* Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins, and Streptococcal C5a peptidase (SCP) or a variant thereof.
- 5 6. The composition according to any one of claims 1-5, wherein the polypeptide is derived from *E. coli* fimbrial H (FimH).
7. The composition according to claim 6, wherein the polypeptide comprises SEQ ID NO: 2.
8. The composition according to claim 6, wherein the polypeptide comprises SEQ ID NO: 3.
- 10 9. The composition according to claim 6, wherein the polypeptide comprises amino acids having an amino acid sequence selected from the group consisting of SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 27, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113.
10. The composition of claim 6, wherein the polypeptide comprises amino acids having an amino acid sequence having at least 70% identity to the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 20, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, and SEQ ID NO: 113.
- 15 11. The composition according to claim 6, wherein the polypeptide comprises a fragment of FimH.
- 20 12. The composition according to claim 11, wherein the polypeptide comprises a lectin domain of FimH.
13. The composition according to claim 6, wherein the polypeptide is complexed with a FimC polypeptide or a fragment thereof.
- 25 14. The composition according to claim 6, wherein the polypeptide is stabilized by a donor-strand peptide of FimG (DsG).
15. The composition according to claim 14, wherein the polypeptide further comprises a linker.
16. The composition of any one of claims 1-15, further comprising one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from O1 and O2.
- 30 17. A composition comprising
- (i) one or more conjugates comprising a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from the group consisting of serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and
- 35

(ii) one or more conjugates, wherein the conjugate comprises a carrier protein covalently bound to a saccharide, said saccharide comprising a structure selected from the group consisting of Formula O1, Formula O1A, Formula O1B, Formula O1C, Formula O2, Formula O3, Formula O4, Formula O4:K52, Formula O4:K6, Formula O5, Formula O5ab, Formula O5ac, Formula O6, Formula O6:K2; K13; K15, Formula O6:K54, Formula O7, Formula O8, Formula O9, Formula O10, Formula O11, Formula O12, Formula O13, Formula O14, Formula O15, Formula O16, Formula O17, Formula O18, Formula O18A, Formula O18ac, Formula O18A1, Formula O18B, Formula O18B1, Formula O19, Formula O20, Formula O21, Formula O22, Formula O23, Formula O23A, Formula O24, Formula O25, Formula O25a, Formula O25b, Formula O26, Formula O27, Formula O28, Formula O29, Formula O30, Formula O32, Formula O33, Formula O34, Formula O35, Formula O36, Formula O37, Formula O38, Formula O39, Formula O40, Formula O41, Formula O42, Formula O43, Formula O44, Formula O45, Formula O45, Formula O45rel, Formula O46, Formula O48, Formula O49, Formula O50, Formula O51, Formula O52, Formula O53, Formula O54, Formula O55, Formula O56, Formula O57, Formula O58, Formula O59, Formula O60, Formula O61, Formula O62, Formula O62D₁, Formula O63, Formula O64, Formula O65, Formula O66, Formula O68, Formula O69, Formula O70, Formula O71, Formula O73, Formula O73, Formula O74, Formula O75, Formula O76, Formula O77, Formula O78, Formula O79, Formula O80, Formula O81, Formula O82, Formula O83, Formula O84, Formula O85, Formula O86, Formula O87, Formula O88, Formula O89, Formula O90, Formula O91, Formula O92, Formula O93, Formula O95, Formula O96, Formula O97, Formula O98, Formula O99, Formula O100, Formula O101, Formula O102, Formula O103, Formula O104, Formula O105, Formula O106, Formula O107, Formula O108, Formula O109, Formula O110, Formula O111, Formula O112, Formula O113, Formula O114, Formula O115, Formula O116, Formula O117, Formula O118, Formula O119, Formula O120, Formula O121, Formula O123, Formula O124, Formula O125, Formula O126, Formula O127, Formula O128, Formula O129, Formula O130, Formula O131, Formula O132, Formula O133, Formula O134, Formula O135, Formula O136, Formula O137, Formula O138, Formula O139, Formula O140, Formula O141, Formula O142, Formula O143, Formula O144, Formula O145, Formula O146, Formula O147, Formula O148, Formula O149, Formula O150, Formula O151, Formula O152, Formula O153, Formula O154, Formula O155, Formula O156, Formula O157, Formula O158, Formula O159, Formula O160, Formula O161, Formula O162, Formula O163, Formula O164, Formula O165, Formula O166, Formula O167, Formula O168, Formula O169, Formula O170, Formula O171, Formula O172, Formula O173, Formula O174, Formula O175, Formula O176, Formula O177, Formula O178, Formula O179, Formula O180, Formula O181,

Formula O182, Formula O183, Formula O184, Formula O185, Formula O186, and Formula O187,

wherein n is an integer from 1 to 100.

18. The composition according to claim 17, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, Formula O6, Formula O8 and Formula O9.
19. The composition according to claim 18, wherein the saccharide is selected from the group consisting of Formula O25b, Formula O1A, Formula O2, and Formula O6.
20. A composition comprising one or more conjugates comprising a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from the group consisting of serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and
 - (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,
 - (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,
 - (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90,
 - (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90,
 - (e) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O8, wherein n is an integer from 31 to 90, and
 - (f) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O9, wherein n is an integer from 31 to 90.
21. The composition of claim 20, comprising (i) one or more conjugates comprising a carrier protein covalently bound to a *K. pneumoniae* O-antigen selected from the group consisting of serotype O1 subtype v1 (O1v1), serotype O1 subtype v2 (O1v2), serotype O2 subtype v1 (O2v1), and serotype O2 subtype v2 (O2v2); and
 - (a) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O25b, wherein n is an integer from 31 to 90,
 - (b) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O1A, wherein n is an integer from 31 to 90,
 - (c) a conjugate comprising a carrier protein covalently bound to a saccharide comprising Formula O2, wherein n is an integer from 31 to 90, and
 - (d) a conjugate comprising a carrier protein covalently bound to a saccharide comprising and Formula O6, wherein n is an integer from 31 to 90.

22. The composition according to claims 20 or 21, further comprising one or more conjugates having a saccharide selected from the group consisting of Formula O15, Formula O16, Formula O17, Formula O18 and Formula O75, wherein n is an integer from 31 to 90.
23. The composition according to any one of claims 17-22, wherein the carrier protein is
- 5 selected from the group consisting of CRM₁₉₇, diphtheria toxin fragment B (DTFB), DTFB C8, Diphtheria toxoid (DT), tetanus toxoid (TT), fragment C of TT, pertussis toxoid, cholera toxoid, or exotoxin A from *Pseudomonas aeruginosa*; detoxified Exotoxin A of *P. aeruginosa* (EPA), maltose binding protein (MBP), detoxified hemolysin A of *S. aureus*, clumping factor A, clumping factor B, Cholera toxin B subunit (CTB), *Streptococcus pneumoniae*
- 10 Pneumolysin and detoxified variants thereof, *C. jejuni* AcrA, *C. jejuni* natural glycoproteins, and Streptococcal C5a peptidase (SCP) or a variant thereof.
24. The composition according to claim 23, wherein the carrier protein is CRM₁₉₇.
25. The composition of any one of claims 1-24, further comprising an adjuvant.
26. A method of eliciting an immune response against *Escherichia coli* in a mammal, comprising
- 15 administering to the mammal an effective amount of the composition according to any one of claims 1-25.
27. A method of eliciting an immune response against *K. pneumoniae* in a mammal, comprising administering to the mammal an effective amount of the composition according to any one of claims 1-25.
- 20 28. A nucleic acid comprising nucleotides encoding the polypeptides of any one of claims 1-23.