BIOSYNTHETIC SYSTEM THAT PRODUCES IMMUNOGENIC POLYSACCHARIDES IN PROKARYOTIC CELLS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Patent Application No. 61/272,931, filed November 19, 2009, herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates to the use of a biosynthetic system and proteins for preparing a vaccine. In addition, the invention relates to a recombinant prokaryotic biosynthetic system having an epimerase that initiates the synthesis of an oligo- or polysaccharide with a specified monosaccharide at the reducing terminus. The invention further relates to *N*-glycosylated proteins produced with glycans in an expression system and bioconjugate vaccines made from said *N*-glycosylated proteins comprising immunogenic glycans, and provides methods for producing *N*-glycosylated proteins.

BACKGROUND OF THE INVENTION

[0003] Glycoproteins are proteins that have one or more covalently attached sugar polymers. *N*-linked protein glycosylation is an essential and conserved process occurring in the endoplasmic reticulum of eukaryotic organisms. It is important for protein folding, oligomerization, stability, quality control, sorting and transport of secretory and membrane proteins (Helenius, A., and Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. 73, 1019-1049).

[0004] Protein glycosylation has a profound influence on the immunogenicity, the stability and the half-life of a protein. In addition, glycosylation can assist the purification of proteins by chromatography, e.g. affinity chromatography with lectin ligands bound to a solid phase interacting with glycosylated moieties of the protein. It is therefore established practice to produce

many glycosylated proteins recombinantly in eukaryotic cells to provide biologically and pharmaceutically useful glycosylation patterns.

[0005] WO 2003/07467 (Aebi *et al.*) demonstrated that the food-borne pathogen *Campylobacter jejuni,* which is a bacterium, could *N*-glycosylate its proteins, which was a unique feature among known prokaryotic organisms except for certain species of archaea. The machinery required for glycosylation is encoded by 12 genes that are clustered in the so-called *pgl* locus. Disruption of *N*-glycosylation affects invasion and pathogenesis of *C. jejuni* but is not lethal as in most eukaryotic organisms (Burda P. and M. Aebi, (1999). The dolichol pathway of *N*-linked glycosylation. Biochem Biophys Acta 1426(2):239-57). It is possible to reconstitute the *N*-glycosylation of *C. jejuni* proteins by recombinantly expressing the *pgl* locus and acceptor glycoprotein in *E. coli* at the same time (Wacker et al. (2002). *N*-linked glycosylation in *Campylobacter* jejuni and its functional transfer into *E. coli.* Science 298, 1790-1793).

[0006] *N*-glycans have a glycan attached to a consensus sequence in a protein. The known *N*-glycosylation consensus sequence in a protein allows for the *N*-glycosylation of recombinant target proteins in prokaryotic organisms. Such organisms comprise an oligosaccharyl transferase ("OT"; "OTase"), such as, for example, an oligosaccharyl transferase of *C. jejuni*, which is an enzyme that transfers the glycan to the consensus sequence of the protein.

[0007] WO 2003/07467 (Aebi *et al.*) teaches a prokaryotic organism into which is introduced a nucleic acid encoding for (i) specific glycosyltransferases for the assembly of an oligosaccharide on a lipid carrier, (ii) a recombinant target protein comprising a consensus sequence "N - X - S/T", wherein X can be any amino acid except proline, and (iii) an oligosaccharyl transferase, such as, for example, an oligosaccharyl transferase of *C. jejuni* that covalently links said oligosaccharide to the consensus sequence of the target protein. Said prokaryotic organism produces *N*-glycans with a specific structure which is defined by the type of the specific glycosyltransferases.

[0008] WO 2006/119987 (Aebi *et al.*) describes proteins, as well as means and methods for producing proteins, with efficiency for *N*-glycosylation in prokaryotic organisms *in vivo.* It further describes an efficient introduction of *N*-

glycans into recombinant proteins for modifying immunogenicity, stability, biological, prophylactic and/or therapeutic activity of said proteins, and the provision of a host cell that efficiently displays recombinant *N*-glycosylated proteins of the present invention on its surface. In addition, it describes a recombinant *N*-glycosylated protein comprising one or more of the following *N*-glycosylated optimized amino acid sequence(s):

D/E - X - N - Z - S/T (optimized consensus sequence),

wherein X and Z may be any natural amino acid except Pro, and wherein at least one of said *N*-glycosylated partial amino acid sequence(s) is introduced. The introduction of specific partial amino acid sequence(s) (optimized consensus sequence(s)) into proteins leads to proteins that are efficiently *N*- glycosylated by an oligosaccharyl transferase in these introduced positions.

[0009] The biosynthesis of different polysaccharides is conserved in bacterial cells. The polysaccharides are assembled on carrier lipids from common precursors (activated sugar nucleotides) at the cytoplasmic membrane by different glycosyltransferases with defined specificity. Lipopolysaccharides ("LPS") are provided in gram-negative bacteria only, e.g. *Shigella* spp., *Pseudomonas* spp. and *E. coli* (ExPEC, EHEC).

[0010] The synthesis of LPS starts with the addition of a monosaccharide to the carrier lipid undecaprenyl phosphate ("Und-P-P") at the cytoplasmic side of the membrane. The antigen is built up by sequential addition of monosaccharides from activated sugar nucleotides by different glycosyltransferases, and the lipid-linked polysaccharide is flipped through the membrane by a flippase. The antigen-repeating unit is polymerized by an enzymatic reaction. The polysaccharide is then transferred to the Lipid A by the Ligase WaaL forming the LPS that is exported to the surface, whereas the capsular polysaccharide is released from the carrier lipid after polymerization and exported to the surface. The biosynthetic pathway of these polysaccharides enables the production of LPS bioconjugates *in vivo*, capturing the polysaccharides in the periplasm to a protein carrier.

[0011] Such synthesized complexes of oligo- or polysaccharides (*i.e.*, sugar residues) and proteins (*i.e.*, protein carriers) can be used as conjugate vaccines

to protect against a number of bacterial infections. Conjugate vaccines have been successfully used to protect against bacterial infections. The conjugation of an antigenic polysaccharide to a protein carrier is required for protective memory response, as polysaccharides are T-cell independent immunogens. Polysaccharides have been conjugated to protein carriers by different chemical methods, using activation reactive groups in the polysaccharide as well as the protein carrier.

[0012] Conjugate vaccines can be administered to children to protect against bacterial infections and also can provide a long lasting immune response to adults. Constructs of WO 2009/104074 (Fernandez, *et al.*) have been found to generate an IgG response in animals. It has been found that an IgG response to a *Shigella* O-specific polysaccharide-protein conjugate vaccine in humans correlates with immune protection in humans. (Passwell, J.H. et al., "Safety and Immunogenicity of Improved Shigella O-Specific Polysaccharide-Protein Conjugate Vaccines in Adults in Israel" Infection and Immunity, 69(3):1351-1357 (Mar. 2001).) It is believed that the polysaccharide (*i.e.* sugar residues) triggers a short-term immune response that is sugar-specific. Indeed, the human immune system generates a strong response to specific polysaccharide surface structures of bacteria, such as O-antigens and capsular polysaccharides. However, since the immune response to polysaccharides is IgM dependent, the immune system develops no memory. The protein carrier that carries the polysaccharide triggers an IgG response that is T-cell dependent and that provides long lasting protection since the immune system develops memory.

[0013] *E. coli* O157 is an enterohemorrhagic strain responsible for approximately two-thirds of all recent cases of hemolytic-uremic syndrome and poses serious human health concerns (Law, D. (2000) *J. App. Microbiol.*, 88, 729-745; Wang, L., and Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551).

[0014] *Escherichia coli* strain O157 produces an O-antigen containing the repeating tetrasaccharide unit (4-N-acetyl perosamine→fucose→glucose→GalNAc) (α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc) (Perry, M. B., MacLean, L. and Griffith, D. W. (1986) *Biochem. Cell. Biol.*, 64, 21-28). The tetrasaccharide is preassembled on undecaprenyl pyrophosphate. The *E. coli* cell envelope contains an inner plasma membrane, a stress-bearing peptidoglycan layer and an asymmetric outer membrane consisting of a phospholipid inner monolayer and an outer

monolayer composed of bacterial LPS. LPS contains three components, the lipid A anchor, the 3-deoxy-D-manno-oct-2-ulosonic acid-containing core, and the O-antigen region (see: Raetz, C. R. H. and Whitfield, C. (2002) *Annu. Rev. Biochem.*, 71, 635-700; Whitfield, C. (2006) *Ann. Rev. Biochem.* 75, 39-68; Samuel, G. and Reeves, P. R. (2003) *Carbohydrate Research*, 338, 2503-2519; and refs. therein for reviews on the assembly of O-antigens of bacterial LPS).

[0015] The O-antigen components of bacterial LPS are large, extremely diverse polysaccharides that can be either homopolymeric, composed of a single repeating monosaccharide, or heteropolymeric, containing 10–30 repeats of 3–6 sugar units (Reeves, P. R., Hobbs, M.., Valvano, M. A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C. R. H., and Rick, P. D. (1996) *Trends Microbiol.*, 4, 495-503). O-antigens are, thus, the dominant feature of the bacterial cell surface and constitute important determinants of virulence and pathogenicity (Law, D. (2000) *J. App. Microbiol.*, 88, 729-745; Spears, K. J., Roe, A. J. and Gally, D. L. (2006) *FEMS Microbiol. Lett.*, 255, 187-202; Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. and Wang, L (2008) *FEMS Microbiol. Rev.* 32, 627-653; Stenutz, R., Weintraub, A. and Widmalm, G. (2006) *FEMS Microbiol. Rev.* 30, 382-403). *E. coli* strains with more than 180 individual O-serotypes, attributed to unique O-antigen structures, have been identified (Stenutz, R., Weintraub, A. and Widmalm, G. (2006) *FEMS Microbiol. Rev.* 30, 382-403).

[0016] O-antigen repeat units are pre-assembled on the cytosolic face of the inner membrane attached to undecaprenyl pyrophosphate. The lipid-linked repeat units diffuse transversely (flip-flop) to the periplasmic surface of the inner membrane and are polymerized before transport to the outer membrane and ligation to LPS. Most heteropolymeric O-antigen repeat units have either *N*-acetylglucosamine ("GlcNAc") or *N*-acetylgalactosamine ("GalNAc") at the reducing terminus.

[0017] It had been assumed that the biosynthesis of the lipid intermediates is initiated by the transfer of GlcNAc-P or GalNAc-P from their respective sugar nucleotide derivatives to undecaprenyl monophosphate ("Und-P") catalyzed by WecA (Samuel, G. and Reeves, P. R. (2003) *Carbohydrate Research*, 338, 2503-2519; Alexander, D. C. and Valvano, M. A. (1994) *J. Bacteriol.*, 176, 7079-7084; Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Tolvanen, P. and Skurkik, M.

(1997) *Mol. Microbiol.* 23, 63-76; Amor, P. A. and Whitfield, C. (1997) *Mol. Microbiol.* 26 (145-161); Wang, L. and Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551). Although the properties and specificity of the GlcNAc-phosphotransferase activity of WecA have been characterized (Rush, J. S., Rick, P. D. and Waechter, C. J. (1997) Glycobiology, 7, 315-322), the conclusion that WecA catalyzes the synthesis of GalNAc-P-P-Und was based on genetic studies (Wang, L. and Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551). Such earlier genetic studies indicated that the biosynthesis of the lipid-linked tetrasaccharide intermediate was initiated by the enzymatic transfer of GalNAc-P from UDP-GalNAc to Und-P catalyzed by WecA (Wang, L. and Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551). However, there was no direct enzymological evidence demonstrating that WecA utilizes UDP-GalNAc as a GalNAc-P donor.

[0018] Furthermore, the *E. coli* O55 *gne* and *gne1* genes were previously proposed to encode a UDP-GlcNAc 4-epimerase (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) *J.* Bacteriol. 184, 2620-2625; Guo, H., Yi, W., Li, L. and Wang, P. G. (2007) *Biochem. Biophys. Res. Commun.*, 356, 604-609). Previous reports identified two genes from *E. coli* O55 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) *J.* Bacteriol. 184, 2620-2625) and *E. coli* O86 (Guo, H., Yi, W., Li, L. and Wang, P. G. (2007) *Biochem. Biophys. Res. Commun.*, 356, 604-609), *E. coli* O55 *gne* and *E. coli* O86 *gne1*, respectively, that are 100% identical to a *Z3206* gene within the same gene family.

[0019] Accordingly, one of skill would have been led to believe that the *Z3206* gene also encodes a UDP-GlcNAc/UDP-GalNAc epimerase.

BRIEF SUMMARY OF THE INVENTION

[0020] It has now been surprisingly discovered that an epimerase encoded by the *Z3206* gene in *E. coli* O157 catalyzes a reaction that synthesizes *N*-acetylgalactosamine ("GalNAc") on undecaprenyl pyrophosphate, which initiates the formation of an oligo- or polysaccharide.

[0021] In one aspect, the present invention relates to a recombinant prokaryotic biosynthetic system that produces all or a portion of a polysaccharide comprising an epimerase that synthesizes GalNAc on undecaprenyl pyrophosphate.

The invention further includes glycosyltransferases that synthesize all or a portion of a polysaccharide having GalNAc at the reducing terminus, and still further includes glycosyltransferases that synthesize all or a portion of an antigenic polysaccharide having GalNAc at the reducing terminus.

[0022] In another aspect, the invention is directed to an epimerase to produce GalNAc on undecaprenyl pyrophosphate, and, in a further aspect, the epimerase is encoded by the *Z3206* gene.

[0023] In an additional aspect, the present invention is directed to an expression system for producing an *N*-glycosylated protein comprising: a nucleotide sequence encoding an oligosaccharyl transferase; a nucleotide sequence encoding a protein carrier; at least one oligo- or polysaccharide gene cluster from at least one bacterium, wherein the polysaccharide contains GalNAc at the reducing terminus; and a nucleic acid sequence encoding an epimerase.

[0024] In a still further aspect, the instant invention is directed to a recombinant prokaryotic biosynthetic system comprising *Z3206* gene which encodes an epimerase that converts GlcNAc-P-P-Und to GalNAc-P-P-Und.

[0025] In yet an additional aspect, the present invention is directed to a recombinant prokaryotic biosynthetic system comprising *E. coli* O55 *gne* gene or *E. coli* O86 *gne1* gene which encodes an epimerase that converts GlcNAc-P-P-Und to GalNAc-P-P-Und.

[0026] In yet another aspect, the present invention relates to an *N*-glycosylated protein comprising at least one introduced consensus sequence, D/E – X – N – Z – S/T, wherein X and Z can be any natural amino acid except proline, and a glycan having *N*-acetylgalactosamine at the reducing terminus.

[0027] In still another aspect, the present invention is directed to a bioconjugate vaccine comprising an *N*-glycosylated protein having at least one introduced consensus sequence, D/E – X – N – Z – S/T, wherein X and Z can be any natural amino acid except proline; an immunogenic glycan having *N*-acetylgalactosamine at the reducing terminus; and an adjuvant.

[0028] In an addition aspect, the invention relates to method for producing an *N*-linked glycosylated protein in a host cell comprising nucleic acids encoding: glycosyltransferases that assemble at least one oligo- or polysaccharide from at least

one bacterium containing GalNAc at the reducing terminus; a protein carrier; an oligosaccharyl transferase; and an epimerase.

[0029] In a further aspect, the present invention relates to the use of a biosynthetic system and proteins for preparing a bioconjugate vaccine.

[0030] In an additional aspect, the present invention is directed to methods for producing mono-, oligo- and polysaccharides, and in a still further aspect the invention directed to methods for producing antigenic glycans and *N*-glycosylated proteins.

BRIEF DESCRIPTION O F THE DRAWINGS

[0031] **FIG. 1** shows the time course of [^{3}H]GlcNAc/GalNAc-P-P-Und synthesis by membrane fractions from *E. coli* O157. The membrane fraction from *E. coli* strain O157 was incubated with UDP-[^{3}H]GlcNAc for the indicated times at 37°C. The [^{3}H]lipid products were extracted and the incorporation of [^{3}H]GlcNAc into [^{3}H]GlcNAc-P-P-Und (O) and [^{3}H]GalNAc-P-P-Und (●) was assayed as described in Example 2.

[0032] **FIG. 2** shows the proposed biosynthetic pathway for the formation of GalNAc-P-P-Und from GlcNAc-P-P-Und.

[0033] **FIG. 3** shows purification and characterization of [^{3}H]GalNAc-P-P-Und synthesized by membrane fractions from *E. coli* strain O157. Membrane fractions from *E. coli* O157 were incubated with UDP-[^{3}H]GlcNAc, and the [^{3}H]GalNAc lipids were purified as described in Example 3. FIG. 3A, preparative thin layer chromatogram of [^{3}H]HexNAc lipids on borate-impregnated silica gel G (Quantum 1) after purification on DEAE-cellulose is shown. FIG. 3B, thin layer chromatography of purified [^{3}H]GalNAc-P-P-Und on borate-impregnated silica gel G (Baker, Si250) after recovery from the preparative plate in panel A is shown. FIG. 3C, descending paper chromatogram (borate-impregnated Whatman No. 1 paper) of the [^{3}H]-amino sugar recovered after mild acid hydrolysis of [^{3}H]GalNAc-P-P-Und purified in FIG. 3B is shown. FIG. 3D, descending paper chromatogram (Whatman No. 3MM) of the [^{3}H]HexNAc-alditol produced by reduction of the [^{3}H] amino sugar from FIG. 3C with $NaBH_4$.

[0034] **FIG. 4** shows metabolic labeling of *E. coli* 21546 cells and *E. coli* 21546 cells after transformation with pMLBAD:Z3206. *E. coli* 21546 (FIG. 4A) and *E. coli* 21546:pMLBAD/Z3206 (FIG. 4B) were labeled metabolically with [^{3}H]GlcNAc for 5 min at 37°C. [^{3}H]GlcNAc/GalNAc-P-P-Und were extracted, freed of water soluble contaminants and separated by thin layer chromatography on borate-impregnated silica gel plates (Baker Si250) as described in Example 3. Radioactive lipids were detected using a Bioscan chromatoscanner. The chromatographic positions of GalNAc-P-P-Und and GlcNAc-P-P-Und are indicated by arrows.

[0035] **FIG. 5** shows thin layer chromatography of [^{3}H]GlcNAc/GalNAc-P-P-Und formed by incubation of membrane fractions from *E. coli* strains with UDP-[^{3}H]GlcNAc. Membrane fractions from *E. coli* strains K12 (FIG. 5A), O157 (FIG. 5B), 21546 (FIG. 5C), and 21546:pMLBAD/Z3206 (FIG. 5D) were incubated with UDP-[^{3}H]GlcNAc for 10 min at 37°C, and the [^{3}H]lipid products were extracted, freed of water-soluble contaminants by partitioning, and separated by thin layer chromatography on borate-impregnated silica gel plates (Baker Si250) as described in Example 3. The chromatographic positions of GalNAc-P-P-Und and GlcNAc-P-P-Und are indicated by arrows.

[0036] **FIG. 6** shows discharge of GlcNAc-P by incubation with UMP. Membrane fractions from *E. coli* 21546:Z3206 were preincubated with UDP-[^{3}H]GlcNAc to enzymatically label GlcNAc-P-P-Und for 10 min (FIG. 6A) at 37°C followed by a second incubation period with 1 mM UMP included for either 1 min (FIG. 6B) or 2 min (FIG. 6C). After the indicated incubation periods [^{3}H]GlcNAc/GalNAc-P-P-Und were extracted and resolved by thin layer chromatography on borate-impregnated silica gel plates (Baker Si250) as described in Example 3. The chromatographic positions of GalNAc-P-P-Und and GlcNAc-P-P-Und are indicated by arrows.

[0037] **FIG. 7** shows conversion of exogenous [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und to the pertinent [^{3}H]HexNAc-P-P-Und product catalyzed by membranes from strain 21546 expressing Z3206. Membrane fractions from *E. coli* strain 21546 (FIG. 7B and FIG. 7E) and 21546:pMLBAD/Z3206 (FIG. 7C and FIG. 7F) were incubated with purified [^{3}H]GlcNAc-P-P-Und (FIG. 7A, FIG. 7B, and FIG. 7C) or [^{3}H]GalNAc-P-P-Und (panels at FIG. 7D, FIG. 7E, and FIG. 7F) (dispersed

ultrasonically in 1% Triton X-100) for 1 min at 37°C. [^{3}H]GlcNAc/GalNAc-P-P-Und were extracted, resolved by thin layer chromatography on borate-impregnated silica gel plates (Baker Si250) and detected with a Bioscan AR2000 radiochromatoscanner as described in Example 3.

[0038] **FIG. 8** shows SDS-PAGE analysis of unglycosylated and glycosylated AcrA protein. Periplasmic extracts prepared from *E. coli* DH5α cells carrying the AcrA expression plasmid and the pgl operon Δgne complemented with pMLBAD:Z3206 (lane 1), pMLBAD:gne (lane 2) or the vector control pMLBAD (lane 3) were separated by 10% SDS-PAGE and transferred to nitrocellulose membranes. AcrA and its glycosylated forms were detected with anti AcrA antisera. The position of bands corresponding to unglycosylated (AcrA) and glycosylated AcrA (gAcrA) is indicated.

[0039] **FIG. 9** shows the genes that have been identified by Liu B *et al.* (*Structure and genetics of Shigella O antigens* FEMS Microbiology Review, 2008. 32: p. 27).

[0040] **FIG. 10** is a scheme showing the DNA region containing the genes required to synthesize the *S. flexneri* 6 O antigen.

[0041] **FIG. 11** shows expression of the *S. flexneri* 6 O antigen in *E. coli*. LPS was visualized by either silver staining or by transfer to nitrocellulose membranes and detection by antibodies directed against *S. flexneri* 6.

[0042] **FIG. 12** shows HPLC of O antigen. LLO analysis of *E. coli* cells (SCM3) containing *S. flexneri* – Z3206, *E. coli* cells (SCM3) containing *S. flexneri* + Z3206 or empty *E. coli* (SCM3) cells.

[0043] **FIG. 13** shows Western blot of Nickel purified proteins from *E. coli* cells expressing EPA, pglB and *S. flexneri* 6 O-antigen +/- Z3206

DETAILED DESCRIPTION OF THE INVENTION

[0044] The present invention encompasses a recombinant prokaryotic biosynthetic system comprising nucleic acids encoding an epimerase that synthesizes an oligo- or polysaccharide having *N*-acetylgalactosamine at the reducing terminus, and *N*-glycosylated proteins having *N*-acetylgalactosamine at the reducing terminus of the glycan.

[0045] The term "partial amino acid sequence(s)" is also referred to as "optimized consensus sequence(s)" or "consensus sequence(s)." The optimized consensus sequence is *N*-glycosylated by an oligosaccharyl transferase ("OST," "OTase"), much more efficiently than the regular consensus sequence "N - X - S/T."

[0046] In general, the term "recombinant *N*-glycosylated protein" refers to any poly- or oligopeptide produced in a host cell that does not naturally comprise the nucleic acid encoding said protein. In the context of the present invention, this term refers to a protein produced recombinantly in a prokaryotic host cell, for example, *Escherichia* spp., *Campylobacter* spp., *Salmonella* spp., *Shigella* spp., *Helicobacter* spp., *Pseudomonas* spp., *Bacillus* spp., and in further embodiments *Escherichia coli, Campylobacter jejuni, Salmonella typhimurium* etc., wherein the nucleic acid encoding said protein has been introduced into said host cell and wherein the encoded protein is *N*-glycosylated by the OTase, said transferase enzyme naturally occurring in or being introduced recombinantly into said host cell.

[0047] In accordance with the internationally accepted one letter code for amino acids the abbreviations D, E, N, S and T denote aspartic acid, glutamic acid, asparagine, serine, and threonine, respectively.

[0048] Proteins according to the invention comprise one or more of an optimized consensus sequence(s) D/E - X - N - Z - S/T that is/are introduced into the protein and *N*-glycosylated. Hence, the proteins of the present invention differ from the naturally occurring *C. jejuni N*-glycoproteins which also contain the optimized consensus sequence but do not comprise any additional (introduced) optimized consensus sequences.

[0049] The introduction of the optimized consensus sequence can be accomplished by the addition, deletion and/or substitution of one or more amino acids. The addition, deletion and/or substitution of one or more amino acids for the

purpose of introducing the optimized consensus sequence can be accomplished by chemical synthetic strategies, which, in view of the instant invention, would be well known to those skilled in the art such as solid phase-assisted chemical peptide synthesis. Alternatively, and preferred for larger polypeptides, the proteins of the present invention can be prepared by recombinant techniques that would be art-standard techniques in light of the invention.

[0050] The proteins of the present invention have the advantage that they may be produced with high efficiency and in any host. In one embodiment of the invention, the host comprises a functional *pgl* operon from *Campylobacter* spp., for example, from *C. jejuni*. In further embodiments, oligosaccharyl transferases from *Campylobacter* spp. for practicing the invention are from *Campylobacter coli* or *Campylobacter lari*. In view of the invention, oligosaccharyl transferases would be apparent to one of skill in the art. For example, oligosaccharyl transferases are disclosed in references such as Szymanski, C.M. and Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens, Nat. Rev. Microbiol. 3:225-237. The functional *pgl* operon may be present naturally when said prokaryotic host is *Campylobacter* spp., or, for example, *C. jejuni*. However, as demonstrated before in the art and mentioned above, the *pgl* operon can be transferred into cells and remain functional in said new cellular environment.

[0051] The term "functional *pgl* operon from *Campylobacter* spp., preferably *C. jejuni*" is meant to refer to the cluster of nucleic acids encoding the functional oligosaccharyl transferase (OTase) of *Campylobacter* spp., for example, *C. jejuni*, and one or more specific glycosyltransferases capable of assembling an oligosaccharide on a lipid carrier, and wherein said oligosaccharide can be transferred from the lipid carrier to the target protein having one or more optimized amino acid sequence(s): D/E - X N - Z - S/T by the OTase. It to be understood that the term "functional *pgl* operon from *Campylobacter* spp., preferably *C. jejuni*" in the context of this invention does not necessarily refer to an operon as a singular transcriptional unit. The term merely requires the presence of the functional components for *N*-glycosylation of the recombinant protein in one host cell. These components may be transcribed as one or more separate mRNAs and may be regulated together or separately. For example, the term also encompasses functional components positioned in genomic DNA and plasmid(s) in one host cell. For the purpose of

efficiency, in one embodiment all components of the functional *pgl* operon are regulated and expressed simultaneously.

[0052] The oligosaccharyl transferase can originate, in some embodiments, from *Campylobacter* spp., and in other embodiments, from *C. jejuni*. In additional embodiments, the oligosaccharyl transferase can originate from other organisms which are known to those of skill in the art as having an oligosaccharyl transferase, such as, for example, *Wolinella* spp. and eukaryotic organisms.

[0053] The one or more specific glycosyltransferases capable of assembling an oligosaccharide on a lipid carrier may originate from the host cell or be introduced recombinantly into said host cell, the only functional limitation being that the oligosaccharide assembled by said glycosyltransferases can be transferred from the lipid carrier to the target protein having one or more optimized consensus sequences by the OTase. Hence, the selection of the host cell comprising specific glycosyltransferases naturally and/or replacing specific glycosyltransferases naturally present in said host as well as the introduction of heterologous specific glycosyltransferases will enable those skilled in the art to vary the *N*-glycans bound to the optimized *N*-glycosylation consensus site in the proteins of the present invention.

[0054] As a result of the above, the present invention provides for the individual design of *N*-glycan-patterns on the proteins of the present invention. The proteins can therefore be individualized in their *N*-glycan pattern to suit biological, pharmaceutical and purification needs.

[0055] In embodiments of the present invention, the proteins may comprise one but also more than one, such as at least two, at least 3 or at least 5 of said *N*-glycosylated optimized amino acid sequences.

[0056] The presence of one or more *N*-glycosylated optimized amino acid sequence(s) in the proteins of the present invention can be of advantage for increasing their immunogenicity, increasing their stability, affecting their biological activity, prolonging their biological half-life and/or simplifying their purification.

[0057] The optimized consensus sequence may include any amino acid except proline in position(s) X and Z. The term "any amino acids" is meant to encompass common and rare natural amino acids as well as synthetic amino acid derivatives and analogs that will still allow the optimized consensus sequence to be *N*-

glycosylated by the OTase. Naturally occurring common and rare amino acids are preferred for X and Z. X and Z may be the same or different.

[0058] It is noted that X and Z may differ for each optimized consensus sequence in a protein according to the present invention.

[0059] The *N*-glycan bound to the optimized consensus sequence will be determined by the specific glycosyltransferases and their interaction when assembling the oligosaccharide on a lipid carrier for transfer by the OTase. In view of the instant invention, those skilled in the art would be able to design the *N*-glycan by varying the type(s) and amount of the specific glycosyltransferases present in the desired host cell.

[0060] "Monosaccharide" as used herein refers to one sugar residue. "Oligo- and polysaccharide" refer to two or more sugar residues. The term "glycans" as used herein refers to mono-, oligo- or polysaccharides. "*N*-glycans" are defined herein as mono-, oligo- or polysaccharides of variable compositions that are linked to an ε-amide nitrogen of an asparagine residue in a protein via an *N*-glycosidic linkage. In an embodiment, the *N*-glycans transferred by the OTase are assembled on an undecaprenol pyrophosphate ("Und-P-P") lipid-anchor that is present in the cytoplasmic membrane of gram-negative or positive bacteria. They are involved in the synthesis of O antigen, O polysaccharide and peptidoglycan (Bugg, T. D., and Brandish, P. E. (1994). From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis. FEMS Microbiol Lett *119*, 255-262; Valvano, M. A. (2003). Export of O-specific lipopolysaccharide. Front Biosci *8*, s452-471).

[0061] Studies were conducted to determine whether the biosynthesis of a lipid-linked repeating tetrasaccharide (4-*N*-acetyl perosamine→fucose→glucose→GalNAc) was initiated by the formation of GalNAc-P-P-Und by WecA. When membrane fractions from *E. coli* strains K12, O157, and PR4019, a WecA-overexpressing strain, were incubated with UDP-[^{3}H]GalNAc, neither the enzymatic synthesis of [^{3}H]GlcNAc-P-P-Und nor [^{3}H]GalNAc-P-P-Und was detected. However, when membrane fractions from strain O157 were incubated with UDP-[^{3}H]GlcNAc, two enzymatically labeled products were observed with the chemical and chromatographic properties of [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-

P-P-Und, confirming that strain O157 contained an epimerase capable of interconverting GlcNAc-P-P-Und and GalNAc-P-P-Und. The presence of an epimerase was also confirmed by showing that exogenous [^{3}H]GlcNAc-P-P-Und was converted to [^{3}H]GalNAc-P-P-Und when incubated with membranes from strain O157. When strain O157 was metabolically labeled with [^{3}H]GlcNAc, both [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und were detected. Transformation of *E. coli* strain 21546 with the *Z3206* gene enabled these cells to synthesize GalNAc-P-P-Und *in vivo* and *in vitro*. The reversibility of the epimerase reaction was demonstrated by showing that [^{3}H]GlcNAc-P-P-Und was reformed when membranes from strain O157 were incubated with exogenous [^{3}H]GalNAc-P-P-Und. The inability of Z3206 to complement the loss of the *gne* gene in the expression of the *Campylobacter jejuni* N-glycosylation system in *E. coli* indicated that it does not function as a UDP-GlcNAc/UDP-GalNAc epimerase. Based on these results, it was confirmed that GalNAc-P-P-Und is synthesized reversibly by a GlcNAc-P-P-Und epimerase following the formation of GlcNAc-P-P-Und by WecA in *E. coli* O157.

[0062] The initiating reaction of *E. coli* O157 O-antigen subunit assembly was investigated to confirm that GalNAc-P-P-Und synthesis is catalyzed by some previously unknown mechanism rather than by WecA. The evidence presented herein shows that GalNAc-P-P-Und is not synthesized by GalNAc-P transfer from UDP-GalNAc catalyzed by WecA but rather by the reversible epimerization of the 4-OH of GlcNAc-P-P-Und catalyzed by an epimerase encoded by the *Z3206* gene in *E. coli* O157.

[0063] Accordingly, the invention encompasses a novel biosynthetic pathway for the assembly of an important bacterial cell surface component as well as a new biosynthetic route for the synthesis of GalNAc-P-P-Und. A further embodiment of the invention includes the bacterial epimerase as a new target for antimicrobial agents.

[0064] *E. coli* O157 synthesizes an O-antigen with the repeating tetrasaccharide structure (4-*N*-acetyl perosamine→fucose→glucose→GalNAc). It is shown herein that the biosynthesis of the lipid-linked tetrasaccharide intermediate was not initiated by the enzymatic transfer of GalNAc-P from UDP-GalNAc to Und-P catalyzed by WecA, contrary to earlier genetic studies (Wang, L. and Reeves, P. R. (1998) *Infect. Immun.* 66, 3545-3551). The invention described herein, obtained by

homology searches and then confirmed by results from genetic, enzymology, and metabolic labeling experiments, demonstrates that WecA does not utilize UDP-GalNAc as a substrate, but that WecA is required to synthesize GlcNAc-P-P-Und which is then reversibly converted to GalNAc-P-P-Und by an epimerase encoded by the *Z3206* gene in strain O157.

[0065] The *Z3206* gene of the present invention belongs to a family of genes present in several strains that produce surface O-antigen repeat units containing GalNAc residues at their reducing termini (Table 1). The *Z3206* gene sequence is shown in SEQ ID NO: 1. Previous reports identified two genes from *E. coli* O55 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) *J.* Bacteriol. 184, 2620-2625) and *E. coli* O86 (Guo, H., Yi, W., Li, L. and Wang, P. G. (2007) *Biochem. Biophys. Res. Comm.*, 356, 604-609), *E. coli* O55 *gne* and *E. coli* O86 *gne1*, respectively, that are 100% identical to a *Z3206* gene (Table 1). The *E. coli* O55 *gne* gene sequence is shown as SEQ ID NO: 3, and *E. coli* O86 *gne1* gene sequence is shown as SEQ ID NO: 5.

TABLE 1

Correlation of *Z3206* gene in bacterial strains expressing O-antigen chains with GalNAc at the reducing termini.

	% Identity with Z3206	GalNAc at the reducing terminus of O-antigen repeat unit
E. coli O55 *gne* (SEQ ID NO: 3)	100	Yes
E. coli O86 *gne1* (SEQ ID NO: 5)	100	Yes
Shigella boydii O18 *gne* (SEQ ID NO: 7)	88	Yes
Salmonella enterica O30 *gne* (SEQ ID NO: 9)	94	Yes
C. jejuni gne (SEQ ID NO: 11)	21	No
E. coli K12 *galE* (SEQ ID NO: 13)	27	No
E. coli O86 *gne2* (SEQ ID NO: 15)	18	Yes

[0066] Accordingly, we conclude that *E. coli* O55 *gne* and *E. coli* O86 *gne1* also encode epimerases capable of converting GlcNAc-P-P-Und to GalNAc-P-

P-Und in strains O55 and O86, respectively, which also produce O-antigen repeat units with GalNAc at the reducing termini (Table 1).

[0067] Two experimental approaches in this study indicate that the Z3206 protein does not catalyze the epimerization of UDP-GlcNAc to UDP-GalNAc in strain O157. First, when membranes from strain O157 were incubated with [^{3}H]UDP-GalNAc, neither [^{3}H]GlcNAc-P-P-Und nor [^{3}H]GalNAc-P-P-Und was detected (Table 3). If Z3206 catalyzed the conversion of [^{3}H]UDP-GalNAc to [^{3}H]UDP-GlcNAc, it would be expected that [^{3}H]GlcNAc-P-P-Und should be observed. Second, we have shown that hemagglutinin-tagged Z3206 was incapable of complementing the UDP-GalNAc-dependent *C. jejuni N*-glycosylation reporter system (FIG. 8).

[0068] *E. coli* O55 *gne* gene from strain O55 (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) *J.* Bacteriol. 184, 2620-2625) was also assayed for epimerase activity by incubating crude extracts with UDP-GalNAc and indirectly assaying the conversion to UDP-GlcNAc by measuring an increase in reactivity with p-dimethylaminobenzaldehyde after acid hydrolysis. In both studies, the formation of the product was based on changes in reactivity with p-dimethylaminobenzaldehyde, and not a definitive characterization of the sugar nucleotide end product. A 90% pure polyhistidine-tagged *E. coli* O86 *gne1* was also shown to have a low level of UDP-glucose epimerase activity relative to Gne2 in a coupled assay.

[0069] Accordingly, an embodiment of the invention is directed to a recombinant prokaryotic biosynthetic system containing *Z3206* gene, *E. coli* O55 *gne* gene or *E. coli* O86 *gne1* gene that converts GlcNAc-P-P-Und to GalNAc-P-P-Und.

[0070] It is significant that *E. coli* O86, which synthesizes an O-antigen containing two GalNAc residues, which would presumably require UDP-GalNAc as the glycosyl donor for the additional, non-reducing terminal GalNAc, also possesses an additional GlcNAc 4-epimerase gene, termed *gne2*, within the O-antigen gene cluster (Guo, H., Yi, W., Li, L. and Wang, P. G. (2007) *Biochem. Biophys. Res. Commun.*, 356, 604-609). This additional epimerase gene has high homology with the *galE* gene of the colanic acid gene cluster and appears to be a UDP-GlcNAc 4-epimerase capable of synthesizing UDP-GalNAc.

[0071] The *Z3206* gene appears to be highly conserved in *E. coli* O-serotypes initiated with GalNAc. In a recent study, 62 *E. coli* strains, with established O-antigen repeat unit structures, were screened for expression of *Z3206* by a polymerase chain reaction based method using nucleotide primers designed to specifically detect the *E. coli* O157 *Z3206* gene (Wang, L., Huskic, S., Cisterne, A., Rothemund, D. and Reeves, P.R. (2002) *J.* Bacteriol. 184, 2620-2625). In this study *Z3206* was detected in 16 of the 22 *E. coli* strains that were known to contain GalNAc, and in only 4 of the 40 strains lacking GalNAc. Moreover, a similar screen of the 22 GalNAc-containing strains with primers designed to detect an alternative epimerase with UDP-GlcNAc 4-epimerase activity (the *GalE* gene of *E. coli* O113) detected no strains carrying this gene, indicating that *Z3206* is the GlcNAc 4-epimerase gene most commonly associated with the presence of a reducing-terminal GalNAc in O-antigen repeat units of *E. coli*.

[0072] Analysis of the Z3206 protein sequence by a variety of web-based topological prediction algorithms indicates that the Z3206 protein is not highly hydrophobic. The majority of the topological prediction algorithms indicate that Z3206 is a soluble 37 kDa protein, although TMPred (Hofmann, K., and Stoffel, W. (1993) *Biol. Chem. Hoppe-Seyler* 374, 166 (abstr.)) predicted a single weak N-terminal transmembrane helix. However, Western blotting after SDS-PAGE of cellular fractions from *E. coli* cells expressing hemagglutinin-tagged Z3206 clearly shows that the tagged protein is associated with the particulate fraction following hypotonic lysis of the cells. Preliminary experiments show that the protein remains associated with the particulate fraction following incubation of the membrane fraction with 1 M KCl, but is solubilized in an active form by incubation with 0.1% Triton X-100.

[0073] *E. coli* O157 Z3206 has significant sequence homology with the short-chain dehydrogenase/reductase family of oxido-reductases including the G*X*X*G*X*X*G motif (Rossman fold), consistent with the NAD(P) binding pocket (Allard, S. T. M., Giraud, M. F., and Naismith, J. H. (2001) *Cell. Mol. Life Sci.* 58, 1650-1655) and the conserved S X_{24}YX_3K sequence, involved in proton abstraction and donation (Field, R. A. and Naismith, J. H. (2003) *Biochemistry* 42, 7637-7647). Molecular modeling based on crystal structures of UDP-Glc 4-epimerase, another member of the short-chain dehydrogenase/reductase family, suggests that, after

hydride abstraction, the 4-keto intermediate rotates around the β phosphate of UDP to present the opposite face of the keto intermediate and allow re-insertion of hydride from the opposite side, thus inverting the configuration of the hydroxyl at carbon 4. The presence of these conserved sequences suggests that Z3206 likely functions *via* a similar mechanism. Although the equilibrium distribution of the epimerase products, seen in FIG. 7, seems to favor the formation of GlcNAc-P-P-Und, the utilization of GalNAc-P-P-Und for O-antigen repeat unit assembly would drive the epimerization reaction in the direction of GalNAc-P-P-Und by mass action.

[0074] Epimerization of the glycosyl moieties of polyisoprenoid lipid intermediates has not been widely reported in nature. In one previous study the 2-epimerization of ribosyl-P-decaprenol to form arabinosyl-P-decaprenol, an arabinosyl donor in arabinogalactan biosynthesis in mycobacteria, was reported (Mikusová, K., Huang, H., Yagi, T., Holsters, M., Vereecke, D., D'Haeze, W., Scherman, M. S., Brennan, P. J., McNeil, M. R., and Crick, D. C. (2005) *J. Bacteriol.* 187, 8020-8025). Arabinosyl-P-decaprenol is formed *via* a two-step oxidation/reduction reaction requiring two mycobacterial proteins, Rv3790 and Rv3791. Although epimerization was modestly stimulated by the addition of NAD and NADP, neither Rv3790 nor Rv3791 contain either the Rossman fold or the $SX_{24}YXXXK$ motif, characteristic of the short-chain dehydrogenase/reductase family (Allard, S. T. M., Giraud, M.-F. and Naismith, J. H. (2001) *Cell. Mol. Life Sci.* 58, 1650-1655; Field, R. A. and Naismith, J. H. (2003) *Biochemistry* 42, 7637-7647).

[0075] In summary, a novel biosynthetic pathway for the formation of GalNAc-P-P-Und by the epimerization of GlcNAc-P-P-Und, is described.

[0076] Several antibiotics have been shown to inhibit the synthesis of GlcNAc-P-P-Und, but are limited in their utility because they also block the synthesis of GlcNAc-P-P-dolichol, the initiating dolichol-linked intermediate of the protein *N*-glycosylation pathway. Although GlcNAc-P-P-dolichol is a structurally related mammalian counterpart of the bacterial glycolipid intermediate, GlcNAc-P-P-Und, there is no evidence for a similar epimerization reaction converting GlcNAc-P-P-dolichol to GalNAc-P-P-dolichol in eukaryotic cells. Thus, this raises the possibility that in strains where the surface O-antigen containing GalNAc at the reducing termini are involved in a pathological process, O-antigen synthesis could potentially be blocked by inhibiting the bacterial epimerases.

[0077] An embodiment of the present invention involves an epimerase that converts GlcNAc-P-P-Und (*N*-acetylglucosaminylpyrophosphorylundecaprenol) to GalNAc-P-P-Und (*N*-acetylgalactosaminylpyrophosphorylundecaprenol) in *E. coli O157*. A still further exemplary aspect of the invention involves the initiation of synthesis of lipid-bound repeating tetrasaccharide having GalNAc at the reducing terminus.

[0078] The basis of another aspect of the invention includes the discovery that *Campylobacter jejuni* contains a general *N*-linked protein glycosylation system. Various proteins of *C. jejuni* have been shown to be modified by a heptasaccharide. This heptasaccharide is assembled on undecaprenyl pyrophosphate, the carrier lipid, at the cytoplasmic side of the inner membrane by the stepwise addition of nucleotide activated monosaccharides catalyzed by specific glycosyltransferases. The lipid-linked oligosaccharide then flip-flops (diffuses transversely) into the periplasmic space by a flipppase, *e.g.*, PglK. In the final step of *N*-linked protein glycosylation, the oligosaccharyltransferase (e.g., PglB) catalyzes the transfer of the oligosaccharide from the carrier lipid to asparagine (Asn) residues within the consensus sequence D/E – X – N – Z – S/T, where the X and Z can be any amino acid except Pro. The glycosylation cluster for the heptasaccharide had been successfully transferred into *E. coli* and *N*-linked glycoproteins of *Campylobacter* had been produced.

[0079] It had been demonstrated that PglB does not have a strict specificity for the lipid-linked sugar substrate. The antigenic polysaccharides assembled on undecaprenyl pyrophosphate are captured by PglB in the periplasm and transferred to a protein carrier (Feldman, 2005; Wacker, M., *et al.*, Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proc Natl Acad Sci U S A, 2006. 103(18): p. 7088-93.) The enzyme will also transfer a diverse array of undecaprenyl pyrophosphate (UPP) linked oligosaccharides if they contain an *N*-acetylated hexosamine at the reducing terminus. The nucleotide sequence for *pglB* and the amino acid sequence for PglB are published at WO2009/104074.

[0080] Accordingly, one embodiment of the invention involves a recombinant N-glycosylated protein comprising: one or more of an introduced consensus sequence, D/E – X – N – Z – S/T, wherein X and Z can be any natural amino acid except proline; and an oligo- or polysaccharide having *N*-

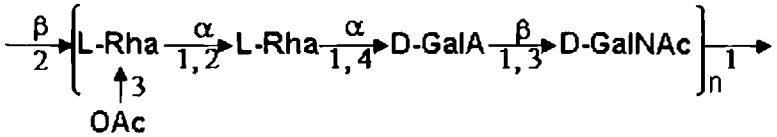
acetylgalactosamine at the reducing terminus and *N*-linked to each of said one or more introduced consensus sequences by an *N*-glycosidic linkage.

[0081] In a further embodiment, the present invention is directed to a recombinant prokaryotic biosynthetic system for producing all or a portion of a polysaccharide comprising an epimerase that synthesizes *N*-acetylgalactosamine ("GalNAc") on undecaprenyl pyrophosphate. In a further embodiment, all or a portion of the polysaccharide is antigenic.

[0082] In another embodiment, the present invention is directed to a recombinant prokaryotic biosynthetic system comprising: an epimerase that synthesizes GalNAc on undecaprenyl pyrophosphate; and glycosyltransferases that synthesize a polysaccharide having GalNAc at the reducing terminus.

[0083] An embodiment of the invention further comprises a recombinant prokaryotic biosynthetic system comprising an epimerase that synthesizes GalNAc on undecaprenyl pyrophosphate and glycosyltransferases that synthesize a polysaccharide, wherein said polysaccharide has the following structure: α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc; and wherein GalNAc is at the reducing terminus of said polysaccharide.

[0084] The recombinant prokaryotic biosynthetic system can produce mono-, oligo- or polysaccharides of various origins. Embodiments of the invention are directed to oligo- and polysaccharides of various origins. Such oligo- and polysaccharides can be of prokaryotic or eukaryotic origin. Oligo- or polysaccharides of prokaryotic origin may be from gram-negative or gram-positive bacteria. In one embodiment of the invention, the oligo- or polysaccharide is from *E. coli*. In a further aspect of the invention, said oligo- or polysaccharide is from *E. coli* O157. In another embodiment, said oligo- or polysaccharide comprises the following structure: α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc. In a further embodiment of the invention, the oligo- or polysaccharide is from *Shigella flexneri*. In a still further embodiment, the oligo- or polysaccharide is from *Shigella flexneri* 6. In a still further aspect, said oligo- or polysaccharide comprises the following structure:



[0085] Embodiments of the invention further include proteins of various origins. Such proteins include proteins native to prokaryotic and eukaryotic organisms. The protein carrier can be, for example, AcrA or a protein carrier that has been modified to contain the consensus sequence for protein glycosylation, *i.e.*, D/E – X – N – Z – S/T, wherein X and Z can be any amino acid except proline (*e.g.*, a modified Exotoxin *Pseudomonas aeruginosa* ("EPA")). In one embodiment of the invention, the protein is *Pseudomonas aeruginosa* EPA.

[0086] A further aspect of the invention involves novel bioconjugate vaccines having GalNAc at the reducing terminus of the *N*-glycan. An additional embodiment of the invention involves a novel approach for producing such bioconjugate vaccines that uses recombinant bacterial cells that contain an epimerase which produces GalNAc on undecaprenyl pyrophosphate. In one embodiment, bioconjugate vaccines can be used to treat or prevent bacterial diseases. In further embodiments, bioconjugate vaccines may have therapeutic and/or prophylactic potential for cancer or other diseases.

[0087] A typical vaccination dosage for humans is about 1 to 25 μg, preferably about 1 μg to about 10μg, most preferably about 10 μg. Optionally, a vaccine, such as a bioconjugate vaccine of the present invention, includes an adjuvant.

[0088] In an additional embodiment, the present invention is directed to an expression system for producing a bioconjugate vaccine against at least one bacterium comprising: a nucleotide sequence encoding an oligosaccharyl transferase; a nucleotide sequence encoding a protein carrier; at least one polysaccharide gene cluster from the at least one bacterium, wherein the polysaccharide contains GalNAc at the reducing terminus; and a nucleic acid sequence encoding an epimerase. In a further embodiment, the polysaccharide gene cluster encodes an antigenic polysaccharide.

[0089] In still a further embodiment, the present invention is directed to an expression system for producing a bioconjugate vaccine against at least one bacterium comprising: a nucleotide sequence encoding an oligosaccharyl transferase; a nucleotide sequence encoding a protein carrier comprising at least one inserted consensus sequence, D/E – X – N – Z – S/T, wherein X and Z may be any natural

amino acid except proline; at least one polysaccharide gene cluster from the at least one bacterium, wherein the polysaccharide contains GalNAc at the reducing terminus; and the *Z3206* gene. In a further embodiment, the polysaccharide gene cluster encodes an antigenic polysaccharide.

[0090] In yet another embodiment, the present invention is directed to a bioconjugate vaccine comprising: a protein carrier; at least one immunogenic polysaccharide chain linked to the protein carrier, wherein said polysaccharide has GalNAc at the reducing terminus, and further wherein said GalNAc is directly linked to the protein carrier; and an adjuvant.

[0091] In yet an additional embodiment, the present invention is directed to a bioconjugate vaccine comprising: a protein carrier comprising at least one inserted consensus sequence, D/E – X – N – Z – S/T, wherein X and Z may be any natural amino acid except proline; at least one immunogenic polysaccharide from at least one bacterium, linked to the protein carrier, wherein the at least one immunogenic polysaccharide contains GalNAc at the reducing terminus directly linked to the protein carrier; and, optionally, an adjuvant.

[0092] Another embodiment of the invention is directed to a method of producing a bioconjugate vaccine, said method comprising: assembling a polysaccharide having GalNAc at the reducing terminus in a recombinant organism through the use of glycosyltransferases; linking said GalNAc to an asparagine residue of one or more target proteins in said recombinant organism, wherein said one or more target proteins contain one or more T-cell epitopes.

[0093] In a further embodiment, the present invention is directed to a method of producing a bioconjugate vaccine, said method comprising: introducing genetic information encoding for a metabolic apparatus that carries out *N*-glycosylation of a target protein into a prokaryotic organism to produce a modified prokaryotic organism; wherein the genetic information required for the expression of one or more recombinant target proteins is introduced into said prokaryotic organism; wherein the genetic information required for the expression of *E. coli* strain O157 epimerase is introduced into said prokaryotic organism; and wherein the metabolic apparatus comprises glycosyltransferases of a type that assembles a polysaccharide having GalNAc at the reducing terminus on a lipid carrier, and an

oligosaccharyltransferase, the oligosaccharyltransferase covalently linking GalNAc of the polysaccharide to an asparagine residue of the target protein, and the target protein containing at least one T-cell epitope; producing a culture of the modified prokaryotic organism; and obtaining glycosylated proteins from the culture medium.

[0094] A further aspect of the present invention relates to a pharmaceutical composition. An additional aspect of the invention involves a pharmaceutical composition comprising at least one *N*-glycosylated protein according to the invention. In light of the disclosure herein, the preparation of medicaments comprising proteins would be well known in the art. A still further aspect of the invention relates to a pharmaceutical composition comprising an antibiotic that inhibits an epimerase that converts GlcNAc-P-P-Und to GalNAc-P-P-Und. In a preferred embodiment, the pharmaceutical composition of the invention comprises a pharmaceutically acceptable excipient, diluent and/or adjuvant.

[0095] Suitable excipients, diluents and/or adjuvants are well-known in the art . An excipient or diluent may be a solid, semi-solid or liquid material which may serve as a vehicle or medium for the active ingredient. One of ordinary skill in the art in the field of preparing compositions can readily select the proper form and mode of administration depending upon the particular characteristics of the product selected, the disease or condition to be treated, the stage of the disease or condition, and other relevant circumstances (Remington's Pharmaceutical Sciences, Mack Publishing Co. (1990)). The proportion and nature of the pharmaceutically acceptable diluent or excipient are determined by the solubility and chemical properties of the pharmaceutically active compound selected, the chosen route of administration, and standard pharmaceutical practice. The pharmaceutical preparation may be adapted for oral, parenteral or topical use and may be administered to the patient in the form of tablets, capsules, suppositories, solution, suspensions, or the like. The pharmaceutically active compounds of the present invention, while effective themselves, can be formulated and administered in the form of their pharmaceutically acceptable salts, such as acid addition salts or base addition salts, for purposes of stability, convenience of crystallization, increased solubility, and the like.

[0096] In instances where specific nucleotide or amino acid sequences are noted, it will be understood that the present invention encompasses homologous sequences that still embody the same functionality as the noted sequences. In an

embodiment of the invention, such sequences are at least 85% homologous. In another embodiment, such sequences are at least 90% homologous. In still further embodiments, such sequences are at least 95% homologous.

[0097] The determination of percent identity between two nucleotide or amino acid sequences is known to one of skill in the art.

[0098] Nucleic acid sequences described herein, such as those described in the sequence listing below, are examples only, and it will be apparent to one of skill in the art that the sequences can be combined in different ways. Additional embodiments of the invention include variants of nucleic acids. A variant of a nucleic acid (e.g., a codon-optimized nucleic acid) can be substantially identical, that is, at least 80% identical, for example, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 99.5% identical, to SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29. Nucleic acid variants of a sequence that contains SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29 include nucleic acids with a substitution, variation, modification, replacement, deletion, and/or addition of one or more nucleotides (for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175 or 200 nucleotides) from a sequence that contains SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28 or SEQ ID NO: 29, or parts thereof.

[0100] For example, in an embodiment of the instant invention, such variants include nucleic acids that encode an epimerase which converts GlcNAc-P-P-Und to GalNAc-P-P-Und and that i) are expressed in a host cell, such as, for example, *E. coli* and ii) are substantially identical to SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9, or parts thereof.

[0101] Nucleic acids described herein include recombinant DNA and synthetic (e.g., chemically synthesized) DNA. Nucleic acids can be double-stranded or single-stranded. In the case of single-stranded nucleic acids, the nucleic acid can be a sense strand or antisense strand. Nucleic acids can be synthesized using oligonucleotide analogs or derivatives.

[0102] Plasmids that include a nucleic acid described herein can be transfected or transformed into host cells for expression. Techniques for transfection and transformation are known to those of skill in the art.

[0103] All publications mentioned herein are incorporated by reference in their entirety. It is to be understood that the term "or," as used herein, denotes alternatives that may, where appropriate, be combined; that is, the term "or" includes each listed alternative separately as well as their combination. As used herein, unless the context clearly dictates otherwise, references to the singular, such as the singular forms "a," an," and "the," include the plural, and references to the plural include the singular.

[0104] The invention is further defined by reference to the following examples that further describe the compositions and methods of the present invention, as well as its utility. It will be apparent to those skilled in the art that modifications, both to compositions and methods, may be practiced which are within the scope of the invention.

Examples

[0105] *Bacterial Strains and Plasmids- E. coli* strains PR4019 (Rush, J. S., Rick, P. D. and Waechter, C. J. (1997) Glycobiology, 7, 315-322) and PR21546 (Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. and Rick, P. D. (1990) *J. Biol. Chem.*, 265, 13490-13497) were generous gifts from Dr. Paul Rick, Bethesda, MD, and *E. coli* O157:H45 (Stephan, R., Borel, N., Zweifel, C., Blanco, M., and Blanco, J.E. (2004) *BMC Microbiol* 4:10) was a gift from Dr. Claudio Zweifel, Veterinary Institute, University of Zurich. *E. coli DH5α* (Invitrogen) was used as the host for cloning experiments and for protein glycosylation analysis. Plasmids used are listed in Table 2.

TABLE 2

Plasmids used in Examples

Plasmid	Description	Ref
pMLBAD	Cloning vector, Tmp^R	Lefebre & Valvano (2002)
pMLBAD:Z3206 (SEQ ID NO: 23)	Z3206 in pMLBAD, Tmp^R ,expression controlled by arabinose-inducible promoter	Examples herein
pMLBAD:gne (SEQ ID NO: 24)	*gne* in pMLBAD, Tmp^R,expression controlled by arabinose-inducible promoter	Examples herein
pACYCpgl	*C. jejuni pgl* cluster Cm^R	Wacker, *et al.* (2002)
pACYC*gne::kan*	*C. jejuni pgl* cluster containing a kan cassette in *gne*, Cm^R, Kan^R	Linton, *et al.* (2005)
pWA2	Soluble periplasmic hexa-His-tagged AcrA under control of Tet promoter in pBR322, Amp^R	Feldman, *et al.* (2005)

[0106] *Materials*— [1,6-^{3}H]GlcNAc (30 Ci/mmol), UDP-[1-^{3}H]GlcNAc (20 Ci/mmol) and UDP-[6-^{3}H]GalNAc (20 Ci/mmol) were obtained from American Radiolabeled Chemicals (St. Louis, MO). Quantum 1 silica gel G thin layer plates are a product of Quantum Industries (Fairfield, NJ), and Baker Si250 Silica Gel G plates are manufactured by Mallinckrodt Chemical Works. Yeast extract and Bacto-peptone were products of BD Biosciences. All other chemicals were obtained from standard commercial sources. Trimethoprim (50 μg/ml), chloramphenicol (20 μg/ml), ampicillin (100 μg/ml), and kanamycin (50 μg/ml) were added to the media as needed.

[0107] *Construction of Recombinant Plasmids— E. coli* strain DH5α was used for DNA cloning experiments and constructed plasmids were verified by DNA sequencing. The *Z3206* gene was amplified from *E. coli* O157:H45 by PCR with oligonucleotides Z3206-Fw and Z3206-RvHA (AAACCCGGGATGAACGATAACG TTTTGCTC (SEQ ID NO: 17) and AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTACTCAGAAACAA ACGTTATGTC (SEQ ID NO: 18); restriction sites are underlined). The PCR fragment was digested with SmaI and XbaI and ligated into SmaI-XbaI cleaved

pMLBAD vector (Lefebre, M. D. and Valvano M. A. (2002) *Appl Environ Microbiol* 68: 5956-5964). This resulted in plasmid pMLBAD:Z3206 (SEQ ID NO: 23) encoding Z3206 with a C-terminal hemagglutinin tag.

[0108] The *gne* gene was amplified from pACYCpgl (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) *Science* 298, 1790-1793), encoding *Campylobacter jejuni pgl* cluster, with oligonucleotides gne-Fw and gne-RV (AAACCATGGATGAAAATTCTTATTAGCGG (SEQ ID NO: 19) and AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTAGCACTGTTTTTC CCAATC (SEQ ID NO: 20); restriction sites are underlined). The PCR product was digested with NcoI and XbaI and ligated into the same sites of pMLBAD to generate plasmid pMLBAD:gne (SEQ ID NO: 24) which encodes Gne with a C-terminal hemagglutinin tag (Table 2).

[0109] *Growth Conditions, Protein Expression and Immunodetection— E. coli* strains were cultured in Luria-Bertani medium (1% yeast extract, 2% Bacto-peptone, 0.6% NaCl) at 37°C with vigorous shaking. Arabinose inducible expression was achieved by adding arabinose at a final concentration of 0.02-0.2% (w/v) to *E. coli* cells grown up to an A_{600} of 0.05-0.4. The same amount of arabinose was added again 5 h post-induction, and incubation continued for 4-15 h.

[0110] *Analytical Procedures*— Protein concentrations were determined using the BCA protein assay (Pierce) after precipitation of membrane proteins with deoxycholate and trichloroacetic acid according to the Pierce Biotechnology bulletin "Eliminate Interfering Substances from Samples for BCA Protein Assay." Samples were analyzed for radioactivity by scintillation spectrometry in a Packard Tri-Carb 2100TR liquid scintillation spectrometer after the addition of 0.5 ml of 1% SDS and 4 ml of Econosafe Economical Biodegradable Counting Mixture (Research Products International, Corp., Mount Prospect, IL).

Example 1: Identification of an *E. coli* O157 Gene Encoding GlcNAc-P-P-Und 4-Epimerase

[0111] We describe herein the surprising discovery of a new biosynthetic pathway in which GalNAc-P-P-Und is formed by the epimerization of the 4-OH of GlcNAc-P-P-Und catalyzed by the previously unknown action of a 4-epimerase. In

this pathway, GlcNAc-P-P-Und is formed by the transfer of GlcNAc-P from UDP-GlcNAc, catalyzed by WecA, and then GlcNAc-P-P-Und is epimerized to GalNAc-P-P-Und by GlcNAc-P-P-Und -4-epimerase, which was a previously unknown pathway (FIG. 2).

[0112] The gene encoding a candidate for the GlcNAc-P-P-Und 4-epimerase was identified by DNA homology searches. Homology searches were performed using the U.S. National Library of Medicine databases found at http://blast.ncbi.nlm.nih.gov/Blast.cgi. Genomic sequences of different bacteria encoding O antigen repeating units having a GalNAc at the reducing terminus were screened. One group with a repeating unit containing a GalNAc at the reducing terminus, and a second group lacking a terminal GalNAc in the repeating unit were compared to identify potential epimerases. Using these criteria, *Z3206* was identified as a candidate GlcNAc-P-P-Und 4-epimerase (Table 1).

[0113] The GlcNAc 4-epimerase genes present in *E. coli* strains with O-antigen repeat units containing GalNAc can be separated into two homology groups as shown in Table 1. It was surprisingly discovered that one homology group (containing *gne1*) clearly was correlated with the presence of GalNAc as the initiating sugar on the O-antigen repeat unit. It was further surprisingly discovered that the second group (containing *gne2*) exhibits a high degree of similarity to the UDP-Glc epimerase, *GalE*, and is found in *E. coli* strains that do not initiate O-antigen repeat unit synthesis with GalNAc. *Z3206* in *E. coli* O157, a gene with a high degree of homology to *gne1*, was identified as a candidate GlcNAc-P-P-Und 4-epimerase. The genomic location of the *Z3206* gene is consistent with a role in this pathway, as it resides between *galF* of the O-antigen cluster and *wcaM* which belongs to the colanic acid cluster.

[0114] The research described in Examples 2-11 further confirms the above discoveries, including identifying the GlcNAc 4-epimerase (*E. coli* O157 *Z3206*) as catalyzing the formation of GalNAc-P-P-Und.

Example 2: UDP-GalNAc Is Not a Substrate for *E. coli* WecA (GlcNAc-phosphotransferase)

[0115] To determine if *E. coli* WecA will utilize UDP-GalNAc as a GalNAc-P donor to form GalNAc-P-P-Und, membrane fractions from *E. coli* strains K12, PR4019, a WecA-overexpressing strain, and O157, which synthesize a tetrasaccharide O-antigen repeat unit with GalNAc at the reducing terminus presumably initiated by the synthesis of GalNAc-P-P-Und, were incubated with UDP-[^{3}H]GalNAc.

[0116] *Preparation of E. coli membranes*— Bacterial cells were collected by centrifugation at 1,000 X *g* for 10 min, washed once in ice-cold phosphate-buffered saline, once with cold water, and once with 10 mM Tris-HCl, pH 7.4, 0.25 M sucrose. The cells were resuspended to a density of ~ 200 A_{600} units/ml in 10 mM Tris-HCl, pH 7.4, 0.25 M sucrose, 10 mM EDTA containing 0.2 mg/ml lysozyme, and incubated at 30°C for 30 min. Bacterial cells were recovered by centrifugation at 1,000 X *g* for 10 min, quickly resuspended in 40 volumes of ice-cold 10 mM Tris-HCl, pH 7.4, and placed on ice. After 10 min the cells were homogenized with 15 strokes with a tight-fitting Dounce homogenizer and supplemented with 0.1 mM phenylmethylsulfonyl fluoride and sucrose to a final concentration of 0.25 M. Unbroken cells were removed by centrifugation at 1,000 X *g* for 10 min, and cell envelopes were recovered by centrifugation at 40,000 X *g* for 20 min. The membrane fraction was resuspended in 10 mM Tris-HCl, pH 7.4, 0.25 M sucrose, 1 mM EDTA and again sedimented at 40,000 X *g* and resuspended in the same buffer to a protein concentration of ~ 20 mg/ml. Membrane fractions were stored at –20°C until needed.

[0117] *Assay For the Biosynthesis of [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und in E. coli Membranes In Vitro*— Reaction mixtures for the synthesis of GlcNAc-P-P-Und and GalNAc-P-P-Und contained 50 mM Tris-HCl, pH 8, 40 mM $MgCl_2$, 5 mM dithiothreitol, 5 mM 5′AMP, *E. coli* membrane fraction (50-200 μg membrane protein, and either 5 μM UDP-[^{3}H]GlcNAc/GalNAc (500-2500 dpm/pmol) in a total volume of 0.05 ml. After incubation at 37°C, reactions were terminated by the addition of 40 volumes of $CHCl_3/CH_3OH$ (2:1), and the total lipid extract containing [^{3}H]HexNAc-P-P-undecaprenols was prepared as described previously (Waechter, C. J., Kennedy, J. L. and Harford, J. B. (1976) *Arch. Biochem. Biophys.* 174, 726-737). After partitioning, the organic phase was dried under a stream of

nitrogen and redissolved in 1 ml $CHCl_3/CH_3OH$ (2:1), and an aliquot (0.2 ml) was removed, dried in a scintillation vial, and analyzed for radioactivity by liquid scintillation spectrometry in a Packard Tri-Carb 2100 TR liquid scintillation spectrometer. To determine the rate of synthesis of [^{3}H]GlcNAc-P-P-Und or [^{3}H]GalNAc-P-P-Und, the lipid extract was dried under a stream of nitrogen, redissolved in a small volume of $CHCl_3/CH_3OH$ (2:1), and spotted on a 10 X 20-cm borate-impregnated Baker Si250 silica gel plate, and the plate was developed with $CHCl_3$, CH_3OH, H_2O, 0.2 M sodium borate (65:25:2:2). Individual glycolipids were detected with a Bioscan AR2000 Imaging Scanner (Bioscan, Washington, D.C.). The biosynthetic rates for each glycolipid were calculated by multiplying the total amount of radioactivity in [^{3}H]GlcNAc/GalNAc-P-P-Und by the percentage of the individual [^{3}H] glycolipids.

[0118] Membrane fractions from different *E. coli* strains (K12, PR4019 and O157) were incubated with either UDP-[^{3}H]GlcNAc or UDP-[^{3}H]GalNAc and the incorporation into [^{3}H]GlcNAc/GalNAc-P-P-Und was determined as described above. As seen in Table 3, no labeled glycolipids were detected after the incubation with UDP-[^{3}H]GalNAc, only GlcNAc-P-P-Und was detectable when membrane fractions were incubated with UDP-[^{3}H]GlcNAc

TABLE 3

Synthesis of [^{3}H]GlcNAc/GalNAc-P-P-undecaprenol in *E. coli* membrane fractions using either UDP-[^{3}H]GlcNAc or UDP-[^{3}H]GalNAc as substrate

Source of membranes	Sugar nucleotide added	[^{3}H]Glycolipid formed	
		GlcNAc-P-P-Und	GalNAc-P-P-Und
		(*pmol/mg*)	(*pmol/mg*)
K12	UDP-[^{3}H]GlcNAc	6.4	<0.01
K12	UDP-[^{3}H]GalNAc	<0.01	<0.01
PR4019	UDP-[^{3}H]GlcNAc	44	<0.01
PR4019	UDP-[^{3}H]GalNAc	<0.01	<0.01
O157	UDP-[^{3}H]GlcNAc	1.5	0.5
O157	UDP-[^{3}H]GalNAc	<0.01	<0.01

[0119] Moreover, neither the addition of exogenous Und-P to incubations with membranes from PR4019, the WecA-overexpressing strain, or the addition of

cytosolic fractions from O157 cells resulted in the formation of GalNAc-P-P-Und from UDP-GalNAc. These results demonstrate that UDP-GalNAc is not a substrate for WecA and suggest that GalNAc-P-P-Und is formed by an alternative mechanism.

[0120] When membranes from strain K12 were incubated with UDP-[^{3}H]GlcNAc, [^{3}H]GlcNAc-P-P-Und was synthesized as expected (Rush, J. S., Rick, P. D. and Waechter, C. J. (1997) Glycobiology, 7, 315-322). However, when membranes from strain O157 were incubated with UDP-[^{3}H]GlcNAc, in addition to [^{3}H]GlcNAc-P-P-Und, a second labeled lipid shown to be [^{3}H]GalNAc-P-P-Und (see below) was observed. When the time course for the formation of the two glycolipids was examined, the incorporation of radioactivity into [^{3}H]GlcNAc-P-P-Und (FIG. 1, O) occurred more quickly and to a higher extent than into [^{3}H]GalNAc-P-P-Und (FIG. 1, ●), compatible with a precursor-product relationship (FIG. 2).

[0121] The observation that *E. coli* O157 membranes do not utilize UDP-GalNAc as a GalNAc-P donor for the synthesis of GalNAc-P-P-Und is one example which confirms the biosynthetic pathway for the formation of GalNAc-P-P-Und illustrated in FIG. 2. In this scheme, GlcNAc-P-P-Und is formed by the transfer of GlcNAc-P from UDP-GlcNAc, catalyzed by WecA, and then GlcNAc-P-P-Und is epimerized by the action of a previously unknown 4-epimerase to produce GalNAc-P-P-Und.

Example 3: Characterization of [^{3}H]GalNAc-P-P-Und Formed *in Vitro* with Membrane Fractions from *E. coli* Strain O157

[0122] Consistent with the additional O157-specific glycolipid product detected in FIG. 1, as GalNAc-P-P-Und, it was stable to mild alkaline methanolysis (toluene/methanol 1:3, containing 0.1 N KOH, 0°C, 60 min), retained by DEAE-cellulose equilibrated in $CHCl_3/CH_3OH/H_2O$ (10:10:3), and eluted with $CHCl_3/CH_3OH/H_2O$ (10:10:3) containing 20 mM ammonium acetate as reported previously for [^{3}H]$GlcNAc_{1-2}$-P-P-Dol (Waechter, C. J. and Harford, J. B. (1977) *Arch. Biochem. Biophys.* 181, 185-198).

[0123] [^{3}H]GalNAc-P-P-Und was clearly resolved from [^{3}H]GlcNAc-P-P-Und by thin layer chromatography on borate-impregnated silica gel G (Kean, E. L. (1966) *J. Lipid Res.* 7, 449-452) and purified by preparative TLC as shown in FIG. 3A and FIG. 3B.

[0124] *Preparation of Borate-impregnated Thin Layer Plates and Whatman No. 1 Paper*— Silica gel thin layer plates were impregnated with sodium borate by briefly immersing the plates in 2.5% $Na_2B_4O_7 \cdot 10\ H_2O$ in 95% methanol as described by Kean (Kean, E. L. (1966) *J. Lipid Res.* 7, 449-452). The borate-impregnated TLC plates were dried overnight at room temperature and stored in a vacuum dessicator over Drierite until use. Immediately before chromatography, the plates were activated by heating briefly (~10-15 min) to 100°C. Whatman No. 1 paper was impregnated with sodium borate by dipping 20 X 30-cm sheets of Whatman 1 paper in 0.2 M $Na_2B_4O_7 \cdot 10\ H_2O$. The Whatman No. 1 paper sheets were pressed firmly between two sheets of Whatman No. 3MM paper and allowed to dry at room temperature for several days, as described by Cardini and Leloir (Cardini, C. E. and Leloir, L. F. (1957) *J. Biol. Chem.* 225, 317-324).

[0125] *Characterization of Glycan Products Formed in in Vitro Reactions*— The glycans of the individual glycolipids ([^{3}H]GalNAc-P-P-Und and [^{3}H]GlcNAc-P-P-Und) were characterized by descending paper chromatography after release by mild acid hydrolysis. The GlcNAc/GalNAc lipids were dried under a stream of nitrogen in a conical screw-cap tube and heated to 100°C, 15 min in 0.2 ml 0.01 M HCl. After hydrolysis the samples were applied to a 0.8-ml mixed-bed ion-exchange column containing 0.4 ml of AG50WX8 (H^+) and 0.4 ml AG1X8 (acetate form) and eluted with 1.5 ml water. The eluate was dried under a stream of nitrogen, redissolved in a small volume of H_2O (0.02 ml), spotted on a 30-cm strip of borate-impregnated Whatman No. 1 paper, and developed in descending mode with butanol/pyridine/water (6:4:3) for 40-50 h. After drying, the paper strips were cut into 1-cm zones and analyzed for radioactivity by scintillation spectrometry. GlcNAc and GalNAc standards were detected using an aniline-diphenylamine dip reagent (Schwimmer, S. and Benvenue, A. (1956) *Science* 123, 543-544).

[0126] Glycan products were converted to their corresponding alditols by reduction with 0.1 M $NaBH_4$ in 0.1 M NaOH (final volume 0.1 ml) following mild acid hydrolysis as described above. After incubation at room temperature overnight, the reactions were quenched with several drops of glacial acetic acid and dried under a stream of nitrogen out of methanol containing 1 drop of acetic acid, several times. The alditols were dissolved in water, desalted by passage over 0.5 ml columns of AG50WX8 (H+) and AG1X8 (acetate), dried under nitrogen, and spotted on 30-cm

strips of Whatman No. 3MM paper. The Whatman No. 3 MM strips were developed overnight in descending mode with ethyl acetate, pyridine, 0.1 M boric acid (65:25:20), dried, cut into 1-cm zones, and analyzed for radioactivity by scintillation spectrometry. GlcNAcitol and GalNAcitol standards were visualized using a modification of the periodate-benzidine dip procedure (Gordon, H. T., Thornburg, W. and Werum, L. N. (1956) *Anal. Chem.* 28, 849-855). The paper strips were dipped in acetone, 0.1 M $NaIO_4$ (95:5), allowed to air dry for 3 min, and then dipped in acetone/acetic acid/H_2O/o-tolidine (96:0.6:4.4:0.2 gm). Alditols containing *cis*-diols stain as yellow spots on a blue background.

[0127] *Mass Spectrometry ("MS") of Glycolipids*— Purified glycolipids were analyzed using an ABI/MDS Sciex 4000 Q-Trap hybrid triple quadrupole linear ion trap mass spectrometer with an ABI Turbo V electrospray ionsource (ABI/MDS-Sciex, Toronto, Canada). In brief, samples were infused at 10 µl/min with ion source settings determined empirically, and MS/MS (mass spectroscopy in a second dimension) information was obtained by fragmentation of the molecular ion in linear ion trap mode.

[0128] When the glycolipid was treated with mild acid (0.01 N HCl, 100°C, 15 min), the water-soluble product co-chromatographed with [^{3}H]GalNAc on descending paper chromatography with borate-impregnated Whatman No. 1 paper (FIG. 3C). In addition, when the labeled sugar was reduced, it was converted to [^{3}H]alditol, GalNAc-OH (FIG. 3D). Moreover, negative-ion MS analysis yielded the [M-H]- ion of m/z = 1128, expected for GalNAc-P-P-Und, and the MS/MS daughter ion spectrum showed a prominent ion at m/z = 907, expected for a glycolipid containing P-P-Und (Guan, Z., Breazeale, S. D. and Raetz, C. R. (2005) *Anal. Biochem.* 345, 336-339). The identification of the glycolipid product formed by strain O157 as GalNAc-P-P-Und is also supported by its formation from exogenous GlcNAc-P-P-Und (see Example 7).

Example 4: Metabolic Labeling of [^{3}H]GalNAc-P-P-Und (*in vivo*) with [^{3}H]GlcNAc in *E. coli* Cells Expressing the Z3206 Gene

[0129] To investigate whether expression of the *E. coli* O157 *Z3206* gene enabled cells to synthesize GalNAc-P-P-Und, *E. coli* strain 21546 (Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. and Rick, P. D. (1990) *J. Biol. Chem.*, 265, 13490-

13497) expressing the *Z3206* gene was labeled metabolically with [^{3}H]GlcNAc and analyzed for [^{3}H]GlcNAc/GalNAc-P-P-Und formation.

[0130] *Metabolic Labeling of Bacterial Cells— E. coli* cells were cultured with vigorous shaking in Luria-Bertani medium at 37°C to an A_{600} of 0.5-1. [^{3}H]GlcNAc was added to a final concentration of 1 μCi/ml and the incubation was continued for 5 min at 37°C. The incorporation of radiolabel into glycolipids was terminated by the addition of 0.5 gm/ml crushed ice, and the cultures were thoroughly mixed. The bacterial cells were recovered by centrifugation at 4000 X *g* for 10 min, and the supernatant was discarded. The cells were washed with ice-cold phosphate-buffered saline two times, resuspended by vigorous vortex mixing in 10 volumes (cell pellet) of methanol, and sonicated briefly with a probe sonicator at 40% full power. After sonication, 20 volumes of chloroform were added, and the extracts were mixed vigorously and allowed to stand at room temperature for 15 min. The insoluble material was sedimented by centrifugation, and the pellet was re-extracted with a small volume of $CHCl_3/CH_3OH$ (2:1) twice. The combined organic extracts were then processed as described below.

[0131] *Purification of GlcNAc-P-P-Und and GalNAc-P-P-Und—* GlcNAc/GalNAc-P-P-Und was extracted with $CHCl_3/CH_3OH$ (2:1) and freed of water-soluble material by partitioning as described elsewhere (Waechter, C. J., Kennedy, J. L. and Harford, J. B. (1976) *Arch. Biochem. Biophys.* 174, 726-737). The organic extract was then dried under a stream of nitrogen, and the bulk glycerophospholipids were destroyed by deacylation in toluene/methanol (1:3) containing 0.1 N KOH at 0°C for 60 min. The deacylation reaction was neutralized with acetic acid, diluted with 4 volumes of $CHCl_3/CH_3OH$ (2:1), and washed with 1/5 volume of 0.9% NaCl. The organic (lower) phase was washed with 1/3 volume of $CHCl_3$, CH_3OH, 0.9 % NaCl (3:48:47), and the aqueous phase was discarded. The organic phase was diluted with sufficient methanol to accommodate the residual aqueous phase in the organic phase and applied to a DEAE-cellulose column (5 ml) equilibrated with $CHCl_3/CH_3OH$ (2:1). The column was washed with 20 column volumes of $CHCl_3/CH_3OH/H_2O$ (10:10:3) and then eluted with $CHCl_3/CH_3OH/H_2O$ (10:10:3) containing 20 mM ammonium acetate. Fractions (2 ml) were collected and monitored for either radioactivity, or GlcNAc/GalNAc-P-P-Und using an anisaldehyde spray reagent (Dunphy, P. J., Kerr, J. D., Pennock, J. F., Whittle, K. J.,

and Feeney, J. (1967) *Biochim. Biophys. Acta* 136, 136-147) after resolution by thin layer chromatography on borate-impregnated silica plates (as described earlier).

[0132] *E. coli* strain 21546 was selected as the host for the *Z3206* expression studies because a mutation in UDP-ManNAcA synthesis results in a block in the utilization of GlcNAc-P-P-Und for the synthesis of the enterobacterial common antigen. Because *E. coli* 21546 is derived from *E. coli* K12 it does not synthesize an O-antigen repeat as well (Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. and Reeves, P. (1994) *J. Bacteriol.*, 176, 4144-4156), and thus, larger amounts of GlcNAc-P-P-Und accumulate for the conversion to GalNAc-P-P-Und. When strain 21546 and the transformant expressing the *Z3206* gene were labeled with [^{3}H]GlcNAc and the radiolabeled lipids were analyzed by thin layer chromatography on borate-impregnated silica gel plates, the parental strain (FIG. 4A) synthesized only one labeled lipid, GlcNAc-P-P-Und. However, 21546 cells expressing the *Z3206* gene (FIG. 4B) also synthesized an additional labeled lipid shown to be GalNAc-P-P-Und.

Example 5: Membrane Fractions from *E. coli* Cells Expressing the Z3206 Gene Synthesize GalNAc-P-P-Und *in Vitro*

[0133] To corroborate that the protein encoded by the *E. coli* O157 *Z3206* gene catalyzed the synthesis of GalNAc-P-P-Und, membrane fractions from *E. coli* cells expressing the *Z3206* gene were incubated with [^{3}H]UDP-GlcNAc and the [^{3}H]glycolipid products were analyzed by thin layer chromatography (chromatographic preparation and characterization methods are described in Example 3) on borate-impregnated silica gel plates as shown in FIG. 5. When membrane fractions from *E. coli* K12 or the host strain *E. coli* 21546 cells were incubated with UDP-[^{3}H]GlcNAc, only [^{3}H]GlcNAc-P-P-Und was observed (FIG. 5A and FIG. 5C). However, membrane fractions from *E. coli* O157 and *E. coli* 21546 expressing *Z3206* formed GalNAc-P-P-Und as well (FIG. 5B and FIG. 5D).

Example 6: Formation of GlcNAc-P-P-Und, but Not GalNAc-P-P-Und, Is Reversed in the Presence of UMP

[0134] To provide additional evidence that GalNAc-P-P-Und is synthesized from GlcNAc-P-P-Und, and not by the action of WecA using UDP-GalNAc as a

glycosyl donor, the effect of discharging endogenous, pre-labeled [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und with UMP was examined. The GlcNAc-phosphotransferase reaction catalyzed by WecA is freely reversible by the addition of excess UMP re-synthesizing UDP-GlcNAc and releasing Und-P.

[0135] In this experiment membrane fractions from *E. coli* strain 21546 expressing Z3206 were pre-labeled for 10 min with UDP-[^{3}H]GlcNAc followed by the addition of 1 mM UMP, and the amount of each labeled glycolipid remaining was determined. The results illustrated in FIG. 6A show the relative amounts of [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und at the end of the 10 min labeling period. After incubation with 1 mM UMP for 1 min it can be seen that there is a substantial loss of [^{3}H]GlcNAc-P-P-Und, whereas the [^{3}H]GalNAc-P-P-Und peak is relatively unchanged (FIG. 6B) (chromatographic preparation and characterization methods are described in Example 5). This observation is consistent with the results in Table 3 indicating that WecA does not catalyze the transfer of GalNAc-P into GalNAc-P-P-Und from UDP-GalNAc. It is noteworthy that during the second minute of incubation with UMP (FIG. 6C), the loss of GlcNAc-P-P-Und slows, and there is a slight reduction in the peak of [^{3}H]GalNAc-P-P-Und, suggesting that [^{3}H]GalNAc-P-P-Und is re-equilibrating with the [^{3}H]GlcNAc-P-P-Und pool by reversal of the epimerase reaction (see Example 7).

Example 7: Interconversion of Exogenous, Purified [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und Catalyzed by Membranes from *E. coli* Cells Expressing Z3206

[0136] To provide direct evidence that GlcNAc-P-P-Und and GalNAc-P-P-Und can be directly interconverted by membrane fractions from *E. coli* cells expressing Z3260, purified [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und were tested as exogenous substrates.

[0137] Purified [^{3}H]GlcNAc-P-P-Und/[^{3}H]GalNAc-P-P-Und were prepared as in Example 4 (*Metabolic Labeling of Bacterial Cells* and *Purification of GlcNAc-P-P-Und and GalNAc-P-P-Und*). [^{3}H]HexNAc-P-P-undecaprenols (2000 dpm/pmol, dispersed in 1 % Triton X-100, final concentration 0.1%) were incubated with *E. coli* membranes as in Example 2 in *Assay For the Biosynthesis of [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und in E. coli Membranes In Vitro.*

[0138] Preliminary experiments showed that the epimerase was active when exogenous [^{3}H]GlcNAc-P-P-Und was added to the reaction mixtures dispersed in Triton X-100, CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid), Nonidet P-40, or octylglucoside and exhibited a pH optimum in the range 7-8.5. The chromatographic mobility of the purified [^{3}H]GlcNAc-P-P-Und and [^{3}H]GalNAc-P-P-Und before incubation with membrane fractions is shown in FIG. 7A and FIG. 7D. As seen in FIG. 7B and FIG. 7E, the glycolipids are unaffected by incubation with membrane fractions from *E. coli* 21546. However, incubation of the purified glycolipids with membrane fractions from *E. coli* 21546 expressing Z3206 catalyzes the conversion of exogenous [^{3}H]GlcNAc-P-P-Und to [^{3}H]GalNAc-P-P-Und (FIG. 7C) and the conversion of [^{3}H]GalNAc-P-P-Und to [^{3}H]GlcNAc-P-P-Und (FIG. 7F). These results demonstrate directly that GlcNAc-P-P-Und and GalNAc-P-P-Und can be enzymatically interconverted in *E. coli* strains expressing the Z3206.

Example 8: *E. coli* Z3206 Is Not a UDP-GlcNAc 4-Epimerase

[0139] To determine if Z3206 can catalyze the formation of UDP-GalNAc, the *N*-glycosylation apparatus from *C. jejuni* was expressed in *E. coli*. In this reporter system, glycosylation of the target protein AcrA is dependent on the presence of the *pgl* locus (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) *Science* 298, 1790-1793), including a functional Gne UDP-Glc/UDP-GlcNAc epimerase (Bernatchez, S., Szymanski, C.M., Ishiyama, N., Li, J., Jarrell, H.C., Lau, P.C., Berghuis, A.M., Young, N.M., Wakarchuk, W.W. (2005) *J. Biol. Chem.* 280, 4792-4802). Glycosylation of AcrA is lost if the *pgl* cluster contains a deletion of *gne* (Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M. and Wren, B.W. (2005) *Mol Microbiol.* 55, 1695-1703). The ability of Z3206 to restore AcrA-glycosylation in the presence of the *pgl* operon *Δgne* was investigated *in vivo* by expressing AcrA (pWA2) together with the *pgl* locus *Δgne* complemented by either Gne (pMLBAD:gne) or *Z3206* (pMLBAD:Z3206).

[0140] Total *E. coli* cell extracts were prepared for immunodetection analysis using cells at a concentration equivalent to 1 A_{600} unit that were resuspended in 100 µl of SDS loading buffer (Laemmli, U. (1970) *Nature* 227, 680-685). Aliquots of 10 µl

were loaded on 10% SDS-PAGE. Periplasmic extracts of *E. coli* cells were prepared by lysozyme treatment (Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M. (2005) *Proc Natl Acad Sci USA* 102, 3016-3021), and 10 µl of the final sample (corresponding to 0.2 A_{600} units of cells) was analyzed by SDS-PAGE. After being blotted on nitrocellulose membrane, sample was immunostained with the specific antiserum (Aebi, M., Gassenhuber, J., Domdey, H., and te Heesen, S. (1996) *Glycobiology* 6, 439-444). Anti-AcrA (Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. (2002) *Science* 298, 1790-1793) antibodies were used. Anti-rabbit IgG-HRP (Bio-Rad) was used as secondary antibody. Detection was carried out with ECL™ Western blotting detection reagents (Amersham Biosciences).

[0141] As shown in FIG. 8, the glycosylated protein, which migrates slower than the unglycosylated form, was formed only when cells expressing *pgl* locus *Δgne* were complemented by Gne (*lane 2*). *Z3206* was unable to restore glycosylation of the reporter glycoprotein (FIG. 8, *lane 1*). Accordingly, Z3206 does not complement glycosylation of AcrA in a Gne dependent glycosylation system. Expression of *Gne* and membrane-associated Z3206 were confirmed by immunodetection.

Example 9: Analysis of *S. flexneri* 6 +/- Z3206 LPS

[0142] In Fig. 9 are depicted some of the genes required for the biosynthesis of the *Shigella flexneri 6* O-antigen: genes encoding enzymes for biosynthesis of nucleotide sugar precursors; genes encoding glycosyltransferases; genes encoding O antigen processing proteins; and genes encoding proteins responsible for the O-acetylation. The structure of the O antigen has been elucidated by Dmitriev, B.A. et al (Dmitriev, B.A., et al *Somatic Antigens of Shigella* Eur JBiochem, 1979. 98: p. 8; Liu B et al *Structure and genetics of Shigella O antigens* FEMS Microbiology Review, 2008. 32: p. 27).

[0143] To identify all the genes required for the biosynthesis of the *Shigella flexneri* 6 O-antigen a genomic library was constructed.

[0144] *Cloning of S. flexneri 6 genomic DNA*_ S. flexneri 6 genomic *DNA* was isolated using a Macherey-Nagel NucleoSpin® Tissue Kit following the protocol for DNA isolation from bacteria. DNA was isolated from five *S. flexneri* 6 overnight

cultures at 2 ml each and final elution was done with 100 µl elution buffer (5 mM Tris/HCl, pH 8.5). The eluted fractions were pooled, precipitated by isopropanol and the final pellet was resuspended in 52 µl TE buffer of which the total volume was subjected to end-repair according to the protocol given by CopyControlTM Fosmid Library Production Kit (EPICENTRE). End-repaired DNA was purified on a 1 % low melting point agarose gel run with 1 X TAE buffer, recovered and precipitated by ethanol as described in the kit protocol. Resuspension of the precipitated DNA was done in 7 µl TE buffer of which 0.15 µl DNA was ligated into pCC1FOS (SEQ ID NO: 27) according to the EPICENTRE protocol. Packaging of the ligation product into phage was performed according to protocol and the packaged phage was diluted 1:1 in phage dilution buffer of which 10 µl were used to infect 100 µl EPI300-T1 cells that were previous grown as described by EPICENTRE. Cells (110 µl) were plated six times with approximately 100 colonies per plate such that the six plates contain the entire *S. flexneri* 6 genomic library. Plates were developed by colony blotting and positive/negative colonies were western blotted and silver stained.

[0145] *Colony blotting_* For colony blots a nitrocellulose membrane was laid over the solid agar plate, removed, washed three times in 1 X PBST and treated in the same manner. The membrane was first blocked in 10 % milk for one hour at room temperature after which it was incubated for one hour at room temperature in 2 ml 1 % milk (in PBST) with the anti-type VI antiserum (primary antibody). After three washes in PBST at 10 minutes each, the membrane was incubated for another hour at room temperature in the secondary antibody, 1:20000 peroxidase conjugated goat-anti-rabbit IgG (BioRad) in 2 ml 1 % milk (in PBST). After a final three washes with PBST (10 minutes each) the membrane was developed in a UVP Chemi Doc Imaging System with a 1:1 mix of luminol and peroxide buffer provided by the SuperSignal® West Dura Extended Duration Substrate Kit (Thermo Scientific).

[0146] The clone reacting with *S. flexneri* 6 antiserum following production of a *S. flexneri* 6 genomic library was sequenced by primer walking out of the region previously sequenced by Liu *et al.* (Liu et al., 2008) reaching from rmlB to wfbZ (FIG. 9). Primers rmlB_rev and wfbZ_fwd (*S. flexneri* - Z3206) annealed in rmlB and wfbZ and were used to sequence the insert of the clone until wcaM and hisI/F were reached (*S. flexneri* + Z3206), respectively (FIG. 10).

[0147] In order to establish whether O antigen synthesis is maintained in clones lacking *Z3206* (thus hindering epimerization of und-GlcNAc to und-GalNAc), two plasmids were constructed (SEQ ID NO. 28 and SEQ ID NO. 29) (Fig. 10), transformed into *E. coli* cells and analyzed by silver staining and western blot.

[0148] As shown in FIG. 11, LPS is produced in E. coli cells + or – Z3206. The O antigen can be produced without Z3206 however with lower production yield, which indicates that the efficiency of polysaccharide production without the epimerase (Z3206) is lower.

Example 10: Analysis of *S. flexneri* 6 +/- Z3206 LLO

[0149] *Purification of undecaprenol-PP-O antigen by C18 column chromatography_E. coli* cells expressing S. flexeneri antigen +/- Z3206 were pelleted, washed once in 50 ml 0.9 % NaCl and the final pellets were lyophilized overnight. The pellets were washed once in 30 ml 85 – 95 % methanol, reextracted with 10:10:3 chloroform-methanol-water (v/v/v) and the extracts were converted to a two-phase Bligh/Dyer system by addition of water, resulting in a final ratio of 10:10:9 (C:M:W). Phases were separated by centrifugation and the upper aqueous phases were loaded each on a C18 Sep-Pak cartridge conditioned with 10 ml methanol and equilibrated with 10 ml 3:48:47 (C:M:W). Following loading, the cartridges were washed with 10 ml 3:48:47 (C:M:W) and eluted with 5 ml 10:10:3 (C:M:W). 20 OD samples of the loads, flow-throughs, washes and elutions of the C18 column were dried in an Eppendorf Concentrator Plus, washed with 250 µl methanol, reevaporated and washed a further three times with 30 µl ddH2O.

[0150] *Glycolipid hydrolysis* The glycolipid samples from the wash of the C18 column were hydrolysed by dissolving the dried samples in 2 ml *n*-propanol:2 M trifluoroacetic acid (1:1), heating to 50°C for 15 minutes and evaporating to dryness under N2.

[0151] *Oligosaccharide labeling with 2-aminobenzoate and HPLC* Labeling was done according to Bigge *et al.* (Bigge, 1995) and glycan cleanup was performed using the paper disk method described in Merry *et al.* (2002) (Merry et al., 2002). Separation of 2-AB labeled glycans was performed by HPLC using a GlycoSep-N normal phase column according to Royle *et al.* (Royle, 2002) but modified to a three solvent system. Solvent A was 10 mM ammonium formate pH 4.4 in 80 %

acetonitrole. Solvent B was 30 mM ammonium formate pH 4.4. in 40 % acetonitrile. Solvent C was 0.5 % formic acid. The column temperature was 30 °C and 2-AB labeled glycans were detected by fluorescence (λex = 330 nm, λem = 420 nm). Gradient conditions were a linear gradient of 100 % A to 100 % B over 160 minutes at a flow rate of 0.4 ml/min, followed by 2 minutes 100 % B to 100 % C, increasing the flow rate to 1 ml/min. The column was washed for 5 minutes with 100 % C, returning to 100 % A over 2 minutes and running for 15 minutes at 100 % A at a flow rate of 1 ml/min, then returning the flow rate to 0.4 ml/min for 5 minutes. All samples were injected in water.

[0152] The plasmids expressing the *S. flexneri* O antigen with (SEQ ID NO: 29) or without (SEQ ID NO: 28) *Z3206* were transformed into SCM3 cells (FIG. 10). Traces at late elution volumes shows a difference between the curves of the two samples containing the *S. flexneri* O antigen +/- Z3206 (FIG. 12). This difference in the elution pattern can be explained by a different oligosaccharide structure carrying a different monosaccharide at the reducing end: GlcNAc or GalNAc depending on the presence of the epimerase (Z3206).

Example 11: Analysis of pglB specificity by production and characterization of bioconjugate produced from *S. flexneri* 6 +/- Z3206

[0153] To assess whether PglB can transfer oligosaccharides having GlcNAc (*S. flexneri* 6 O-antigen) at the reducing end to the carrier protein EPA Nickel purified extracts from *E. coli* cells expressing EPA (SEQ ID NO: 25), PglB (SEQ ID NO: 26) and *S. flexneri* 6 O-antigen +/- Z3206 (SEQ ID NO: 29/SEQ ID NO: 28) were analyzed by western blot using anti EPA and anti type VI antibodies. The *S. flexneri* O6 antigen with and without GalNAc at the reducing end was transferred to EPA by PglB as detected by antiEPA and anti VI antisera (FIG. 13).

[0154] The O antigen is still produced and detected, but with lower production yield, which indicates that the efficiency of polysaccharide production without the epimerase is lower.

[0155] While this invention has been particularly shown and described with references to embodiments thereof, it will be understood by those skilled in the art that various changes in form and detail may be made therein without departing from

the scope of the invention encompassed by the claims. Such various changes that will be understood by those skilled in the art as covered within the scope of the invention include, in particular, *N*-glycosylated proteins and bioconjugates comprising a glycan other than those from *E. coli* and *S. flexneri* with GalNAc at the reducing terminus.

Sequence Listing

[0156] Applicant: GlycoVaxyn AG

[0157] Title: Biosynthetic System That Produces Immunogenic Polysaccharides In Prokaryotic Cells

[0158] Number of SEQ ID NOs: 29

[0159] **SEQ ID NO: 1 - Nucleotide Sequence for *E. coli* O157 *Z3206***

[0160] Length: 993

[0161] Type: DNA

[0162] Organism: *E. coli* O157

[0163] Sequence:

ATGAACGATAACGTTTTGCTCATAGGAGCTTCCGGATTCGTAGGAACCCGACTACTTGAAACGG
CAATTGCTGACTTTAATATCAAGAACCTGGACAAACAGCAGAGCCACTTTTATCCAGAAATCAC
ACAGATTGGCGATGTTCGCGATCAACAGGCACTCGACCAGGCGTTAGTCGGTTTTGACACTGTT
GTACTACTGGCAGCGGAACACCGCGATGACGTCAGCCCTACTTCTCTCTATTATGATGTCAACG
TTCAGGGTACCCGCAATGTGCTGGCGGCCATGGAAAAAAATGGCGTTAAAAATATCATCTTTAC
CAGTTCCGTTGCTGTTTATGGTTTGAACAAACACAACCCTGACGAAAACCATCCACACGACCCT
TTCAACCACTACGGCAAAAGTAAGTGGCAGGCAGAGGAAGTGCTGCGTGAATGGTATAACAAAG
CACCAACAGAACGTTCATTAACCATCATCCGTCCTACCGTTATCTTCGGTGAACGCAACCGCGG
TAACGTCTATAACTTGCTGAAACAGATCGCTGGCGGCAAGTTTATGATGGTGGGCGCAGGGACT
AACTATAAGTCCATGGCTTATGTTGGAAACATTGTTGAGTTTATCAAGTACAAACTGAAGAATG
TTGCCGCAGGTTATGAGGTTTATAACTACGTTGATAAGCCAGACCTGAACATGAACCAGTTGGT
TGCTGAAGTTGAACAAAGCCTGAACAAAAAGATCCCTTCTATGCACTTGCCTTACCCACTAGGA
ATGCTGGGTGGATATTGCTTTGATATCCTGAGCAAAATTACGGGCAAAAAATACGCTGTCAGCT
CAGTGCGCGTGAAAAAATTCTGCGCAACAACACAGTTTGACGCAACGAAAGTGCATTCTTCAGG
TTTTGTGGCACCGTATACGCTGTCGCAAGGTCTGGATCGAACACTGCAGTATGAATTCGTTCAT
GCCAAAAAAGACGACATAACGTTTGTTTCTGAG

[0164] **SEQ ID NO: 2 - Amino Acid Sequence for Z3206**

[0165] Length: 331

[0166] Type: PRT

[0167] Organism: *E coli* O157

[0168] Sequence:

MNDNVLLIGASGFVGTRLLETAIADFNIKNLDKQQSHFYPEITQIGDVRDQQALDQALV
GFDTVVLLAAEHRDDVSPTSLYYDVNVQGTRNVLAAMEKNGVKNIIFTSSVAVYGLNKHNPDEN
HPHDPFNHYGKSKWQAEEVLREWYNKAPTERSLTIIRPTVIFGERNRGNVYNLLKQIAGGKFMM
VGAGTNYKSMAYVGNIVEFIKYKLKNVAAGYEVYNYVDKPDLNMNQLVAEVEQSLNKKIPSMHL
PYPLGMLGGYCFDILSKITGKKYAVSSVRVKKFCATTQFDATKVHSSGFVAPYTLSQGLDRTLQ
YEFVHAKKDDITFVSE

[0169] **SEQ ID NO: 3 - Nucleotide Sequence for *E. coli* O55 *gne***

[0170] Locus AF461121_1 BCT 02-MAY-2002

[0171] Definition (UDP-GlcNAc 4-epimerase Gne [Escherichia coli])

[0172] Accession AAL67550

[0173] Length: 993

[0174] Type: DNA

[0175] Organism: *E. coli* O55

[0176] Sequence:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC
AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTGCTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA
ACACAACCCT GACGAAAACC ATCCACACGA TCCTTTCAAC CACTACGGCA
AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA
CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA
ACGGAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT

GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

[0177] **SEQ ID NO: 4 - Amino Acid Sequence for *E. coli* O55 UDP-GlcNAc 4-epimerase Gne**

[0178] Locus AF461121_1

[0179] Definition (UDP-GlcNAc 4-epimerase Gne [Escherichia coli])

[0180] Accession AAL67550

[0181] Length: 331 aa linear

[0182] Type: PRT

[0183] Organism: *E. coli* O55

[0184] Sequence:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp eitqigdvrd
qqaldqalag fdtvvllaae hrddvsptsl yydvnvqgtr nvlaameknq
vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv
gnivefikyk lknvaagyev ynyvdkpdln mnqlvaeveq slnkkipsmh
lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf
vapytlsqgl drtlqyefvh akkdditfvs e

[0185] **SEQ ID NO. 5 - Nucleotide Sequence for *E . coli* O86 gne1**

[0186] Locus AAO37706 BCT 06-DEC-2005

[0187] Definition UDP-GlcNAc C4-epimerase [Escherichia coli O86].

[0188] Accession AAO37706

[0189] Length: 993

[0190] Type: DNA

[0191] Organism: *E. coli* O86

[0192] Sequence:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC
AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA
ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA
AAAGCAAGTG GCAGGCGGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA
CCAACAGAAC GTTCATTAAC TATCATCCGT CCTACCGTTA TCTTCGGTGA
ACGCAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT
GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

[0193] **SEQ ID NO: 6 - Amino Acid Sequence for *E . coli* O86 UDP-GlcNAc C4-epimerase**

[0194] Locus AAO37706

[0195] Definition UDP-GlcNAc C4-epimerase [Escherichia coli O86].

[0196] Accession AAO37706

[0197] Length: 331 aa linear

[0198] Type: PRT

[0199] Organism: *E. coli* O86

[0200] Sequence:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp eitqigdvrd
qqaldqalag fdtvvllaae hrddvsptsl yydvnvqgtr nvlaameknq
vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv
gnivefikyk lknvaagyev ynyvdkpdln mnqlvaeveq slnkkipsmh
lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf
vapytlsqgl drtlqyefvh akkdditfvs e

[0201] **SEQ ID NO: 7 - Nucleotide Sequence for *Shigella boydii* O18 gne**

[0202] Locus ACD09753 BCT 05-MAY-2008

[0203] Definition UDP-N-acetylglucosamine 4-epimerase [Shigella boydii CDC 3083-94].

[0204] Accession ACD09753

[0205] Length: 993

[0206] Type: DNA

[0207] Organism: *Shigella boydii* O18

[0208] Sequence:

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC
AGCAGAGCCA TTTTTATCCA GCAATCACAC AGATTGGCGA TGTTCGTGAT
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA
ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA
AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA
CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA

ACGCAACCGC GGTAACGTCT ATAACTTGCT GAAACAGATC GCTGGCGGCA
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT
GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTATGAGGTT TATAACTATG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

[0209] **SEQ ID NO: 8 - Amino Acid Sequence for *Shigella boydii* O18 UDP-N-acetylglucosamine 4-epimerase**

[0210] Locus ACD09753

[0211] Definition UDP-N-acetylglucosamine 4-epimerase [Shigella boydii CDC 3083-94].

[0212] Accession ACD09753

[0213] Length: 331 aa linear

[0214] Type: PRT

[0215] Organism: *Shigella boydii* O18

[0216] Sequence:

mndnvlliga sgfvgtrlle taiadfnikn ldkqqshfyp aitqigdvrd
qqaldqalag fdtvvllaae hrddvsptsl yydvnvqgtr nvlaamekng
vkniiftssv avyglnkhnp denhphdpfn hygkskwqae evlrewynka
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv
gnivefikyk lknvaagyev ynyvdkpdln mnqlvaeveq slnkkipsmh
lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf
vapytlsqgl drtlqyefvh akkdditfvs e

[0217] **SEQ ID NO: 9 - Nucleotide Sequence for *Salmonella enterica* O30 *gne***

[0218] Locus AAV34516 BCT 25-OCT-2004

[0219] Definition UDP-GlcNAc 4-epimerase [Salmonella enterica subsp. salamae serovar Greenside].

[0220] Accession AAV34516

[0221] Length: 993

[0222] Type: DNA

[0223] Organism: *Salmonella enterica* O30

[0224] Sequence:

ATGAACGATA ACGTTTTGCT CATTGGTGCT TCCGGATTCG TAGGAACCCG
ACTCCTTGAA ACGGCAGTGG ATGATTTTAA TATCAAGAAC CTGGATAAAC
AGCAAAGCCA TTTCTACCCA GAGATTACAC ACATTGGCGA TGTTCGTGAC
CAACAAATCC TTGACCAGAC GTTGGTGGGT TTTGACACCG TAGTACTATT
GGCTGCGGAG CATCGTGATG ATGTTAGTCC TACCTCGCTT TATTATGATG
TCAACGTCCA GGGAACGCGT AATGTACTGG CGGCGATGGA AAAAAATGGT
GTAAAAAATA TCATTTTTAC CAGTTCCGTT GCAGTTTATG GACTCAACAA
GAAAAATCCT GACGAAACGC ACCCTCACGA TCCCTTTAAT CATTACGGAA
AAAGTAAATG GCAAGCAGAA GAAGTTCTGC GTGAGTGGCA TGCTAAAGCG
CCGAATGAGC GTTCTTTGAC CATAATTCGT CCTACCGTTA TTTTCGGGGA
GCGTAACCGC GGTAATGTAT ACAATCTCTT GAAACAGATC GCTGGTGGTA
AATTTGCGAT GGTTGGTCCG GGAACTAACT ATAAATCAAT GGCTTATGTT
GGTAATATCG TTGAGTTTAT CAAATTCAAA CTCAAGAATG TTACGGCGGG
CTATGAAGTT TATAATTATG TTGATAAACC TGATCTGAAT ATGAATCAAT
TGGTTGCTGA AGTAGAGCAG AGCCTGGGCA AAAAAATACC ATCGATGCAC
CTTCCATATC CATTAGGTAT GCTGGGGGGT TACTGTTTCG ATATCCTGAG
CAAAGTAACG GGCAAGAAGT ACGCTGTAAG TTCGGTTCGT GTTAAAAAAT
TCTGTGCGAC AACGCAGTTT GATGCAACAA AAGTGCATTC TTCTGGTTTT
GTTGCGCCAT ACACCTTATC TCAGGGGTTG GATCGTACAC TGCAATATGA
ATTTGTTCAT GCAAAGAAAG ATGACATTAC ATTCGTTTCA GAG

[0225] **SEQ ID NO: 10 - Amino Acid Sequence for *Salmonella enterica* O30 UDP-GlcNAc 4-epimerase**

[0226] Locus AAV34516

[0227] Definition UDP-GlcNAc 4-epimerase [Salmonella enterica subsp. salamae serovar Greenside].

[0228] Accession AAV34516

[0229] Length: 331 aa linear

[0230] Type: PRT

[0231] Organism: *Salmonella enterica* O30

[0232] Sequence:

```
mndnvlliga sgfvgtrlle tavddfnikn ldkqqshfyp eithigdvrd
qqildqtlvg fdtvvllaae hrddvsptsl yydvnvqgtr nvlaameknq
vkniiftssv avyglnkknp dethphdpfn hygkskwqae evlrewhaka
pnersltiir ptvifgernr gnvynllkqi aggkfamvgp gtnyksmayv
gnivefikfk lknvtagyev ynyvdkpdln mnqlvaeveq slgkkipsmh
lpyplgmlgg ycfdilskvt gkkyavssvr vkkfcattqf datkvhssgf
vapytlsqgl drtlqyefvh akkdditfvs e
```

[0233] **SEQ ID NO: 11 - Nucleotide Sequence for *C. jejuni gne***

[0234] Locus YP_002344524 BCT 14-SEP-2010

[0235] Definition UDP-GlcNAc/Glc 4-epimerase [Campylobacter jejuni subsp. jejuni

[0236] Accession YP_002344524

[0237] Length: 987

[0238] Type: DNA

[0239] Organism: *C. jejuni*

[0240] Sequence:

```
ATGAAAATTCTTATTAGCGGTGGTGCAGGTTATATAGGTTCTCATACTTTAAGACAATT
TTTAAAAACAGATCATGAAATTTGTGTTTTAGATAATCTTTCTAAGGGTTCTAAAATCG
CAATAGAAGATTTGCAAAAAACAAGAGCTTTTAAATTTTTCGAACAAGATTTAAGTGAT
TTTCAAGGCGTAAAAGCATTGTTTGAGAGAGAAAAATTTGACGCTATTGTGCATTTTGC
AGCAAGCATTGAAGTTTTTGAAAGTATGCAAAATCCTTTAAAATATTATATGAACAACA
CTGTTAATACGACAAATCTCATCGAAACTTGTTTGCAAACTGGAGTGAATAAATTTATA
```

TTTTCTTCAACGGCGGCCACTTATGGCGAACCACAAACTCCCGTTGTGAGCGAAACAAG
TCCTTTAGCACCTATTAATCCTTATGGGCGTAGTAAGCTTATGAGTGAAGAAGTTTTGC
GTGATGCAAGTATGGCAAATCCTGAATTTAAGCATTGTATTTTAAGATATTTTAATGTT
GCAGGTGCTTGTATGGATTATACTTTAGGACAACGCTATCCAAAAGCGACTTTGCTTAT
AAAAGTTGCAGCTGAATGTGCCGCAGGAAAACGTGATAAACTTTTCATATTTGGCGATG
ATTATGATACAAAAGATGGTACTTGCATAAGAGATTTTATCCATGTAGATGATATTTCA
AGTGCACATTTAGCGGCTTTGGATTATTTAAAAGAGAATGAAAGCAATGTTTTTAATGT
AGGTTATGGACATGGTTTTAGCGTAAAAGAAGTGATTGAAGCGATGAAAAAAGTTAGCG
GAGTGGATTTTAAAGTAGAACTTGCCCCACGCCGTGCGGGTGATCCTAGTGTATTGATT
TCTGATGCAAGTAAAATCAGAAATCTTACTTCTTGGCAGCCTAAATATGATGATTTAGA
GCTTATTTGTAAATCTGCTTTTGATTGGGAAAAACAGTGTTAA

[0241] **SEQ ID NO: 12 - Amino Acid Sequence for *C. jejuni* UDP-GlcNAc/Glc 4-epimerase**

[0242] Locus YP_002344524

[0243] Definition UDP-GlcNAc/Glc 4-epimerase [Campylobacter jejuni subsp. jejuni

[0244] Accession YP_002344524

[0245] Length: 328 aa linear

[0246] Type: PRT

[0247] Organism: *C. jejuni*

[0248] Sequence:

mkilisggag yigshtlrqf lktdheicvl dnlskgskia iedlqktraf
kffeqdlsdf qgvkalfere kfdaivhfaa sievfesmqn plkyymnntv
nttnlietcl qtgvnkfifs
staatygepq tpvvsetspl apinpygrsk lmseevlrda smanpefkhc
ilryfnvaga cmdytlgqry pkatllikva aecaagkrdk lfifgddydt
kdgtcirdfi hvddissahl
aaldylkene snvfnvgygh gfsvkeviea mkkvsgvdfk velaprragd
psvlisdask irnltswqpk yddlelicks
afdwekqc

[0249] **SEQ ID NO: 13 - Nucleotide Sequence for *E. coli* K12 *galE***

[0250] Locus AP_001390 BCT 30-APR-2010

[0251] Definition UDP-galactose-4-epimerase [Escherichia coli str. K-12 substr. W3110].

[0252] Accession AP_001390

[0253] Length: 1,017

[0254] Type: DNA

[0255] Organism: *E. coli* K12

[0256] Sequence:

ATGAGAGTTCTGGTTACCGGTGGTAGCGGTTACATTGGAAGTCATACCTGTGTGCAA
TTACTGCAAAACGGTCATGATGTCATCATTCTTGATAACCTCTGTAACAGTAAGCGC
AGCGTACTGCCTGTTATCGAGCGTTTAGGCGGCAAACATCCAACGTTTGTTGAAGGC
GATATTCGTAACGAAGCGTTGATGACCGAGATCCTGCACGATCACGCTATCGACAC
CGTGATCCACTTCGCCGGGCTGAAAGCCGTGGGCGAATCGGTACAAAAACCGCTGG
AATATTACGACAACAATGTCAACGGCACTCTGCGCCTGATTAGCGCCATGCGCGCC
GCTAACGTCAAAAACTTTATTTTTAGCTCCTCCGCCACCGTTTATGGCGATCAGCCC
AAAATTCCATACGTTGAAAGCTTCCCGACCGGCACACCGCAAAGCCCTTACGGCAA
AAGCAAGCTGATGGTGGAACAGATCCTCACCGATCTGCAAAAAGCCCAGCCGGACT
GGAGCATTGCCCTGCTGCGCTACTTCAACCCGGTTGGCGCGCATCCGTCGGGCGATA
TGGGCGAAGATCCGCAAGGCATTCCGAATAACCTGATGCCATACATCGCCCAGGTT
GCTGTAGGCCGTCGCGACTCGCTGGCGATTTTTGGTAACGATTATCCGACCGAAGAT
GGTACTGGCGTACGCGATTACATCCACGTAATGGATCTGGCGGACGGTCACGTCGT
GGCGATGGAAAAACTGGCGAACAAGCCAGGCGTACACATCTACAACCTCGGCGCTG
GCGTAGGCAACAGCGTGCTGGACGTGGTTAATGCCTTCAGCAAAGCCTGCGGCAAA
CCGGTTAATTATCATTTTGCACCGCGTCGCGAGGGCGACCTTCCGGCCTACTGGGCG
GACGCCAGCAAAGCCGACCGTGAACTGAACTGGCGCGTAACGCGCACACTCGATGA
AATGGCGCAGGACACCTGGCACTGGCAGTCACGCCATCCACAGGGATATCCCGATT
AA

[0257] **SEQ ID NO: 14 - Amino Acid Sequence for *E. coli* K12 UDP-galactose-4-epimerase**

[0258] Locus AP_001390

[0259] Definition UDP-galactose-4-epimerase [Escherichia coli str. K-12 substr. W3110].

[0260] Accession AP_001390

[0261] Length: 338 aa linear

[0262] Type: PRT

[0263] Organism: *E. coli* K12

[0264] Sequence:

mrvlvtggsgyigshtcvqllqnghdviildnlcnskrsvlpvierlggkhptfvegdi
rnealmteilhdhaidtvihfaglkavgesvqkpleyydnnvngtlrlisamraanvkn
fifsssatvygdqpkipyvesfptgtpqspygksklmveqiltdlqkaqpdwsiallry
fnpvgahpsgdmgedpqgipnnlmpyiaqvavgrrdslaifgndyptedgtgvrdyihv
mdladghvvameklankpgvhiynlgagvgnsvldvvnafskacgkpvnyhfaprregd
lpaywadaskadrelnwrvtrtldemaqdtwhwqsrhpqgypd

[0265] **SEQ ID NO: 15 - Nucleotide Sequence for *E. coli* O86 *gne2***

[0266] Locus AAV85952 BCT 27-MAR-2005

[0267] Definition Gne [Escherichia coli O86].

[0268] Accession AAV85952

[0269] Length: 1,020

[0270] Type: DNA

[0271] Organism: *E. coli* O86

[0272] Sequence:

ATGGTGATTT TCGTAACAGG CGGTGCAGGA TATATTGGAT CCCATACCAT
ACTTGAGTTA CTTAATAATG GTCATGATGT CGTTTCGATA GATAATTTTG
TCAATTCCTC TATAGAATCA TTAAAAAGAG TAGAGCAAAT AACTAATAAG
AAAATTATTT CTTATCAAGG TGATATCCGT GATAAAAATC TACTTGATGA
GATTTTTTCA AGACACCATA TCGATGCTGT AATTCACTTT GCATCGTTAA
AATCTGTAGG TGAGTCTAAG TTAAAGCCCT TAGAGTATTA TTCTAATAAT
GTTGGTGGAA CTTTAGTATT ACTTGAATGC ATGAAGAGAT ATAACATTAA

TAAAATGATA TTTAGCTCTT CTGCTACTGT TTATGGGAGT AACAGTATCC
CTCCCCATAC GGAAGATAGA CGAATTGGTG AAACTACAAA CCCATATGGG
ACATCGAAAT TTATAATAGA AATAATTTTG AGTGATTATT GTGATAGTGA
TAATAATAAA TCAGTAATTG CACTGCGTTA CTTTAATCCA ATCGGAGCAC
ATAAGTCCGG GATGATTGGT GAAAATCCTA ACGGGATCCC TAATAATCTG
GTTCCTTATA TATCTAAAGT TGCACAAAAT CAACTTCCTG TATTAAATAT
TTATGGCAAC GATTATCCAA CTAAAGATGG TACAGGAGTA AGAGACTATA
TACATGTCTG TGATTTGGCT AAAGGGCATG TTAAAGCATT AGAATATATG
TTTTTAAATG ATGTCAATTA TGAAGCTTTT AATTTAGGTA CTGGTCAAGG
TTATTCTGTT TTAGAGATTG TAAAAATGTT TGAGATAGTC ACTAAAAAGA
GTATACCTGT TGCTATTTGT AATAGACGTG AGGGGGATGT TGCGGAGTCA
TGGGCGTCTG CTGATTTGGC ACATAAAAAG CTTTCCTGGA AAGCGGAAAA
AAATTTGAAA GAAATGATCG AAGATGTATG GCGTTGGCAA ACAAACAATC
CAAATGGATA TAAAAAATAA

[0273] **SEQ ID NO: 16 - Amino Acid Sequence for *E. coli* O86 Gne**

[0274] Locus AAV85952

[0275] Definition Gne [Escherichia coli O86].

[0276] Accession AAV85952

[0277] Length: 339 aa (gne2) linear

[0278] Type: PRT

[0279] Organism: *E. coli* O86

[0280] Sequence:

mvifvtggag yigshtilel lnnghdvvsi dnfvnssies lkrveqitnk
kiisyqgdir dknlldeifs rhhidavihf aslksvgesk lkpleyysnn
vggtlvllec mkryninkmi fsssatvygs nsipphtedr rigettnpyg
tskfiieiil sdycdsdnnk svialryfnp igahksgmig enpngipnnl
vpyiskvaqn qlpvlniygn dyptkdgtgv rdyihvcdla kghvkaleym
flndvnyeaf nlgtgqgysv leivkmfeiv tkksipvaic nrregdvaes
wasadlahkk lswkaeknlk emiedvwrwq tnnpngykk

[0281] **SEQ ID NO: 17 - Nucleotide Sequence for synthetic oligonucleotide Z3206-Fw (primer) encoding an end of Z3206; restriction sites underlined**

[0282] Length: 30

[0283] Type: DNA

[0284] Sequence: AAACCCGGGATGAACGATAACGTTTTGCTC

[0285] **SEQ ID NO: 18 - Nucleotide Sequence for synthetic oligonucleotide Z3206-RvHA (primer) encoding an end of Z3206 with a hemoaglutinin tag (HA tag); restriction sites underlined**

[0286] Length: 60

[0287] Type: DNA

[0288] Organism:

[0289] Sequence: AAATCTAGATTAAGCGTAATCTGGAAC ATCGTATGGGTACTCAGAAACAAACGTTATGTC

[0290] **SEQ ID NO: 19 - Nucleotide Sequence for synthetic oligonucleotide gne-Fw (primer) with restriction sites underlined**

[0291] Length: 29

[0292] Type: DNA

[0293] Organism:

[0294] Sequence: AAACCATGGATGAAAATTCTTATTAGCGG

[0295] **SEQ ID NO: 20 - Nucleotide Sequence for synthetic oligonucleotide gne-RV (primer) with restriction sites underlined**

[0296] Length: 57

[0297] Type: DNA

[0298] Organism:

[0299] Sequence: AAATCTAGATTAAGCGTAATCTGGAAC ATCGTATGGGTAGCACTGTTTTTCCCAATC

[0300] **SEQ ID NO: 21 - Nucleotide Sequence for oligonucleotide containing restriction sites for NheI restriction enzyme**

[0301] Length: 11

[0302] Type: DNA

[0303] Organism:

[0304] Sequence: AAAAAGCTAGC

[0305] **SEQ ID NO: 22 - Nucleotide Sequence for oligonucleotide containing restriction sites for AscI restriction enzyme**

[0306] Length: 8

[0307] Type: DNA

[0308] Organism:

[0309] Sequence: CCGCGCGG

[0310] **SEQ ID NO: 23 - Nucleotide Sequence for plasmid pMLBAD:Z3206 (*E. coli* O157 insert in plasmid) encoding Z3206 with a C-terminal hemagglutinin tag**

[0311] Definition Ligation of product into Z3206-pMLBAD*

[0312] Features Location/Qualifiers

[0313] CDS 2105..3098

/label=Z3206

[0314] CDS 3098..3127

/label=HA

[0315] Length: 7794 bp

[0316] Type: DNA circular UNA

[0317] Sequence:

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA

61 TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG GTAGGCAGTC

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG CTTCCCCTGA GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC AAGTCGTCGG

541 TTGGTATTGC ACAAAACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT CCCACCCAGG

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA TATCTTAGGA

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT ATCATCGATG

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCCTATGC TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA TCATTCACTT

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC GACGGTGGCG ATAGGCATCC

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCCTCGCGC CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC GCTTCCATGC

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG CGCCCTTCCC

1321 CTTGCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA ATGCGGCTGG TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA GTTAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC GCGAGCCTCC GGATGACGAC

1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC ACCCGGTCGG CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAAT CGAGATAACC GTTGGCCTCA ATCGGCGTTA

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG CAGGGGATCA TTTTGCGCTT

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAACCG GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC ACACTTTGCT

1981 ATGCCATAGC ATTTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT TTATCGCAAC

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA CCATGGTACC

2101 CGGGATGAAC GATAACGTTT TGCTCATAGG AGCTTCCGGA TTCGTAGGAA CCCGACTACT

2161 TGAAACGGCA ATTGCTGACT TTAATATCAA GAACCTGGAC AAACAGCAGA GCCACTTTTA

2221 TCCAGAAATC ACACAGATTG GCGATGTTCG CGATCAACAG GCACTCGACC AGGCGTTAGT

2281 CGGTTTTGAC ACTGTTGTAC TACTGGCAGC GGAACACCGC GATGACGTCA GCCCTACTTC

2341 TCTCTATTAT GATGTCAACG TTCAGGGTAC CCGCAATGTG CTGGCGGCCA TGGAAAAAAA

2401 TGGCGTTAAA AATATCATCT TTACCAGTTC CGTTGCTGTT TATGGTTTGA ACAAACACAA

2461 CCCTGACGAA AACCATCCAC ACGACCCTTT CAACCACTAC GGCAAAAGTA AGTGGCAGGC

2521 AGAGGAAGTG CTGCGTGAAT GGTATAACAA AGCACCAACA GAACGTTCAT TAACCATCAT

2581 CCGTCCTACC GTTATCTTCG GTGAACGCAA CCGCGGTAAC GTCTATAACT TGCTGAAACA

2641 GATCGCTGGC GGCAAGTTTA TGATGGTGGG CGCAGGGACT AACTATAAGT CCATGGCTTA

2701 TGTTGGAAAC ATTGTTGAGT TTATCAAGTA CAAACTGAAG AATGTTGCCG CAGGTTATGA

2761 GGTTTATAAC TACGTTGATA AGCCAGACCT GAACATGAAC CAGTTGGTTG CTGAAGTTGA

2821 ACAAAGCCTG AACAAAAAGA TCCCTTCTAT GCACTTGCCT TACCCACTAG GAATGCTGGG

2881 TGGATATTGC TTTGATATCC TGAGCAAAAT TACGGGCAAA AAATACGCTG TCAGCTCAGT

2941 GCGCGTGAAA AAATTCTGCG CAACAACACA GTTTGACGCA ACGAAAGTGC ATTCTTCAGG

3001 TTTTGTGGCA CCGTATACGC TGTCGCAAGG TCTGGATCGA ACACTGCAGT ATGAATTCGT

3061 TCATGCCAAA AAAGACGACA TAACGTTTGT TTCTGAGTAC CCATACGATG TTCCAGATTA

3121 CGCTTAATCT AGAGTCGACC TGCAGGCATG CAAGCTTGGC TGTTTTGGCG GATGAGAGAA

3181 GATTTTCAGC CTGATACAGA TTAAATCAGA ACGCAGAAGC GGTCTGATAA AACAGAATTT

3241 GCCTGGCGGC AGTAGCGCGG TGGTCCCACC TGACCCCATG CCGAACTCAG AAGTGAAACG

3301 CCGTAGCGCC GATGGTAGTG TGGGGTCTCC CCATGCGAGA GTAGGGAACT GCCAGGCATC

3361 AAATAAAACG AAAGGCTCAG TCGAAAGACT GGGCCTTTCG TTTTATCTGT TGTTTGTCGG

3421 TGAACGCTCT CCTGAGTAGG ACAAATCCGC CGGGAGCGGA TTTGAACGTT GCGAAGCAAC

3481 GGCCCGGAGG GTGGCGGGCA GGACGCCCGC CATAAACTGC CAGGCATCAA ATTAAGCAGA

3541 AGGCCATCCT GACGGATGGC CTTTTTGCGT TTCTACAAAC TCTTCCACTC ACTACAGCAG

3601 AGCCATTTAA ACAACATCCC CTCCCCCTTT CCACCGCGTC AGACGCCCGT AGCAGCCCGC

3661 TACGGGCTTT TTCATGCCCT GCCCTAGCGT CCAAGCCTCA CGGCCGCGCT CGGCCTCTCT

3721 GGCGGCCTTC TGGCGCTGAG GTCTGCCTCG TGAAGAAGGT GTTGCTGACT CATACCAGGC

3781 CTGAATCGCC CCATCATCCA GCCAGAAAGT GAGGGAGCCA CGGTTGATGA GAGCTTTGTT

3841 GTAGGTGGAC CAGTTGGTGA TTTTGAACTT TTGCTTTGCC ACGGAACGGT CTGCGTTGTC

3901 GGGAAGATGC GTGATCTGAT CCTTCAACTC AGCAAAAGTT CGATTTATTC AACAAAGCCG

3961 CCGTCCCGTC AAGTCAGCGT AATGCTCTGC CAGTGTTACA ACCAATTAAC CAATTCTGAT

4021 TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT TCATATCAGG ATTATCAATA

4081 CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA ACTCACCGAG GCAGTTCCAT

4141 AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC GTCCAACATC AATACAACCT

4201 ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGCGAGA AATCACCATG AGTGACGACT

4261 GAATCCGGTG AGAATGGCAA AAGCTAAAAA GGCCGTAATA TCCAGCTGAA CGGTCTGGTT

4321 ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT GCCATTGGGA

4381 TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT TAGCTCCTGA

4441 AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTTCATTAT GGTGAAAGTT

4501 GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC AGGGCTTCCC

4561 GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG TCACAGGTAT

4621 TTATTCGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT ATAGGTTAAT GTCATGATAA

4681 TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGGAAA TGTGCGCGCC CGCGTTCCTG

4741 CTGGCGCTGG GCCTGTTTCT GGCGCTGGAC TTCCCGCTGT TCCGTCAGCA GCTTTTCGCC

4801 CACGGCCTTG ATGATCGCGG CGGCCTTGGC CTGCATATCC CGATTCAACG GCCCCAGGGC

4861 GTCCAGAACG GGCTTCAGGC GCTCCCGAAG GTCTCGGGCC GTCTCTTGGG CTTGATCGGC

4921 CTTCTTGCGC ATCTCACGCG CTCCTGCGGC GGCCTGTAGG GCAGGCTCAT ACCCCTGCCG

4981 AACCGCTTTT GTCAGCCGGT CGGCCACGGC TTCCGGCGTC TCAACGCGCT TTGAGATTCC

5041 CAGCTTTTCG GCCAATCCCT GCGGTGCATA GGCGCGTGGC TCGACCGCTT GCGGGCTGAT

5101 GGTGACGTGG CCCACTGGTG GCCGCTCCAG GGCCTCGTAG AACGCCTGAA TGCGCGTGTG

5161 ACGTGCCTTG CTGCCCTCGA TGCCCCGTTG CAGCCCTAGA TCGGCCACAG CGGCCGCAAA

5221 CGTGGTCTGG TCGCGGGTCA TCTGCGCTTT GTTGCCGATG AACTCCTTGG CCGACAGCCT

5281 GCCGTCCTGC GTCAGCGGCA CCACGAACGC GGTCATGTGC GGGCTGGTTT CGTCACGGTG

5341 GATGCTGGCC GTCACGATGC GATCCGCCCC GTACTTGTCC GCCAGCCACT TGTGCGCCTT

5401 CTCGAAGAAC GCCGCCTGCT GTTCTTGGCT GGCCGACTTC CACCATTCCG GGCTGGCCGT

5461 CATGACGTAC TCGACCGCCA ACACAGCGTC CTTGCGCCGC TTCTCTGGCA GCAACTCGCG

5521 CAGTCGGCCC ATCGCTTCAT CGGTGCTGCT GGCCGCCCAG TGCTCGTTCT CTGGCGTCCT

5581 GCTGGCGTCA GCGTTGGGCG TCTCGCGCTC GCGGTAGGCG TGCTTGAGAC TGGCCGCCAC

5641 GTTGCCCATT TTCGCCAGCT TCTTGCATCG CATGATCGCG TATGCCGCCA TGCCTGCCCC

5701 TCCCTTTTGG TGTCCAACCG GCTCGACGGG GGCAGCGCAA GGCGGTGCCT CCGGCGGGCC

5761 ACTCAATGCT TGAGTATACT CACTAGACTT TGCTTCGCAA AGTCGTGACC GCCTACGGCG

5821 GCTGCGGCGC CCTACGGGCT TGCTCTCCGG GCTTCGCCCT GCGCGGTCGC TGCGCTCCCT

5881 TGCCAGCCCG TGGATATGTG GACGATGGCC GCGAGCGGCC ACCGGCTGGC TCGCTTCGCT

5941 CGGCCCGTGG ACAACCCTGC TGGACAAGCT GATGGACAGG CTGCGCCTGC CCACGAGCTT

6001 GACCACAGGG ATTGCCCACC GGCTACCCAG CCTTCGACCA CATACCCACC GGCTCCAACT

6061 GCGCGGCCTG CGGCCTTGCC CCATCAATTT TTTTAATTTT CTCTGGGGAA AAGCCTCCGG

6121 CCTGCGGCCT GCGCGCTTCG CTTGCCGGTT GGACACCAAG TGGAAGGCGG GTCAAGGCTC

6181 GCGCAGCGAC CGCGCAGCGG CTTGGCCTTG ACGCGCCTGG AACGACCCAA GCCTATGCGA

6241 GTGGGGGCAG TCGAAGGCGA AGCCCGCCCG CCTGCCCCCC GAGCCTCACG GCGGCGAGTG

6301 CGGGGGTTCC AAGGGGGCAG CGCCACCTTG GGCAAGGCCG AAGGCCGCGC AGTCGATCAA

6361 CAAGCCCCGG AGGGGCCACT TTTTGCCGGA GGGGGAGCCG CGCCGAAGGC GTGGGGGAAC

6421 CCCGCAGGGG TGCCCTTCTT TGGGCACCAA AGAACTAGAT ATAGGGCGAA ATGCGAAAGA

6481 CTTAAAAATC AACAACTTAA AAAAGGGGGG TACGCAACAG CTCATTGCGG CACCCCCCGC

6541 AATAGCTCAT TGCGTAGGTT AAAGAAAATC TGTAATTGAC TGCCACTTTT ACGCAACGCA

6601 TAATTGTTGT CGCGCTGCCG AAAAGTTGCA GCTGATTGCG CATGGTGCCG CAACCGTGCG

6661 GCACCCTACC GCATGGAGAT AAGCATGGCC ACGCAGTCCA GAGAAATCGG CATTCAAGCC

6721 AAGAACAAGC CCGGTCACTG GGTGCAAACG GAACGCAAAG CGCATGAGGC GTGGGCCGGG

6781 CTTATTGCGA GGAAACCCAC GGCGGCAATG CTGCTGCATC ACCTCGTGGC GCAGATGGGC

6841 CACCAGAACG CCGTGGTGGT CAGCCAGAAG ACACTTTCCA AGCTCATCGG ACGTTCTTTG

6901 CGGACGGTCC AATACGCAGT CAAGGACTTG GTGGCCGAGC GCTGGATCTC CGTCGTGAAG

6961 CTCAACGGCC CCGGCACCGT GTCGGCCTAC GTGGTCAATG ACCGCGTGGC GTGGGGCCAG

7021 CCCCGCGACC AGTTGCGCCT GTCGGTGTTC AGTGCCGCCG TGGTGGTTGA TCACGACGAC

7081 CAGGACGAAT CGCTGTTGGG GCATGGCGAC CTGCGCCGCA TCCCGACCCT GTATCCGGGC

7141 GAGCAGCAAC TACCGACCGG CCCCGGCGAG GAGCCGCCCA GCCAGCCCGG CATTCCGGGC

7201 ATGGAACCAG ACCTGCCAGC CTTGACCGAA ACGGAGGAAT GGGAACGGCG CGGGCAGCAG

7261 CGCCTGCCGA TGCCCGATGA GCCGTGTTTT CTGGACGATG GCGAGCCGTT GGAGCCGCCG

7321 ACACGGGTCA CGCTGCCGCG CCGGTAGCAC TTGGGTTGCG CAGCAACCCG TAAGTGCGCT

7381 GTTCCAGACT ATCGGCTGTA GCCGCCTCGC CGCCCTATAC CTTGTCTGCC TCCCCGCGTT

7441 GCGTCGCGGT GCATGGAGCC GGGCCACCTC GACCTGAATG GAAGCCGGCG GCACCTCGCT

7501 AACGGATTCA CCGTTTTTAT CAGGCTCTGG GAGGCAGAAT AAATGATCAT ATCGTCAATT

7561 ATTACCTCCA CGGGGAGAGC CTGAGCAAAC TGGCCTCAGG CATTTGAGAA GCACACGGTC

7621 ACACTGCTTC CGGTAGTCAA TAAACCGGTA AACCAGCAAT AGACATAAGC GGCTATTTAA

7681 CGACCCTGCC CTGAACCGAC GACCGGGTCG AATTTGCTTT CGAATTTCTG CCATTCATCC

7741 GCTTATTATC ACTTATTCAG GCGTAGCACC AGGCGTTTAA GTCGACCAAT AACC

[0318] **SEQ ID NO: 24 - Nucleotide Sequence for pMLBAD:gne (*E. coli* O157 insert in plasmid) which encodes Gne with a C-terminal hemagglutinin tag**

[0319] Locus gne-pMLBAD

[0320] Definition Ligation of dig galE into pmlbad did (NcoI-XbaI)

[0321] Features Location/Qualifiers

[0322] CDS 2097..3080

/label=galE

[0323] CDS 3081..3107

/label=HA

[0324] Region 3108..3110

/label=stop

[0325] Length: 7776 bp

[0326] Type: DNA circular UNA

[0327] Sequence:

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC CCAGTTGACA

61 TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG CAACTGGTCC

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTCAT GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG GTAGGCAGTC

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG CTTCCCCTGA GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC AAGTCGTCGG

541 TTGGTATTGC ACAAAACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT CCCACCCAGG

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA TATCTTAGGA

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT ATCATCGATG

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCCTATGC TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA TCATTCACTT

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC GACGGTGGCG ATAGGCATCC

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCCTCGCGC CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC GCTTCCATGC

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG CGCCCTTCCC

1321 CTTGCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA ATGCGGCTGG TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA GTTAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC GCGAGCCTCC GGATGACGAC

1501 CGTAGTGATG AATCTCTCCT GGCGGGAACA GCAAAATATC ACCCGGTCGG CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAAT CGAGATAACC GTTGGCCTCA ATCGGCGTTA

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG CAGGGGATCA TTTTGCGCTT

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAACCG GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC ACACTTTGCT

1981 ATGCCATAGC ATTTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT TTATCGCAAC

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA CCATGGATGA

2101 AAATTCTTAT TAGCGGTGGT GCAGGTTATA TAGGTTCTCA TACTTTAAGA CAATTTTTAA

2161 AAACAGATCA TGAAATTTGT GTTTTAGATA ATCTTTCTAA GGGTTCTAAA ATCGCAATAG

2221 AAGATTTGCA AAAAATAAGA ACTTTTAAAT TTTTTGAACA AGATTTAAGT GATTTTCAAG

2281 GCGTAAAAGC ATTGTTTGAG AGAGAAAAAT TTGACGCTAT TGTGCATTTT GCAGCGAGCA

2341 TTGAAGTTTT TGAAAGTATG CAAAACCCTT TAAAGTATTA TATGAATAAC ACTGTTAATA

2401 CGACAAATCT CATCGAAACT TGTTTGCAAA CTGGAGTGAA TAAATTTATA TTTTCTTCAA

2461 CGGCAGCCAC TTATGGCGAA CCACAAACTC CCGTTGTGAG CGAAACAAGT CCTTTAGCAC

2521 CTATTAATCC TTATGGGCGT AGTAAGCTTA TGAGCGAAGA GGTTTTGCGT GATGCAAGTA

2581 TGGCAAATCC TGAATTTAAG CATTGTATTT TAAGATATTT TAATGTTGCA GGTGCTTGCA

2641 TGGATTATAC TTTAGGACAA CGCTATCCAA AAGCGACTTT GCTTATAAAA GTTGCAGCTG

2701 AATGTGCCGC AGAAAAACGT AATAAACTTT TCATATTTGG CGATGATTAT GATACAAAAG

2761 ATGGCACTTG CATAAGAGAT TTTATCCATG TGGATGATAT TTCAAGTGCG CATTTATCGG

2821 CTTTGGATTA TTTAAAAGAG AATGAAAGCA ATGTTTTTAA TGTAGGTTAT GGACATGGTT

2881 TTAGCGTAAA AGAAGTGATT GAAGCGATGA AAAAAGTTAG CGGAGTGGAT TTTAAAGTAG

2941 AACTTGCCCC ACGCCGTGCG GGTGATCCTA GTGTATTGAT TTCTGATGCA AGTAAAATCA

3001 GAAATCTTAC TTCTTGGCAG CCTAAATATG ATGATTTAGG GCTTATTTGT AAATCTGCTT

3061 TTGATTGGGA AAAACAGTGC TACCCATACG ATGTTCCAGA TTACGCTTAA TCTAGAGTCG

3121 ACCTGCAGGC ATGCAAGCTT GGCTGTTTTG GCGGATGAGA GAAGATTTTC AGCCTGATAC

3181 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAAACAGAA TTTGCCTGGC GGCAGTAGCG

3241 CGGTGGTCCC ACCTGACCCC ATGCCGAACT CAGAAGTGAA ACGCCGTAGC GCCGATGGTA

3301 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC ATCAAATAAA ACGAAAGGCT

3361 CAGTCGAAAG ACTGGGCCTT TCGTTTTATC TGTTGTTTGT CGGTGAACGC TCTCCTGAGT

3421 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC AACGGCCCGG AGGGTGGCGG

3481 GCAGGACGCC CGCCATAAAC TGCCAGGCAT CAAATTAAGC AGAAGGCCAT CCTGACGGAT

3541 GGCCTTTTTG CGTTTCTACA AACTCTTCCA CTCACTACAG CAGAGCCATT TAAACAACAT

3601 CCCCTCCCCC TTTCCACCGC GTCAGACGCC CGTAGCAGCC CGCTACGGGC TTTTTCATGC

3661 CCTGCCCTAG CGTCCAAGCC TCACGGCCGC GCTCGGCCTC TCTGGCGGCC TTCTGGCGCT

3721 GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG ACTCATACCA GGCCTGAATC GCCCCATCAT

3781 CCAGCCAGAA AGTGAGGGAG CCACGGTTGA TGAGAGCTTT GTTGTAGGTG GACCAGTTGG

3841 TGATTTTGAA CTTTTGCTTT GCCACGGAAC GGTCTGCGTT GTCGGGAAGA TGCGTGATCT

3901 GATCCTTCAA CTCAGCAAAA GTTCGATTTA TTCAACAAAG CCGCCGTCCC GTCAAGTCAG

3961 CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA ACTCATCGAG

4021 CATCAAATGA AACTGCAATT TATTCATATC AGGATTATCA ATACCATATT TTTGAAAAAG

4081 CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC CATAGGATGG CAAGATCCTG

4141 GTATCGGTCT GCGATTCCGA CTCGTCCAAC ATCAATACAA CCTATTAATT TCCCCTCGTC

4201 AAAAATAAGG TTATCAAGCG AGAAATCACC ATGAGTGACG ACTGAATCCG GTGAGAATGG

4261 CAAAAGCTAA AAAGGCCGTA ATATCCAGCT GAACGGTCTG GTTATAGGTA CATTGAGCAA

4321 CTGACTGAAA TGCCTCAAAA TGTTCTTTAC GATGCCATTG GGATATATCA ACGGTGGTAT

4381 ATCCAGTGAT TTTTTTCTCC ATTTTAGCTT CCTTAGCTCC TGAAAATCTC GATAACTCAA

4441 AAAATACGCC CGGTAGTGAT CTTATTTCAT TATGGTGAAA GTTGGAACCT CTTACGTGCC

4501 GATCAACGTC TCATTTTCGC CAAAAGTTGG CCCAGGGCTT CCCGGTATCA ACAGGGACAC

4561 CAGGATTTAT TTATTCTGCG AAGTGATCTT CCGTCACAGG TATTTATTCG AAGACGAAAG

4621 GGCCTCGTGA TACGCCTATT TTTATAGGTT AATGTCATGA TAATAATGGT TTCTTAGACG

4681 TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GCCCGCGTTC CTGCTGGCGC TGGGCCTGTT

4741 TCTGGCGCTG GACTTCCCGC TGTTCCGTCA GCAGCTTTTC GCCCACGGCC TTGATGATCG

4801 CGGCGGCCTT GGCCTGCATA TCCCGATTCA ACGGCCCCAG GGCGTCCAGA ACGGGCTTCA

4861 GGCGCTCCCG AAGGTCTCGG GCCGTCTCTT GGGCTTGATC GGCCTTCTTG CGCATCTCAC

4921 GCGCTCCTGC GGCGGCCTGT AGGGCAGGCT CATACCCCTG CCGAACCGCT TTTGTCAGCC

4981 GGTCGGCCAC GGCTTCCGGC GTCTCAACGC GCTTTGAGAT TCCCAGCTTT TCGGCCAATC

5041 CCTGCGGTGC ATAGGCGCGT GGCTCGACCG CTTGCGGGCT GATGGTGACG TGGCCCACTG

5101 GTGGCCGCTC CAGGGCCTCG TAGAACGCCT GAATGCGCGT GTGACGTGCC TTGCTGCCCT

5161 CGATGCCCCG TTGCAGCCCT AGATCGGCCA CAGCGGCCGC AAACGTGGTC TGGTCGCGGG

5221 TCATCTGCGC TTTGTTGCCG ATGAACTCCT TGGCCGACAG CCTGCCGTCC TGCGTCAGCG

5281 GCACCACGAA CGCGGTCATG TGCGGGCTGG TTTCGTCACG GTGGATGCTG GCCGTCACGA

5341 TGCGATCCGC CCCGTACTTG TCCGCCAGCC ACTTGTGCGC CTTCTCGAAG AACGCCGCCT

5401 GCTGTTCTTG GCTGGCCGAC TTCCACCATT CCGGGCTGGC CGTCATGACG TACTCGACCG

5461 CCAACACAGC GTCCTTGCGC CGCTTCTCTG GCAGCAACTC GCGCAGTCGG CCCATCGCTT

5521 CATCGGTGCT GCTGGCCGCC CAGTGCTCGT TCTCTGGCGT CCTGCTGGCG TCAGCGTTGG

5581 GCGTCTCGCG CTCGCGGTAG GCGTGCTTGA GACTGGCCGC CACGTTGCCC ATTTTCGCCA

5641 GCTTCTTGCA TCGCATGATC GCGTATGCCG CCATGCCTGC CCCTCCCTTT TGGTGTCCAA

5701 CCGGCTCGAC GGGGGCAGCG CAAGGCGGTG CCTCCGGCGG GCCACTCAAT GCTTGAGTAT

5761 ACTCACTAGA CTTTGCTTCG CAAAGTCGTG ACCGCCTACG GCGGCTGCGG CGCCCTACGG

5821 GCTTGCTCTC CGGGCTTCGC CCTGCGCGGT CGCTGCGCTC CCTTGCCAGC CCGTGGATAT

5881 GTGGACGATG GCCGCGAGCG GCCACCGGCT GGCTCGCTTC GCTCGGCCCG TGGACAACCC

5941 TGCTGGACAA GCTGATGGAC AGGCTGCGCC TGCCCACGAG CTTGACCACA GGGATTGCCC

6001 ACCGGCTACC CAGCCTTCGA CCACATACCC ACCGGCTCCA ACTGCGCGGC CTGCGGCCTT

6061 GCCCCATCAA TTTTTTTAAT TTTCTCTGGG GAAAAGCCTC CGGCCTGCGG CCTGCGCGCT

6121 TCGCTTGCCG GTTGGACACC AAGTGGAAGG CGGGTCAAGG CTCGCGCAGC GACCGCGCAG

6181 CGGCTTGGCC TTGACGCGCC TGGAACGACC CAAGCCTATG CGAGTGGGGG CAGTCGAAGG

6241 CGAAGCCCGC CCGCCTGCCC CCCGAGCCTC ACGGCGGCGA GTGCGGGGGT TCCAAGGGGG

6301 CAGCGCCACC TTGGGCAAGG CCGAAGGCCG CGCAGTCGAT CAACAAGCCC CGGAGGGGCC

6361 ACTTTTTGCC GGAGGGGGAG CCGCGCCGAA GGCGTGGGGG AACCCCGCAG GGGTGCCCTT

6421 CTTTGGGCAC CAAAGAACTA GATATAGGGC GAAATGCGAA AGACTTAAAA ATCAACAACT

6481 TAAAAAAGGG GGGTACGCAA CAGCTCATTG CGGCACCCCC CGCAATAGCT CATTGCGTAG

6541 GTTAAAGAAA ATCTGTAATT GACTGCCACT TTTACGCAAC GCATAATTGT TGTCGCGCTG

6601 CCGAAAAGTT GCAGCTGATT GCGCATGGTG CCGCAACCGT GCGGCACCCT ACCGCATGGA

6661 GATAAGCATG GCCACGCAGT CCAGAGAAAT CGGCATTCAA GCCAAGAACA AGCCCGGTCA

6721 CTGGGTGCAA ACGGAACGCA AAGCGCATGA GGCGTGGGCC GGGCTTATTG CGAGGAAACC

6781 CACGGCGGCA ATGCTGCTGC ATCACCTCGT GGCGCAGATG GGCCACCAGA ACGCCGTGGT

6841 GGTCAGCCAG AAGACACTTT CCAAGCTCAT CGGACGTTCT TTGCGGACGG TCCAATACGC

6901 AGTCAAGGAC TTGGTGGCCG AGCGCTGGAT CTCCGTCGTG AAGCTCAACG GCCCCGGCAC

6961 CGTGTCGGCC TACGTGGTCA ATGACCGCGT GGCGTGGGGC CAGCCCCGCG ACCAGTTGCG

7021 CCTGTCGGTG TTCAGTGCCG CCGTGGTGGT TGATCACGAC GACCAGGACG AATCGCTGTT

7081 GGGGCATGGC GACCTGCGCC GCATCCCGAC CCTGTATCCG GGCGAGCAGC AACTACCGAC

7141 CGGCCCCGGC GAGGAGCCGC CCAGCCAGCC CGGCATTCCG GGCATGGAAC CAGACCTGCC

7201 AGCCTTGACC GAAACGGAGG AATGGGAACG GCGCGGGCAG CAGCGCCTGC CGATGCCCGA

7261 TGAGCCGTGT TTTCTGGACG ATGGCGAGCC GTTGGAGCCG CCGACACGGG TCACGCTGCC

7321 GCGCCGGTAG CACTTGGGTT GCGCAGCAAC CCGTAAGTGC GCTGTTCCAG ACTATCGGCT

7381 GTAGCCGCCT CGCCGCCCTA TACCTTGTCT GCCTCCCCGC GTTGCGTCGC GGTGCATGGA

7441 GCCGGGCCAC CTCGACCTGA ATGGAAGCCG GCGGCACCTC GCTAACGGAT TCACCGTTTT

7501 TATCAGGCTC TGGGAGGCAG AATAAATGAT CATATCGTCA ATTATTACCT CCACGGGGAG

7561 AGCCTGAGCA AACTGGCCTC AGGCATTTGA GAAGCACACG GTCACACTGC TTCCGGTAGT

7621 CAATAAACCG GTAAACCAGC AATAGACATA AGCGGCTATT TAACGACCCT GCCCTGAACC

7681 GACGACCGGG TCGAATTTGC TTTCGAATTT CTGCCATTCA TCCGCTTATT ATCACTTATT

7741 CAGGCGTAGC ACCAGGCGTT TAAGTCGACC AATAAC

[0328] **SEQ ID NO: 25 - Amino Acid Sequence for modified EPA with signal sequence**

[0329] Disclosed in WO 2009/104074 (as SEQ ID NO. 6)

[0330] Type: PRT

[0331] Organism: Artificial

[0332] /note="Description of Artificial Sequence: Synthetic polypeptide"

[0333] Length: 643

[0334] Sequence:

```
Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1               5                  10                  15

Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys
            20                  25                  30

Ala Cys Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser
        35                  40                  45

Val Asp Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr
    50                  55                  60

Ser Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp
65                  70                  75                  80

Asn Ala Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly
                85                  90                  95

Gly Val Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala
            100                 105                 110

Arg Gly Ser Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys
        115                 120                 125

Pro Ser Asn Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln
    130                 135                 140

Leu Ser His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu
145                 150                 155                 160
```

Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu
165 170 175

Ser Asn Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser
180 185 190

Val Val Met Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu
195 200 205

Trp Ala Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val
210 215 220

Tyr Asn Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu
225 230 235 240

Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu
245 250 255

Asp Ile Lys Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg
260 265 270

Leu His Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln
275 280 285

Ala Cys His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg
290 295 300

Gly Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val
305 310 315 320

Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val
325 330 335

Ile Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu
340 345 350

Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala

355 360 365

Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu
370 375 380

Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala
385 390 395 400

Lys Asp Gln Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser
405 410 415

Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu
420 425 430

Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp
435 440 445

Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly
450 455 460

Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser
465 470 475 480

Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile
485 490 495

Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr
500 505 510

Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala
515 520 525

Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg
530 535 540

Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg
545 550 555 560

```
Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro
                565                 570                 575

Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu
            580                 585                 590

Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val
        595                 600                 605

Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile
    610                 615                 620

Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu
625                 630                 635                 640

Asp Leu Lys
```

[0335] **SEQ ID NO: 26 - Amino Acid Sequence for PglB**

[0336] Disclosed in WO 2009/104074 (as SEQ ID NO. 2)

[0337] Length: 722

[0338] Type: PRT

[0339] Organism: *Campylobacter jejuni*

[0340] Sequence:

```
Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala
1               5                   10                  15

Met Ile Ile Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp
            20                  25                  30

Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gln Leu
        35                  40                  45

Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp
```

50 55 60

Met Ile Ala Gly Phe His Gln Pro Asn Asp Leu Ser Tyr Tyr Gly Ser
65 70 75 80

Ser Leu Ser Ala Leu Thr Tyr Trp Leu Tyr Lys Ile Thr Pro Phe Ser
85 90 95

Phe Glu Ser Ile Ile Leu Tyr Met Ser Thr Phe Leu Ser Ser Leu Val
100 105 110

Val Ile Pro Thr Ile Leu Leu Ala Asn Glu Tyr Lys Arg Pro Leu Met
115 120 125

Gly Phe Val Ala Ala Leu Leu Ala Ser Ile Ala Asn Ser Tyr Tyr Asn
130 135 140

Arg Thr Met Ser Gly Tyr Tyr Asp Thr Asp Met Leu Val Ile Val Leu
145 150 155 160

Pro Met Phe Ile Leu Phe Phe Met Val Arg Met Ile Leu Lys Lys Asp
165 170 175

Phe Phe Ser Leu Ile Ala Leu Pro Leu Phe Ile Gly Ile Tyr Leu Trp
180 185 190

Trp Tyr Pro Ser Ser Tyr Thr Leu Asn Val Ala Leu Ile Gly Leu Phe
195 200 205

Leu Ile Tyr Thr Leu Ile Phe His Arg Lys Glu Lys Ile Phe Tyr Ile
210 215 220

Ala Val Ile Leu Ser Ser Leu Thr Leu Ser Asn Ile Ala Trp Phe Tyr
225 230 235 240

Gln Ser Ala Ile Ile Val Ile Leu Phe Ala Leu Phe Ala Leu Glu Gln
245 250 255

Lys Arg Leu Asn Phe Met Ile Ile Gly Ile Leu Gly Ser Ala Thr Leu
260 265 270

Ile Phe Leu Ile Leu Ser Gly Gly Val Asp Pro Ile Leu Tyr Gln Leu
275 280 285

Lys Phe Tyr Ile Phe Arg Ser Asp Glu Ser Ala Asn Leu Thr Gln Gly
290 295 300

Phe Met Tyr Phe Asn Val Asn Gln Thr Ile Gln Glu Val Glu Asn Val
305 310 315 320

Asp Leu Ser Glu Phe Met Arg Arg Ile Ser Gly Ser Glu Ile Val Phe
325 330 335

Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser
340 345 350

Met Ile Met Ala Leu Pro Ile Leu Val Leu Gly Phe Leu Ala Leu Lys
355 360 365

Gly Gly Leu Arg Phe Thr Ile Tyr Ser Val Pro Val Met Ala Leu Gly
370 375 380

Phe Gly Phe Leu Leu Ser Glu Phe Lys Ala Ile Met Val Lys Lys Tyr
385 390 395 400

Ser Gln Leu Thr Ser Asn Val Cys Ile Val Phe Ala Thr Ile Leu Thr
405 410 415

Leu Ala Pro Val Phe Ile His Ile Tyr Asn Tyr Lys Ala Pro Thr Val
420 425 430

Phe Ser Gln Asn Glu Ala Ser Leu Leu Asn Gln Leu Lys Asn Ile Ala
435 440 445

Asn Arg Glu Asp Tyr Val Val Thr Trp Ala Ala Tyr Gly Tyr Pro Val
450 455 460

Arg Tyr Tyr Ser Asp Val Lys Thr Leu Val Asp Gly Gly Lys His Leu
465 470 475 480

Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gln
485 490 495

Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser
500 505 510

Phe Tyr Ala Pro Gln Asn Asp Ile Leu Lys Thr Asp Ile Leu Gln Ala
515 520 525

Met Met Lys Asp Tyr Asn Gln Ser Asn Val Asp Leu Phe Leu Ala Ser
530 535 540

Leu Ser Lys Pro Asp Phe Lys Ile Asp Thr Pro Lys Thr Arg Asp Ile
545 550 555 560

Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu Ile Phe Ser Thr Val Ala
565 570 575

Ser Phe Ser Phe Ile Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe
580 585 590

Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu Ile Tyr
595 600 605

Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys Ile
610 615 620

Gly Asp Asn Val Val Ser Val Asn Ser Ile Val Glu Ile Asn Ser Ile
625 630 635 640

Lys Gln Gly Glu Tyr Lys Ile Thr Pro Ile Asp Asp Lys Ala Gln Phe
645 650 655

```
Tyr Ile Phe Tyr Leu Lys Asp Ser Ala Ile Pro Tyr Ala Gln Phe Ile
            660                 665                 670

Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gln Met Phe Phe
        675                 680                 685

Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val Ile Asn Ser Arg
    690                 695                 700

Asp Ala Lys Val Phe Lys Leu Lys Ile Tyr Pro Tyr Asp Val Pro Asp
705                 710                 715                 720

Tyr Ala
```

[0341] **SEQ ID NO: 27 - Nucleotide Sequence for pCC1FOS Empty plasmid**

[0342] Locus pCC1FOS with MCS cassette

[0343] Features Location/Qualifiers

[0344] Region 230..256

[0345] /label="pCC1/pEpiFOS fwd"

[0346] Region 311..330

[0347] /label="T7 promoter"

[0348] Region complement(504..529)

[0349] /label="pCC1pEpiFOS rv"

[0350] CDS complement(805..1464)

[0351] /label=cat

[0352] CDS 1683..2030

[0353] /label=redF

[0354] CDS 3425..4180

[0355] /label=repE

[0356] CDS 4759..5934

[0357] /label=parA

[0358] CDS 5934..6905

[0359] /label=parB

[0360] ORIGIN

[0361] Length: 8171 bp

[0362] Type: DNA circular UNA

[0363] Organism: Artificial

[0364] Sequence:

1 GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG
CTTAACTATG
61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
CCGCACAGAT
121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA
ACTGTTGGGA
181 AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG GCGAAAGGGG
GATGTGCTGC
241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA
AAACGACGGC
301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC
GGGGATCCCA
361 CGTGGCGCGC CACTAGTGCT AGCGACGTCG TGGGATCCTC TAGAGTCGAC
CTGCAGGCAT
421 GCAAGCTTGA GTATTCTATA GTCTCACCTA AATAGCTTGG CGTAATCATG
GTCATAGCTG
481 TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC
CGGAAGCATA
541 AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC
GTTGCGCTCA
601 CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT
CGGCCAACGC
661 GAACCCCTTG CGGCCGCCCG GGCCGTCGAC CAATTCTCAT GTTTGACAGC
TTATCATCGA
721 ATTTCTGCCA TTCATCCGCT TATTATCACT TATTCAGGCG TAGCAACCAG
GCGTTTAAGG
781 GCACCAATAA CTGCCTTAAA AAAATTACGC CCCGCCCTGC CACTCATCGC
AGTACTGTTG
841 TAATTCATTA AGCATTCTGC CGACATGGAA GCCATCACAA ACGGCATGAT
GAACCTGAAT
901 CGCCAGCGGC ATCAGCACCT TGTCGCCTTG CGTATAATAT TTGCCCATGG
TGAAAACGGG
961 GGCGAAGAAG TTGTCCATAT TGGCCACGTT TAAATCAAAA CTGGTGAAAC
TCACCCAGGG

1021 ATTGGCTGAG ACGAAAAACA TATTCTCAAT AAACCCTTTA GGGAAATAGG
CCAGGTTTTC
1081 ACCGTAACAC GCCACATCTT GCGAATATAT GTGTAGAAAC TGCCGGAAAT
CGTCGTGGTA
1141 TTCACTCCAG AGCGATGAAA ACGTTTCAGT TTGCTCATGG AAAACGGTGT
AACAAGGGTG
1201 AACACTATCC CATATCACCA GCTCACCGTC TTTCATTGCC ATACGAAATT
CCGGATGAGC
1261 ATTCATCAGG CGGGCAAGAA TGTGAATAAA GGCCGGATAA AACTTGTGCT
TATTTTTCTT
1321 TACGGTCTTT AAAAAGGCCG TAATATCCAG CTGAACGGTC TGGTTATAGG
TACATTGAGC
1381 AACTGACTGA AATGCCTCAA AATGTTCTTT ACGATGCCAT TGGGATATAT
CAACGGTGGT
1441 ATATCCAGTG ATTTTTTTCT CCATTTTAGC TTCCTTAGCT CCTGAAAATC
TCGATAACTC
1501 AAAAAATACG CCCGGTAGTG ATCTTATTTC ATTATGGTGA AAGTTGGAAC
CTCTTACGTG
1561 CCGATCAACG TCTCATTTTC GCCAAAAGTT GGCCCAGGGC TTCCCGGTAT
CAACAGGGAC
1621 ACCAGGATTT ATTTATTCTG CGAAGTGATC TTCCGTCACA GGTATTTATT
CGCGATAAGC
1681 TCATGGAGCG GCGTAACCGT CGCACAGGAA GGACAGAGAA AGCGCGGATC
TGGGAAGTGA
1741 CGGACAGAAC GGTCAGGACC TGGATTGGGG AGGCGGTTGC CGCCGCTGCT
GCTGACGGTG
1801 TGACGTTCTC TGTTCCGGTC ACACCACATA CGTTCCGCCA TTCCTATGCG
ATGCACATGC
1861 TGTATGCCGG TATACCGCTG AAAGTTCTGC AAAGCCTGAT GGGACATAAG
TCCATCAGTT
1921 CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC TGCCCGGCAC
CGGGTGCAGT
1981 TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA ATTATCCTGA
GAATAAATGC
2041 CTTGGCCTTT ATATGGAAAT GTGGAACTGA GTGGATATGC TGTTTTTGTC
TGTTAAACAG
2101 AGAAGCTGGC TGTTATCCAC TGAGAAGCGA ACGAAACAGT CGGGAAAATC
TCCCATTATC
2161 GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG TTTATAGGAA
GTAGTGTTCT
2221 GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA TGCCTTCAGG
AACAATAGAA
2281 ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT CAGCAATGGA
CAGAACAACC
2341 TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG CCAGTAGTGC
TCGCCGCAGT
2401 CGAGCGACAG GGCGAAGCCC TCGGCTGGTT GCCCTCGCCG CTGGGCTGGC
GGCCGTCTAT
2461 GGCCCTGCAA ACGCGCCAGA AACGCCGTCG AAGCCGTGTG CGAGACACCG
CGGCCGGCCG

2521 CCGGCGTTGT GGATACCTCG CGGAAAACTT GGCCCTCACT GACAGATGAG GGGCGGACGT
2581 TGACACTTGA GGGGCCGACT CACCCGGCGC GGCGTTGACA GATGAGGGGC AGGCTCGATT
2641 TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC GAAAACGCCT GATTTTACGC
2701 GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT GCCCTGCGGT ATTGACACTT
2761 GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG ACACTTGAGG GGCAGAGTGC
2821 TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG TCCACAGGCA GAAAATCCAG
2881 CATTTGCAAG GGTTTCCGCC CGTTTTTCGG CCACCGCTAA CCTGTCTTTT AACCTGCTTT
2941 TAAACCAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC GCCCTGTGCG CGTGACCGCG
3001 CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCCTCCCG GTCGAGTGAG CGAGGAAGCA
3061 CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC CTGAAAAAAC TTCCCTTGGG
3121 GTTATCCACT TATCCACGGG GATATTTTTA TAATTATTTT TTTTATAGTT TTTAGATCTT
3181 CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT CTAGAGAAGG TGTTGTGACA
3241 AATTGCCCTT TCAGTGTGAC AAATCACCCT CAAATGACAG TCCTGTCTGT GACAAATTGC
3301 CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT TTTCACAAAG TTATCCCTGC
3361 TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAAACT TGTCACACTT CACATGGATC
3421 TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT AAAAATAGCC CGCGAATCGT
3481 CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC CGGGATCAAA AACGTATGCT
3541 GTATCTGTTC GTTGACCAGA TCAGAAAATC TGATGGCACC CTACAGGAAC ATGACGGTAT
3601 CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA TTGACCTCTG CGGAAGCCAG
3661 TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG GAAGTGGTTT TTTATCGCCC
3721 TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT CCTTGGTTTA TCAAACGTGC
3781 GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC CCATATCTCA TTCCCTTCTT
3841 TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT GAAACAAAAG AAATCACCAA
3901 TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT CGTAAGCCGG ATGGCTCAGG
3961 CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC CAGCTGCCTC AAAGTTACCA

4021 GCGTATGCCT GACTTCCGCC GCCGCTTCCT GCAGGTCTGT GTTAATGAGA
TCAACAGCAG
4081 AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC CGCCAGACGA
CTCATATCGT
4141 ATTTTCCTTC CGCGATATCA CTTCCATGAC GACAGGATAG TCTGAGGGTT
ATCTGTCACA
4201 GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA GGGTAATTTG
TCACAGTTTT
4261 GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT GAACTGTAAT
TTTTAAGGAA
4321 GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCTTCT CTTTCCCTTC
GTCATGTGAC
4381 CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG ATTATCACAG
TTTATTACTC
4441 TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT TCCCACGGTG
GATATTTCTT
4501 CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT CTTTGCTTCC
TCGCCAGTTC
4561 GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT GATAATAAGT
GACTGAGGTA
4621 TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT TAAACAACTT
TGCGGTTTTT
4681 TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG CAAGATTTAA
TAAAAAAACG
4741 CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAACTCATGG AAACACTTAA
CCAGTGCATA
4801 AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC AGTTTAATGA
TGACAGCCCG
4861 GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG CAGCGGATTT
AGTTGGGGTT
4921 TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC TACCGCACCC
GGATATGGAA
4981 ATTCGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG AACAAATTAA
TCATATGCGT
5041 GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT TTCCACCGGT
GATCGGGGTT
5101 GCTGCCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG TTCATCTTGC
TCAGGATCTG
5161 GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG ACCCCCAGGG
AACAGCCTCA
5221 ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG AAGACACTCT
CCTGCCTTTC
5281 TATCTTGGGG AAAAGGACGA TGTCACTTAT GCAATAAAGC CCACTTGCTG
GCCGGGGCTT
5341 GACATTATTC CTTCCTGTCT GGCTCTGCAC CGTATTGAAA CTGAGTTAAT
GGGCAAATTT
5401 GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC GACTGGCCAT
TGAAACTGTT
5461 GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA ACCTGGGTAT
CGGCACGATT

5521 AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC CTGCTGAGTT
GTTTGACTAC
5581 ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC TCAAGAACGT
TGATCTTAAA
5641 GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA GCAATAGTAA
TGGCTCTCAG
5701 TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA GCATGGTTCT
AAAAAATGTT
5761 GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA TGAGAACTGT
TTTTGAACAG
5821 GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG CTCTTTCTAT
TTGGGAACCT
5881 GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT GGGAGATTAG
ATAATGAAGC
5941 GTGCGCCTGT TATTCCAAAA CATACGCTCA ATACTCAACC GGTTGAAGAT
ACTTCGTTAT
6001 CGACACCAGC TGCCCCGATG GTGGATTCGT TAATTGCGCG CGTAGGAGTA
ATGGCTCGCG
6061 GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA GTTTACTCTT
GAAGTGCTCC
6121 GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG TAATGAACGT
GACCAGGAGC
6181 TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT TCTACTGACT
GGTCAACAGA
6241 CACCGGCGTT CGGTCGAAGA GTATCTGGTG TCATAGAAAT TGCCGATGGG
AGTCGCCGTC
6301 GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTTCT GGTTGGCGAG
CTGGATGATG
6361 AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG CCCAACAAGT
GCTTATGAAC
6421 GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC TGGAAATATT
TCTGCGCTGG
6481 CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG TATCAACACC
GCCAAATTGC
6541 CTAAATCAGT TGTTGCTCTT TTTTCTCACC CCGGTGAACT ATCTGCCCGG
TCAGGTGATG
6601 CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA GCAGCAGGCA
TCTAACCTTC
6661 ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA AGTTATCACT
CTTTTAACTT
6721 CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAG CTCACGACAT
CAGTTTGCTC
6781 CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT TAACCTGGAC
AGGTCTCGTG
6841 TTCCAACTGA GTGTATAGAG AAAATTGAGG CCATTCTTAA GGAACTTGAA
AAGCCAGCAC
6901 CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT ACTTAATGTC
CTTTGTTACA
6961 GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC CACTGTTCCA
CTTGTATCGT

7021 CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT CGTCGGTCTG
ATTATTAGTC
7081 TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA GTCTGGGACC
ACGGTCCCAC
7141 TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC CACTCGTATC
GTCGGTCTGA
7201 TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC TGATTATTAG
TCTGGGACCA
7261 CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA CCACGGTCCC
ACTCGTATCG
7321 TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA TCGTCGGTCT
GATTATTAGT
7381 CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC GGTCTGGGAC
CACGGTCCCA
7441 CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT CCATCAATGC
CTGTCAAGGG
7501 CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA CCTCGGTGTG
CGGTTGTATG
7561 CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT ATCCACAACA TTTTGCGCAC
GGTTATGTGG
7621 ACAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC CGGGCTGCAT
CCGATGCAAG
7681 TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT TGCCCCGTAT
TCAGTGTCGC
7741 TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT GCAGATCAAT
TAATACGATA
7801 CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT CCACGCACGT
TGTGATATGT
7861 AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA TCCGACAGGT
TACGGGGCGG
7921 CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT TTAAGGCGTT
TCCGTTCTTC
7981 TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA AAAGAAAGGA
AACGACAGGT
8041 GCTGAAAGCG AGCTTTTTGG CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC
CGTGGAATGA
8101 ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG CATACATTAT
ACGAAGTTAT
8161 ATTCGATCCA C

[0365] **SEQ ID NO: 28 - Nucleotide Sequence for pCC1FOS cut (pFOS) and *S. flexneri* 6 O-antigen without Z3206**

[0366] Locus pFOS cut and O-antige cut (-Z3206)

[0367] Definition Ligation of inverted pCC1FOS with MCS cassette cut with NheI and AscI into S.flexneri 6 O antigen cluster amplified with galFNheI and wzzAscI cut with NheI and AscI

[0368] FEATURES Location/Qualifiers

[0369] CDS 3..411

[0370] /label='galF

[0371] CDS 784..1869

[0372] /label=rmlB

[0373] CDS 1869..2768

[0374] /label=rmlD

[0375] CDS 2826..3704

[0376] /label=rmlA

[0377] CDS 3709..4266

[0378] /label=rmlC

[0379] CDS 4263..5495

[0380] /label=wzx

[0381] CDS 5551..6738

[0382] /label=wzy

[0383] CDS 6755..7624

[0384] /label=wfbY

[0385] CDS 7621..8454

[0386] /label=wfbZ

[0387] CDS 8559..9965

[0388] /label=gnd

[0389] CDS 10187..11380

[0390] /label=ugd

[0391] CDS complement(11446..12450)

[0392] /label=uge

[0393] CDS 12802..12828

[0394] /label=wzz'

[0395] Region complement(12868..12887)

[0396] /label="T7 promoter"

[0397] Region complement(12942..12968)

[0398] /label="pCC1/pEpiFOS fwd"

[0399] CDS complement(14460..15431)

[0400] /label=parB

[0401] CDS complement(15431..16606)

[0402] /label=parA

[0403] CDS complement(17185..17940)

[0404] /label=repE

[0405] CDS complement(19335..19682)

[0406] /label=redF

[0407] CDS 19901..20560

[0408] /label=cat

[0409] Region 20836..20861

[0410] /label="pCC1pEpiFOS rv"

[0411] Length: 20982 bp

[0412] Type: DNA circular UNA

[0413] Sequence:

1 CTAGCGGCAA AACGTATGCC GGGTGACCTC TCTGAATACT CCGTCATCCA GACCAAAGAA

61 CCGCTGGATC GCGAAGGTAA AGTCAGCCGC ATTGTTGAAT TTATCGAAAA ACCGGATCAG

121 CCGCAGACGC TGGACTCAGA CATCATGGCC GTTGGTCGCT ATGTGCTTTC TGCCGATATT

181 TGGCCGGAAC TTGAACGTAC TCAGCCTGGT GCATGGGGAC GTATTCAGCT GACTGATGCC

241 ATTGCCGAGC TGGCGAAAAA ACAGTCCGTT GATGCAATGC TGATGACCGG CGACAGCTAC

301 GACTGCGGTA AAAAAATGGG CTATATGCAG GCGTTTGTGA AGTATGGGCT GCGCAACCTG

361 AAAGAAGGGG CGAAGTTCCG TAAAGGTATT GAGAAGCTGT TAAGCGAATA ATGAAAATCT

421 GACCGGATGT AACGGTTGAT AAGAAAATTA TAACGGCAGT GAAGATTCGT GGTGAAAGTA

481 ATTTGTTGCG AATATTCCTG CCGTTGTTTT ATATAAACAA TCAGAATAAC AACGAGTTAG

541 CAATAGGATT TTAGTCAAAG TTTTCCAGGA TTTTCCTTGT TTCCAGAGCG GATTGGTAAG

601 ACAATTAGCT TTTGAATTTT TCGGGTTTAG CGCGAGTGGG TAACGCTCGT CACATCGTAG

661 GCATGCATGC AGTGCTCTGG TAGCTGTAAA GCCAGGGGCG GTAGCGTGCA TTAATACTTC

721 TATTAATCAA ACTGAGAGCC GCTTATTTCA CAGCATGCTC TGAAGCAATA TGGAATAAAT

781 TAGGTGAAAA TACTTGTTAC TGGTGGCGCA GGATTTATTG GTTTTGCTGT AGTTCGTCAC

841 ATTATAAATA ATACGCAGGA TAGTGTTGTT AATGTCGATA AATTAACGTA CGCCGGAAAC

901 CTGGAATCAC TTGCTGATGT TTCTGATTCT GAACGCTATG TTTTTGAACA TGCGGATATT

961 TGCGATGCAG CTGCAATGGC ACGGATTTTT GCTCAGCATC AGCCAGATGC AGTGATGCAC

1021 CTGGCTGCTG AAAGCCATGT TGACCGTTCA ATTACAGGTC CTGCGGCATT TATTGAAACC

1081 AATATTGTTG GTACATATGT CCTTTTGGAA GCCGCTCGCA ATTATTGGTC TGCTCTTGAT

1141 AGCGACAAGA AAACTAGATT CCGTTTTCAT CATATTTCTA CTGACGAAGT CTATGGTGAT

1201 TTGCCTCATC CTGACGAGGT AAATAATACA GAAGAATTAC CCTTATTTAC AGAGACAACA

1261 GCTTACGCGC CAAGCAGCCC TTATTCCGCT TCAAAAGCAT CCAGCGATCA TTTAGTCCGC

1321 GCGTGGAAAC GTACCTATGG TTTACCAACC ATTGTGACTA ATTGCTCTAA TAATTATGGT

1381 CCTTATCATT TCCCGGAAAA ATTGATTCCA TTGGTTATTC TGAATGCTCT GGAAGGTAAG

1441 GCATTACCTA TTTATGGCAA AGGGGATCAA ATTCGTGACT GGCTGTATGT TGAAGATCAT

1501 GCGCGTGCGT TATATACCGT CGTAACCGAA GGTAAAGCGG GTGAAACTTA TAACATTGGT

1561 GGACACAACG AAAAGAAAAA CATCGATGTA GTGCTCACTA TTTGTGATTT GCTGGATGAG

1621 ATTGTACCGA AAGAGAAATC TTACCGCGAG CAAATTACTT ATGTTGCCGA TCGCCCGGGA

1681 CACGATCGCC GTTATGCGAT TGATGCAGAG AAGATTAGCC GCGAATTGGG CTGGAAACCG

1741 CAGGAAACGT TTGAGAGCGG GATTCGGAAG ACATTGGAAT GGTACCTGTC CAATACAAAA

1801 TGGGTTGATA ATGTGAAAAG TGGTGCTTAT CAATCGTGGA TTGAACAGAA CTATGAGGGC

1861 CGCCAGTAAT GAATATCCTC CTTTTCGGCA AAACAGGGCA GGTAGGTTGG GAACTACAGC

1921 GTGCTCTGGC ACCTTTGGGT AATTTGATTG CTCTTGATGT TCACTCCACT GATTATTGTG

1981 GTGATTTTAG TAATCCTGAA GGTGTAGCTG AAACAGTCAA AAGAATTCGA CCTGATGTTA

2041 TTGTTAATGC TGCGGCTCAC ACCGCAGTAG ATAAGGCTGA GTCAGAACCC GAATTTGCAC

2101 AATTACTCAA TGCGACTAGT GTTGAATCAA TTGCAAAAGA GGCTAATGAA GTTGGGGCTT

2161 GGGTAATTCA TTACTCAACT GACTACGTAT TCCCTGGAAA TGGCGACACG CCATGGCTGG

2221 AGACGGATGC AACCGCACCG CTAAATGTTT ACGGTGAAAC CAAGTTAGCC GGAGAAAAAG

2281 CGTTACAGGA ACATTGCGCG AAGCATCTTA TTTTCCGTAC CAGCTGGGTA TACGCAGCTA

2341 AAGGAAATAA CTTCGCCAAA ACGATGTTGC GTCTGGCAAA AGAGCGCGAA GAACTGGCTG

2401 TGATAAATGA TCAATTTGGT GCGCCAACAG GTGCTGAGCT GCTGGCTGAT TGTACGGCAC

2461 ATGCTATTCG TGTGGCACTG AATAAACCGG AAGTCGCAGG TTTGTACCAT CTGGTAGCCA

2521 GTGGTACCAC AACCTGGCAC GATTATGCTG CGCTGGTTTT TGAAGAGGCG CGCAAAGCAG

2581 GTATTCCCCT TGCACTCAAC AAGCTCAACG CAGTACCAAC AACAGCCTAT CCTACACCAG

2641 CTCGTCGTCC ACATAACTCT CGCCTTAATA CAGAAAAATT TCAGCAGAAC TTTGCGCTTG

2701 TCTTGCCTGA CTGGCAGGTT GGTGTGAAAC GAATGCTCAA CGAATTAATT ACGACTACAG

2761 CAATTTAATA GTTTTTGCAT CTTGTTCGTG ATGGTGGAGC AAGATGAATT AAAAGGAATG

2821 ATGAAATGAA AACGCGTAAA GGTATTATTT TAGCGGGTGG TTCTGGTACA CGTCTTTATC

2881 CTGTGACTAT GGCTGTCAGT AAACAGCTAT TACCTATTTA TGATAAGCCG ATGATCTATT

2941 ACCCGCTCTC TACACTGATG TTGGCGGGTA TTCGCGATAT TCTGATTATT AGTACGCCAC

3001 AGGATACTCC TCGTTTTCAA CAACTGCTAG GTGACGGTAG CCAGTGGGGG CTAAATCTTC

3061 AGTACAAAGT GCAACCGACT CCAGATGGGC TTGCGCAGGC GTTTATTATC GGTGAAGAGT

3121 TTATCGGTGG TGATGATTGT GCTTTGGTTC TTGGTGATAA TATCTTCTAC GGTCATGATC

3181 TGCCGAAGTT AATGGATGTC GCTGTTAACA AAGAAAGTGG TGCAACGGTA TTTGCCTATC

3241 ACGTTAATGA TCCTGAACGC TACGGCGTCG TTGAGTTTGA TAAAAACGGT ACGGCAATAA

3301 GCCTGGAAGA AAAACCGCTA CAACCAAAAA GTAATTATGC GGTAACCGGG CTTTATTTCT

3361 ATGATAACGA CGTTGTCGAA ATGGCGAAAA ACCTTAAGCC TTCTGCCCGT GGTGAACTGG

3421 AAATTACCGA TATTAACCGT ATTTATATGG AACAGGGGCG TTTATCCGTT GCCATGATGG

3481 GGCGTGGTTA TGCATGGCTG GATACGGGGA CACATCAGAG TCTTATTGAA GCAAGCAACT

3541 TCATTGCCAC CATTGAAGAG CGCCAGGGAC TAAAGGTTTC CTGCCCAGAA GAAATTGCTT

3601 ACCGTAAAGG GTTTATTGAT GCTGAACAGG TGAAAGCATT AGCGGAGCCG CTGAAAAAAA

3661 ATGCTTATGG ACAGTATCTG CTGAAAATGA TTAAAGGTTA TTAATAAAAT GAACGTAATT

3721 AAAACAGAAA TTCCTGATGT GTTAATTTTC GAGCCGAAAG TTTTTGGTGA TGAGCGTGGT

3781 TTCTTTATGG AAAGCTTTAA TCAGAAAGTT TTCGAAGAAG CTGTAGGACG TAAGGTTGAA

3841 TTTGTTCAGG ATAACCATTC GAAGTCTAGT AAAGGTGTTT TACGCGGGCT GCATTATCAG

3901 TTAGAACCTT ATGCGCAAGG GAAACTGGTA CGTTGCGTTG TTGGTGAGGT TTTTGATGTA

3961 GCTGTTGATA TTCGTAAATC GTCGCCTACC TTTGGTAAAT GGGTTGGGGT GAATTTATCT

4021 GCTGAGAATA AGCGGCAATT GTGGATCCCT GAGGGATTTG CACATGGTTT TTTGGTGCTG

4081 AGCGAGACTG CGGAATTTTT ATATAAAACG ACGAACTATT ATCATCCTGA TAGTGATAGA

4141 GGGATTGTAT GGAATGATCC TATTCTGAGC ATAAAATGGC CGACGATAGA ACATAATAAT

4201 TATATTTTAT CGATTAAAGA TGCAAGGGCT AAAGAATTGC ATAACATGAA GGAATTATTT

4261 TTGTGAGTAT TGTAAAGAAT ACTTTATGGA ATATAAGTGG GTATATTATA CCATCATTAA

4321 TAGCAATTCC TGCGTTAGGT ATACTGTCTA GAATTCTAGG GACCGAGCAA TTTGGCCTTT

4381 TTACGTTAGC TATTGCCTTA GTTGGATATG CAAGTATTTT TGATGCTGGA TTGACCAGAG

4441 CTGTTATAAG AGAAGTATCA ATATATAAAA ATGTTCATAA AGAATTAAGA GCGATCATTT

4501 CAACTTCAAC GGTAATTCTA ACTATATTGG GCTTGATTGG CGGTAGTGTA CTATTTTTGA

4561 GTAGCAATGT AATTGTTAAA TTATTAAACA TTAACGCGAA TCATGTTGTA GAATCTGTCA

4621 AAGCAATATA TATTATTTCA GCTACCATAC CCTTATACTT GTTAAACCAA GTCTGGTTGG

4681 GGATTTTTGA GGGGATGGAA AAGTTCAGAA AAGTAAATTT AATAAAATCA ATTAACAACT

4741 CTTTTGTGGC TGGATTACCA GTGATTTTCT GTTTTTTTCA TGGAGGATTA CTAAGTGCTA

4801 TATATGGTTT AGTTATGGCA AGAGTCTTAT CACTTATAGT GACCTTTATA TTTAGTCGAA

4861 AACTAATAAT ATCATCTGGG CTGTCTGTAA AAATTGTAAC AGTTAAAAGA TTAATCGGCT

4921 TTGGAAGCTG GATAACAGTT AGCAATATTA TTAGCCCTAT TATGACATAT ATGGATCGTT

4981 TTATTCTTTC ACACATTGTG GGGGCTGATA AAGTTTCTTT TTATACTGCT CCGTCTGAAG

5041 GTATACAACG CTTAACGATA TTACCAAGTG CGTTGTCCAG AGCTATTTTT CCAAGATTAA

5101 GTTCAGAATT GCAATCGGTA AAGCAAACTA AAATATTATC ATATTTTATA ATGGTTATTG

5161 GTATACTTCC AATTGTAATG TTGATAATTA TTTTATCAGA TTTTATAATG TCCGCTTGGA

5221 TGGGACCTAC ATATCATGGG ACGCCAGGTA TAGTATTAAA AATTCTTGCA ATAGGTTTCT

5281 TTTTTAATTG CATTGCACAA ATCCCATTTG TTTCAGTTCA GGCTAGTGGA AGATCAAAAA

5341 TTACAGCTAT TATTCATTTG CTCGAAGTTA TCCCATATTT ATGCATATTA TATATTTTTA

5401 TTTATCATTG GGGAATTGTT GGAGCCGCAA TAGCATGGTC TGTAAGAACA TCGTTAGATT

5461 TTTTGATATT ATTATTAATT GATACGAAAT ATTAATAGCG AATTGATTTT AGGGATTACT

5521 TCCTCAAGCC CATCTAATTA GAGTGCAAAC ATGACTTCTG ATTTTTATAA CTCAAAAGAC

5581 AAAAGTTTAA GTGTTCTTTT GTTTTTTGGG TTTATATTTT TCCTTACACG TAGCTTTCCA

5641 TTTATTCAAT ATAGTTGGAT TATGGAGGGG TTTTTATGTC TTTGTATCAT GTCATTTACA

5701 AAGAAAATTG CAAACGGAAT ATATCACTAT CCTGTTATTT TAATATTTCT ATTAGCTCTT

5761 TTTATAAATT TTATTTATTC CTATATCAAG GGTAACGATA TAGCGATAAT AATTAGGTTT

5821 TATATTATCA TATTATTTAT ATTATGTGCT TATTTCTGCT CTTATGGAAC CATCTCGATT

5881 GTTAAAATAT TTTTATATTT AATGGTATTA CAGGCGGTTA TTATATCCAT CATTAGTATT

5941 TATATGACAA AAACATATGG TATTGGTGAT TATTCAGCAC TAAGACATTA TTTTTTGGAG

6001 AATGATTATG GTGATGTTTA TACATATGGA AGTGGTTTCT ATAGAGTTCA AATTAAAGGA

6061 AATGCTCTCA TTCCATTTGC CTTTATGTTG CATATAGTCA TAAAAGATTA TTTCTATTAT

6121 CGATTCAAAA ATACAATAAC CGTTATTCTG GCTATAGGTA CTATAGTGGC TGGTAATTTT

6181 GCATATTTTG TTTCGATATG CTTGTTTTTT ATGTATATTA TACTATGTTC TAAATCTAAC

6241 TCACGATACG CTAAATTAAG GAAAATTATT TTTGGGGTTT TTCTTACTGT GATTCTCCCT

6301 TTTTTTATTA CATATTCAAT TGAGTTGATA ATCATGAAAT CAAATGGAGC TGATTCTTCT

6361 TTAGGAGTTA GATGGGATCA GTTTACTGTA TTAATTAATG ATCTTACAGA GTCTGTATCA

6421 AATTTTGTTA TAGGTTCTGG TTTGGGTAAT GTCATCAAAA TTCAAACTCC TATCCGTGAT

6481 TATAGTGCAT ATATATATTA TGAATTGCAG TCAGTTTATT TTTTAAATCA ACTTGGCGTT

6541 ATTTTATTTA CTTTGTTTTT ATTAATTAAT CTCCTTCTCA CGATTAAAAT CATAAAATAC

6601 AGTGAGTTGT GTGTGCTATA TTTTCTATAT GTTTCTTATG CAATTACTAA TCCTTATATT

6661 TTAGACTCTA ACCATGTTGC TGTAATAATT GTATTAGTGA CATTAAGTAA TGTTCTAAAA

6721 AAGATGAAAG CTAAATGAAG GTTTTAAGGT GAAGATGGAC ACTGTATATG CCGTTTTGGT

6781 TGCTTACAAC CCAGAACATA ATGATTTAAA AAATGCGGTT GAATTATTGT TGAGACAAGT

6841 TACTAAAGTT GTCGTTTGCA ATAACTCTAC AAATGGTTAT AAATATGCTG AAAATTCTTC

6901 AGGCGATGTA AAAATATTCA ATTTCAATGA TAATTTAGGC ATAGCAGAAG CCCAAAGTAT

6961 AGGAATGAAA TGGGCTTTTG AAAATGGCGC TGATTTTATA TTGCAAATGG ATCAGGATAG

7021 TATTCCTGAT CCTAAGATGG TAGAGCAGTT ACTTACTTGT TACAAAAAAT TGCTTAAACA

7081 AAATGTCAAT GTTGGTTTAG TTGGTTCACA AGATTTTGAT AAAGTAACTG GTGAATTAAA

7141 TAAAGCAAGG GTAAAAAAAG GGAAACCACT TACAGAAGTT TATTATGAGG TAGATAGTAC

7201 ATTAAGTTCT GGCAGTCTAA TACCAAAAAA TAGTTGGTTG ATTGTTGGAG GAATGAAAGA

7261 TGAGCTTTTT ATCGATGCGG TAGACCATGA ATATTGTTGG AGATTAAGAG CTGCTGGGTT

7321 TAAAGTAATT AGGAATAAAA ATGCGTTACT TGCACATAGA CTTGGAGATG GGCGATTTAA

7381 GATCTTAAAT ATTCTTTCTG TCGGTTTGCC AAGCCCATTT CGTCATTATT ATGCTACTCG

7441 AAATATCTTT CTTTTATTAA ATAAAAATTA TGTACCCATC TACTGGAAAA TTTCTAGTCT

7501 GGTTAAATTA ATTGGAAAGG TTTTTTTATA TCCTATTTTC CTTCCAAATG GTAATAAAAG

7561 GTTATATTTT TTTTTAAAAG GCATTAATGA CGGTTTAATG GGTCGAAGTG GTAAAATGAA

7621 ATGAATCATA GATTAGAAAA ATTCTCAGTT TTAATTAGCA TTTATAAAAA TGATCTACCG

7681 CAATTTTTTG AGGTGGCTCT ACGCTCTATT TTTCACGATC AAACACTTAA GCCAGATCAA

7741 ATAGTAATTG TTGCAGATGG AGAACTCCAT CAAACACACA TCGATATTAT AAATTCATTC

7801 ATTGATGATG TTGGCAATAA AATAGTAACA TTTGTACCTT TACCTAGAAA TGTTGGATTG

7861 GCTAATGCCT TAAATGAAGG ATTAAAGGCT TGTAGGAATG AGTTAGTGGC AAGAATGGAT

7921 GCTGATGATA TTTCTTTGCC TCATCGGTTT GAGAAACAAA TTTCTTTTAT GATTAATAAT

7981 TCAGAAATAG ATGTATGTGG CAGTTTTATT GATGAAATTG AAACTGTTAC TGAGGAGTTT

8041 ATTTCAACAC GCAAAGTGCC TCTCGAACAT AGAGAAATAG TTAAATTCGC GAGGAAACGA

8101 AGCGCAGTTA GCCATCCTTC TGTAATTTTT AGAAAGAATA CAGTATTAGC TGTTGGTGGT

8161 TATCCTCCAT TCAGAAAATC TCAAGATTTT GCATTGTGGA GCCTATTAAT TGTACATAAT

8221 GCAAGATTTG CAAATCTTCC AGATATTTTA TTAAAAATGC GAACTGGTCG TAATCTTATG

8281 GCTCGACGTG GATTGTCATA TTTATTGTAC GAGTATAAAG TATTGTATTA TCAATATAAA

8341 ATTGGTTTTA TTCGAAAAAA TGAATTAATA AGTAATGCTA TGTTGAGAAC ATTTTTTCGT

8401 ATAATGCCAT CTAAATTAAA GGAGCTGATG TATTCAATCG TTAGGAATCG ATAATAATAA

8461 TTTTCTGATT AAGTGTTATG GATTTATTTT TATTAGGCAT ATTCTATAAT TAAGCATAAC

8521 CCGCATACCA CCCAGCGGTA TCCTGACAGG AGTAAACAAT GTCAAAGCAA CAGATCGGCG

8581 TCGTCGGTAT GGCAGTGATG GGGCGCAACC TTGCGCTCAA TATCGAAAGC CGTGGTTATA

8641 CCGTCTCTAT TTTCAACCGT TCCCGTGAAA AGACCGAAGA AGTGATTACC GAAAATCCAG

8701 GCAAGAAACT GGTTCCTTAC TATACGGTGA AAGAATTTGT TGAATCTCTG GAAACGCCTC

8761 GTCGCATCCT GTTAATGGTG AAAGCAGGTG CTGGCACGGA TGCTGCTATT GATTCCCTCA

8821 AGCCATACCT CGATAAAGGT GACATCATCA TTGATGGTGG TAACACCTTC TTCCATGACA

8881 CCATTCGTCG TAACCGTGAG CTTTCTGCAG AAGGCTTTAA CTTTATCGGT ACCGGTGTTT

8941 CCGGTGGTGA AGAAGGTGCG CTGAAAGGTC CTTCCATTAT GCCTGGTGGG CAGAAAGAAG

9001 CTTATGAACT GATTGCGCCG ATCCTGACCA AAATCGCCGC TGTGGCTGAA GACGGCGAAC

9061 CGTGCGTTAC CTATATTGGT GCCGATGGTG CAGGTCATTA TGTGAAGATG GTTCACAACG

9121 GTATTGAATA CGGTGATATG CAGCTGATTG CTGAAGCCTA TTCTCTGCTT AAAGGTGGCT

9181 TGAACCTCAC CAACGAAGAA CTGGCGCAGA CCTTTACCGA GTGGAATAAC GGTGAACTGA

9241 GCAGCTACCT GATCGACATC ACCAAAGATA TCTTCACCAA AAAAGATGAA GAGGGTAACT

9301 ACCTGGTTGA TGTGATTCTG GATGAAGCAG CAAACAAAGG TACGGGCAAA TGGACCAGCC

9361 AGAGCGCGCT GGATCTCGGC GAACCGCTGT CGCTGATTAC CGAGTCTGTG TTTGCACGTT

9421 ATATCTCTTC TCTGAAAGAG CAGCGTGTTG CCGCATCTAA AGTTCTCTCT GGCCCGCAAG

9481 CGCAGCCAGC TGGCGACAAT GCTGAGTTCA TCGAAAAAGT TCGCCGTGCG CTGTATCTGG

9541 GCAAAATCGT TTCTTACGCT CAGGGCTTCT CTCAGCTACG CGCTGCGTCT GAAGAGTACA

9601 ACTGGGATCT GAACTACGGT GAAATCGCGA AGATTTTCCG TGCTGGCTGC ATCATCCGTG

9661 CGCAGTTCCT GCAGAAAATC ACCGATGCTT ATGCCGAAAA TCCGCAGATC GCTAACCTGT

9721 TGCTGGCTCC TTACTTCAAG CAAATTGCCG ATGACTACCA GCAGGCGCTG CGCGATGTCG

9781 TCGCTTACGC AGTACAGAAC GGTATCCCGG TGCCGACCTT CGCCGCTGCG GTTGCCTATT

9841 ACGACAGCTA CCGCGCCGCT GTTCTGCCTG CGAACCTGAT CCAGGCACAG CGTGACTATT

9901 TCGGTGCGCA TACTTATAAG CGCATTGATA AAGAAGGTGT GTTCCATACC GAATGGCTGG

9961 ATTAATCTGA TTTAAATCAA TTAATCAAAG CAAGGCCCGG AGAAACCCTC CGGGCTTTTT

10021 TATTATACAA AGCGGCAGGT TAGGGCCTTT TTTTATAATT TATAGTTAAA AACGCGATAT

10081 AATACAGCGC CGCACAGCAG GATCGCTGCC TTGACAGTTC ATCTACATCA GCGTTAAAAA

10141 TCCCGCAGTA GATGAAGCTG TGGTGGTGGA TTAATGACCA CTCTAAATGT TTAACCGGAA

10201 GAAGTCAGAG CTAATGAAAA TAACAATTTC AGGAACAGGT TATGTTGGTC TTTCAAATGG

10261 TATTCTGATT GCGCAAAACC ACGAAGTGGT TGCACTGGAT ATCGTTCAGG CCAAAGTGGA

10321 CATGCTTAAC AAGAGGCAGT CACCGCTTGT TGATAAGGAG ATTGAAGAGT ATCTGGCGAC

10381 TAAAGATCTC AATTTCCGCG CTACGACAGA TAAGTATGAC GCGTATAAAA ATGCCGATTA

10441 CGTTATTATT GCCACACCTA CCGATTATGA TCCGAAAACA AATTATTTTA ATACCTCAAG

10501 CGTGGAAGCG GTCATTCGTG ATGTGACAGA AATTAATCCC AACGCGGTAA TGATTATAAA

10561 ATCAACTATC CCTGTTGGTT TTACAGAGTC CATTAAAGAA CGTTTTGGTA TTGAAAATGT

10621 GATCTTTTCG CCTGAGTTTT TGCGTGAAGG TAAAGCACTT TATGATAACT TACACCCATC

10681 ACGCATTGTG ATTGGCGAGC AGTCTGAACG CGCTAAACGT TTTGCTGCGT TATTACAGGA

10741 AGGCGCCATT AAGCAAGACA TACCAACATT GTTTACTGAC TCAACCGAGG CTGAGGCGAT

10801 TAAACTTTTT GCGAACACTT ATCTGGCGAT GCGTGTAGCG TATTTCAATG AACTTGATAG

10861 TTATGCTGAA AGCCTGGGAC TTAATTCACG CCAGATTATT GAGGGCGTAT GCCTTGACCC

10921 GCGTATCGGT AATCACTACA ACAACCCGTC ATTCGGTTAT GGTGGTTATT GTCTGCCGAA

10981 AGATACTAAG CAGTTACTGG CAAATTACCA GTCTGTGCCG AATAACCTGA TCTCGGCAAT

11041 TGTTGACGCC AACCGCACGC GCAAAGATTT TATTGCCGAT TCTATCCTTG CACGTAAACC

11101 GAAAGTTGTT GGCGTCTATC GTTTGATTAT GAAGAATGGT TCAGACAATT TTCGTGCTTC

11161 CTCGATTCAG GGTATTATGA AGCGAATCAA GGCGAAAGGT GTGCCTGTAA TCGTTTATGA

11221 GCCAGCTATG AAAGAGGACG ATTTTTTCCG GTCGCGCGTG GTACGTGATC TGGATGCGTT

11281 CAAACAAGAA GCTGATGTTA TTATTTCTAA CCGTATGTCT GCCGATCTGG CTGATGTAGC

11341 AGATAAAGTT TATACGCGCG ACTTGTTTGG CAATGATTAA TTATTTTGTT TCATTCTAAG

11401 AAAAGGCCCT AATAAATTAG GGCCTTTTCT TATGGTTTTG TAAAATCAAA CTTTATAGAA

11461 GTTACGATAC CATTCTACAA AGTTCTTTAC CCCTTCTTTA ACTGACGTTT CAGGTTTGAA

11521 TCCTATTACG TCATACAGTG CTTTTGTATC AGCACTGGTT TCCAGTACAT CACCGGGTTG

11581 GAGAGGCATC ATATTTTTGT TGGCTTCAAT ACCCAGAGCC TCTTCTAACG CATTGATATA

11641 GTCCATCAAC TCCACAGGCG AACTATTACC AATGTTATAG ACACGATATG GTGCTGAACT

11701 TGTTGCAGGC GAGCCTGTTT CTACAGCCCA CTGTGGGTTT TTTTCTGGAA TAACATCCTG

11761 TAAGCGAATA ATAGCTTCGG CAATATCATC AATGTAAGTA AAGTCACGCT TCATTTTGCC

11821 GAAGTTGTAA ACATCAATGC TTTTACCTTC CAGCATGGCT TTAGTGAATT TAAATAATGC

11881 CATATCCGGA CGTCCCCATG GACCATAAAC CGTAAAGAAA CGCAGCCCTG TGGTCGGTAA

11941 GCCATACAAA TGAGAATATG TATGGGCCAT GAGTTCATTC GCTTTTTTAG TTGCTGCATA

12001 AAGCGAAACA GGATGATCTA CAGAGTCATC TGTAGAGAAA GGCATCTTGC GGTTCATGCC

12061 ATAAACAGAA CTGGAGGAAG CGTAAAGTAG ATGCTGAACA TTATTATGGC GACATCCTTC

12121 TAGTATGTTC AGGAATCCAA TCAGGTTTGC ATCTGCATAT GCATTGGGAT TTTCAAGAGA

12181 GTAACGTACA CCGGCTTGCG CAGCGAGGTT TATTACGCGT TCGAACCGCT CGTCTGCAAA

12241 CAGTGCCGCC ATTTTCTCAC GATCGGCCAG GTCAATTTTA TAAAAACTGA AGTTGTCGTG

12301 CTTGAGTAAA TCAAGTCGTG CTTGTTTGAG GTTGACATCG TAATAATCAT TTAAGTTGTC

12361 AATGCCTACA ACCTGATGAC CAGCTGCAAG AAGCCGTTTA CTTAGATAGA AACCGATAAA

12421 GCCAGCAGCT CCCGTAACCA GAAATTTCAT TTATAATCCT CGCTCAGGCT AGAATATAGC

12481 CAATCTTCAT CTGGCATAAC TGAAAGTTAA ATTATACCGT TAGACAAGAA AAAAAGATAA

12541 TCGGTATCAG TTCTAAACTT GGCTGTTTTT TCTGGTAACG TGCTCATTTT ACAATCAAAG

12601 CTGTTCTAAG CTGACTATAC AAGCCGACGT CATTATCTCC AACCGTATGG CAGAAGAGCT

12661 TAAGGATGTG GCAGACAAAG TCTACACCCG CGATCTCTTT GGCAGTGACT AACATCCTGT

12721 TATCATGGCG ATTTTCGCCC TGATTCTCTT ATGTTCCCTT TGTAATAATT CATTATTTTT

12781 ATCATTTATC CTATAGCATT CATGGCGATT ATCGCTAAAC TATGGCGGCG CGCCACGTGG

12841 GATCCCCGGG TACCGAGCTC GAATTCGCCC TATAGTGAGT CGTATTACAA TTCACTGGCC

12901 GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA TCGCCTTGCA

12961 GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA TCGCCCTTCC

13021 CAACAGTTGC GCAGCTGAAT GGCGAATGGC GCCTGATGCG GTATTTTCTC CTTACGCATC

13081 TGTGCGGTAT TTCACACCGC ATATGGTGCA CTCTCAGTAC AATCTGCTCT GATGCCGCAT

13141 AGTTAAGCCA GCCCCGACAC CCGCCAACAC CCGCTGACGC GAACCCCTTG CGGCCGCATC

13201 GAATATAACT TCGTATAATG TATGCTATAC GAAGTTATTA GCGATGAGCT CGGACTTCCA

13261 TTGTTCATTC CACGGACAAA AACAGAGAAA GGAAACGACA GAGGCCAAAA AGCTCGCTTT

13321 CAGCACCTGT CGTTTCCTTT CTTTTCAGAG GGTATTTTAA ATAAAAACAT TAAGTTATGA

13381 CGAAGAAGAA CGGAAACGCC TTAAACCGGA AAATTTTCAT AAATAGCGAA AACCCGCGAG

13441 GTCGCCGCCC CGTAACCTGT CGGATCACCG GAAAGGACCC GTAAAGTGAT AATGATTATC

13501 ATCTACATAT CACAACGTGC GTGGAGGCCA TCAAACCACG TCAAATAATC AATTATGACG

13561 CAGGTATCGT ATTAATTGAT CTGCATCAAC TTAACGTAAA AACAACTTCA GACAATACAA

13621 ATCAGCGACA CTGAATACGG GGCAACCTCA TGTCCGAGCT CGCGAGCTCG TCGACAGCGA

13681 CACACTTGCA TCGGATGCAG CCCGGTTAAC GTGCCGGCAC GGCCTGGGTA ACCAGGTATT

13741 TTGTCCACAT AACCGTGCGC AAAATGTTGT GGATAAGCAG GACACAGCAG CAATCCACAG

13801 CAGGCATACA ACCGCACACC GAGGTTACTC CGTTCTACAG GTTACGACGA CATGTCAATA

13861 CTTGCCCTTG ACAGGCATTG ATGGAATCGT AGTCTCACGC TGATAGTCTG ATCGACAATA

13921 CAAGTGGGAC CGTGGTCCCA GACCGATAAT CAGACCGACA ACACGAGTGG GATCGTGGTC

13981 CCAGACTAAT AATCAGACCG ACGATACGAG TGGGACCGTG GTCCCAGACT AATAATCAGA

14041 CCGACGATAC GAGTGGGACC GTGGTTCCAG ACTAATAATC AGACCGACGA TACGAGTGGG

14101 ACCGTGGTCC CAGACTAATA ATCAGACCGA CGATACGAGT GGGACCATGG TCCCAGACTA

14161 ATAATCAGAC CGACGATACG AGTGGGACCG TGGTCCCAGT CTGATTATCA GACCGACGAT

14221 ACGAGTGGGA CCGTGGTCCC AGACTAATAA TCAGACCGAC GATACGAGTG GGACCGTGGT

14281 CCCAGACTAA TAATCAGACC GACGATACGA GTGGGACCGT GGTCCCAGTC TGATTATCAG

14341 ACCGACGATA CAAGTGGAAC AGTGGGCCCA GAGAGAATAT TCAGGCCAGT TATGCTTTCT

14401 GGCCTGTAAC AAAGGACATT AAGTAAAGAC AGATAAACGT AGACTAAAAC GTGGTCGCAT

14461 CAGGGTGCTG GCTTTTCAAG TTCCTTAAGA ATGGCCTCAA TTTTCTCTAT ACACTCAGTT

14521 GGAACACGAG ACCTGTCCAG GTTAAGCACC ATTTTATCGC CCTTATACAA TACTGTCGCT

14581 CCAGGAGCAA ACTGATGTCG TGAGCTTAAA CTAGTTCTTG ATGCAGATGA CGTTTTAAGC

14641 ACAGAAGTTA AAAGAGTGAT AACTTCTTCA GCTTCAAATA TCACCCCAGC TTTTTTCTGC

14701 TCATGAAGGT TAGATGCCTG CTGCTTAAGT AATTCCTCTT TATCTGTAAA GGCTTTTTGA

14761 AGTGCATCAC CTGACCGGGC AGATAGTTCA CCGGGGTGAG AAAAAAGAGC AACAACTGAT

14821 TTAGGCAATT TGGCGGTGTT GATACAGCGG GTAATAATCT TACGTGAAAT ATTTTCCGCA

14881 TCAGCCAGCG CAGAAATATT TCCAGCAAAT TCATTCTGCA ATCGGCTTGC ATAACGCTGA

14941 CCACGTTCAT AAGCACTTGT TGGGCGATAA TCGTTACCCA ATCTGGATAA TGCAGCCATC

15001 TGCTCATCAT CCAGCTCGCC AACCAGAACA CGATAATCAC TTTCGGTAAG TGCAGCAGCT

15061 TTACGACGGC GACTCCCATC GGCAATTTCT ATGACACCAG ATACTCTTCG ACCGAACGCC

15121 GGTGTCTGTT GACCAGTCAG TAGAAAAGAA GGGATGAGAT CATCCAGTGC GTCCTCAGTA

15181 AGCAGCTCCT GGTCACGTTC ATTACCTGAC CATACCCGAG AGGTCTTCTC AACACTATCA

15241 CCCCGGAGCA CTTCAAGAGT AAACTTCACA TCCCGACCAC ATACAGGCAA AGTAATGGCA

15301 TTACCGCGAG CCATTACTCC TACGCGCGCA ATTAACGAAT CCACCATCGG GGCAGCTGGT

15361 GTCGATAACG AAGTATCTTC AACCGGTTGA GTATTGAGCG TATGTTTTGG AATAACAGGC

15421 GCACGCTTCA TTATCTAATC TCCCAGCGTG GTTTAATCAG ACGATCGAAA ATTTCATTGC

15481 AGACAGGTTC CCAAATAGAA AGAGCATTTC TCCAGGCACC AGTTGAAGAG CGTTGATCAA

15541 TGGCCTGTTC AAAAACAGTT CTCATCCGGA TCTGACCTTT ACCAACTTCA TCCGTTTCAC

15601 GTACAACATT TTTTAGAACC ATGCTTCCCC AGGCATCCCG AATTTGCTCC TCCATCCACG

15661 GGGACTGAGA GCCATTACTA TTGCTGTATT TGGTAAGCAA AATACGTACA TCAGGCTCGA

15721 ACCCTTTAAG ATCAACGTTC TTGAGCAGAT CACGAAGCAT ATCGAAAAAC TGCAGTGCGG

15781 AGGTGTAGTC AAACAACTCA GCAGGCGTGG GAACAATCAG CACATCAGCA GCACATACGA

15841 CATTAATCGT GCCGATACCC AGGTTAGGCG CGCTGTCAAT AACTATGACA TCATAGTCAT

15901 GAGCAACAGT TTCAATGGCC AGTCGGAGCA TCAGGTGTGG ATCGGTGGGC AGTTTACCTT

15961 CATCAAATTT GCCCATTAAC TCAGTTTCAA TACGGTGCAG AGCCAGACAG GAAGGAATAA

16021 TGTCAAGCCC CGGCCAGCAA GTGGGCTTTA TTGCATAAGT GACATCGTCC TTTTCCCCAA

16081 GATAGAAAGG CAGGAGAGTG TCTTCTGCAT GAATATGAAG ATCTGGTACC CATCCGTGAT

16141 ACATTGAGGC TGTTCCCTGG GGGTCGTTAC CTTCCACGAG CAAAACACGT AGCCCCTTCA

16201 GAGCCAGATC CTGAGCAAGA TGAACAGAAA CTGAGGTTTT GTAAACGCCA CCTTTATGGG

16261 CAGCAACCCC GATCACCGGT GGAAATACGT CTTCAGCACG TCGCAATCGC GTACCAAACA

16321 CATCACGCAT ATGATTAATT TGTTCAATTG TATAACCAAC ACGTTGCTCA ACCCGTCCTC

16381 GAATTTCCAT ATCCGGGTGC GGTAGTCGCC CTGCTTTCTC GGCATCTCTG ATAGCCTGAG

16441 AAGAAACCCC AACTAAATCC GCTGCTTCAC CTATTCTCCA GCGCCGGGTT ATTTTCCTCG

16501 CTTCCGGGCT GTCATCATTA AACTGTGCAA TGGCGATAGC CTTCGTCATT TCATGACCAG

16561 CGTTTATGCA CTGGTTAAGT GTTTCCATGA GTTTCATTCT GAACATCCTT TAATCATTGC

16621 TTTGCGTTTT TTTATTAAAT CTTGCAATTT ACTGCAAAGC AACAACAAAA TCGCAAAGTC

16681 ATCAAAAAAC CGCAAAGTTG TTTAAAATAA GAGCAACACT ACAAAAGGAG ATAAGAAGAG

16741 CACATACCTC AGTCACTTAT TATCACTAGC GCTCGCCGCA GCCGTGTAAC CGAGCATAGC

16801 GAGCGAACTG GCGAGGAAGC AAAGAAGAAC TGTTCTGTCA GATAGCTCTT ACGCTCAGCG

16861 CAAGAAGAAA TATCCACCGT GGGAAAAACT CCAGGTAGAG GTACACACGC GGATAGCCAA

16921 TTCAGAGTAA TAAACTGTGA TAATCAACCC TCATCAATGA TGACGAACTA ACCCCCGATA

16981 TCAGGTCACA TGACGAAGGG AAAGAGAAGG AAATCAACTG TGACAAACTG CCCTCAAATT

17041 TGGCTTCCTT AAAAATTACA GTTCAAAAAG TATGAGAAAA TCCATGCAGG CTGAAGGAAA

17101 CAGCAAAACT GTGACAAATT ACCCTCAGTA GGTCAGAACA AATGTGACGA ACCACCCTCA

17161 AATCTGTGAC AGATAACCCT CAGACTATCC TGTCGTCATG GAAGTGATAT CGCGGAAGGA

17221 AAATACGATA TGAGTCGTCT GGCGGCCTTT CTTTTTCTCA ATGTATGAGA GGCGCATTGG

17281 AGTTCTGCTG TTGATCTCAT TAACACAGAC CTGCAGGAAG CGGCGGCGGA AGTCAGGCAT

17341 ACGCTGGTAA CTTTGAGGCA GCTGGTAACG CTCTATGATC CAGTCGATTT TCAGAGAGAC

17401 GATGCCTGAG CCATCCGGCT TACGATACTG ACACAGGGAT TCGTATAAAC GCATGGCATA

17461 CGGATTGGTG ATTTCTTTTG TTTCACTAAG CCGAAACTGC GTAAACCGGT TCTGTAACCC

17521 GATAAAGAAG GGAATGAGAT ATGGGTTGAT ATGTACACTG TAAAGCCCTC TGGATGGACT

17581 GTGCGCACGT TTGATAAACC AAGGAAAAGA TTCATAGCCT TTTTCATCGC CGGCATCCTC

17641 TTCAGGGCGA TAAAAAACCA CTTCCTTCCC CGCGAAACTC TTCAATGCCT GCCGTATATC

17701 CTTACTGGCT TCCGCAGAGG TCAATCCGAA TATTTCAGCA TATTTAGCAA CATGGATCTC

17761 GCAGATACCG TCATGTTCCT GTAGGGTGCC ATCAGATTTT CTGATCTGGT CAACGAACAG

17821 ATACAGCATA CGTTTTTGAT CCCGGGAGAG ACTATATGCC GCCTCAGTGA GGTCGTTTGA

17881 CTGGACGATT CGCGGGCTAT TTTTACGTTT CTTGTGATTG ATAACCGCTG TTTCCGCCAT

17941 GACAGATCCA TGTGAAGTGT GACAAGTTTT TAGATTGTCA CACTAAATAA AAAAGAGTCA

18001 ATAAGCAGGG ATAACTTTGT GAAAAAACAG CTTCTTCTGA GGGCAATTTG TCACAGGGTT

18061 AAGGGCAATT TGTCACAGAC AGGACTGTCA TTTGAGGGTG ATTTGTCACA CTGAAAGGGC

18121 AATTTGTCAC AACACCTTCT CTAGAACCAG CATGGATAAA GGCCTACAAG GCGCTCTAAA

18181 AAAGAAGATC TAAAAACTAT AAAAAAAATA ATTATAAAAA TATCCCCGTG GATAAGTGGA

18241 TAACCCCAAG GGAAGTTTTT TCAGGCATCG TGTGTAAGCA GAATATATAA GTGCTGTTCC

18301 CTGGTGCTTC CTCGCTCACT CGACCGGGAG GGTTCGAGAA GGGGGGGCAC CCCCCTTCGG

18361 CGTGCGCGGT CACGCGCACA GGGCGCAGCC CTGGTTAAAA ACAAGGTTTA TAAATATTGG

18421 TTTAAAAGCA GGTTAAAAGA CAGGTTAGCG GTGGCCGAAA AACGGGCGGA AACCCTTGCA

18481 AATGCTGGAT TTTCTGCCTG TGGACAGCCC CTCAAATGTC AATAGGTGCG CCCCTCATCT

18541 GTCAGCACTC TGCCCCTCAA GTGTCAAGGA TCGCGCCCCT CATCTGTCAG TAGTCGCGCC

18601 CCTCAAGTGT CAATACCGCA GGGCACTTAT CCCCAGGCTT GTCCACATCA TCTGTGGGAA

18661 ACTCGCGTAA AATCAGGCGT TTTCGCCGAT TTGCGAGGCT GGCCAGCTCC ACGTCGCCGG

18721 CCGAAATCGA GCCTGCCCCT CATCTGTCAA CGCCGCGCCG GGTGAGTCGG CCCCTCAAGT

18781 GTCAACGTCC GCCCCTCATC TGTCAGTGAG GGCCAAGTTT TCCGCGAGGT ATCCACAACG

18841 CCGGCGGCCG GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG

18901 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCGAGGGCTT CGCCCTGTCG

18961 CTCGACTGCG GCGAGCACTA CTGGCTGTAA AAGGACAGAC CACATCATGG TTCTGTGTTC

19021 ATTAGGTTGT TCTGTCCATT GCTGACATAA TCCGCTCCAC TTCAACGTAA CACCGCACGA

19081 AGATTTCTAT TGTTCCTGAA GGCATATTCA AATCGTTTTC GTTACCGCTT GCAGGCATCA

19141 TGACAGAACA CTACTTCCTA TAAACGCTAC ACAGGCTCCT GAGATTAATA ATGCGGATCT

19201 CTACGATAAT GGGAGATTTT CCCGACTGTT TCGTTCGCTT CTCAGTGGAT AACAGCCAGC

19261 TTCTCTGTTT AACAGACAAA AACAGCATAT CCACTCAGTT CCACATTTCC ATATAAAGGC

19321 CAAGGCATTT ATTCTCAGGA TAATTGTTTC AGCATCGCAA CCGCATCAGA CTCCGGCATC

19381 GCAAACTGCA CCCGGTGCCG GGCAGCCACA TCCAGCGCAA AAACCTTCGT GTAGACTTCC

19441 GTTGAACTGA TGGACTTATG TCCCATCAGG CTTTGCAGAA CTTTCAGCGG TATACCGGCA

19501 TACAGCATGT GCATCGCATA GGAATGGCGG AACGTATGTG GTGTGACCGG AACAGAGAAC

19561 GTCACACCGT CAGCAGCAGC GGCGGCAACC GCCTCCCCAA TCCAGGTCCT GACCGTTCTG

19621 TCCGTCACTT CCCAGATCCG CGCTTTCTCT GTCCTTCCTG TGCGACGGTT ACGCCGCTCC

19681 ATGAGCTTAT CGCGAATAAA TACCTGTGAC GGAAGATCAC TTCGCAGAAT AAATAAATCC

19741 TGGTGTCCCT GTTGATACCG GGAAGCCCTG GGCCAACTTT TGGCGAAAAT GAGACGTTGA

19801 TCGGCACGTA AGAGGTTCCA ACTTTCACCA TAATGAAATA AGATCACTAC CGGGCGTATT

19861 TTTTGAGTTA TCGAGATTTT CAGGAGCTAA GGAAGCTAAA ATGGAGAAAA AAATCACTGG

19921 ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA CATTTTGAGG CATTTCAGTC

19981 AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT ATTACGGCCT TTTTAAAGAC

20041 CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCCTTTATT CACATTCTTG CCCGCCTGAT

20101 GAATGCTCAT CCGGAATTTC GTATGGCAAT GAAAGACGGT GAGCTGGTGA TATGGGATAG

20161 TGTTCACCCT TGTTACACCG TTTTCCATGA GCAAACTGAA ACGTTTTCAT CGCTCTGGAG

20221 TGAATACCAC GACGATTTCC GGCAGTTTCT ACACATATAT TCGCAAGATG TGGCGTGTTA

20281 CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG AATATGTTTT TCGTCTCAGC

20341 CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTTAAACGTG GCCAATATGG ACAACTTCTT

20401 CGCCCCCGTT TTCACCATGG GCAAATATTA TACGCAAGGC GACAAGGTGC TGATGCCGCT

20461 GGCGATTCAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT GTCGGCAGAA TGCTTAATGA

20521 ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA TTTTTTTAAG GCAGTTATTG

20581 GTGCCCTTAA ACGCCTGGTT GCTACGCCTG AATAAGTGAT AATAAGCGGA TGAATGGCAG

20641 AAATTCGATG ATAAGCTGTC AAACATGAGA ATTGGTCGAC GGCCCGGGCG GCCGCAAGGG

20701 GTTCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTTCCCGAC TGGAAAGCGG

20761 GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC

20821 ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG

20881 GAAACAGCTA TGACCATGAT TACGCCAAGC TATTTAGGTG AGACTATAGA ATACTCAAGC

20941 TTGCATGCCT GCAGGTCGAC TCTAGAGGAT CCCACGACGT CG

[0414] **SEQ ID NO: 29 - Nucleotide Sequence for pCC1FOS cut (pFOS) and *S. flexneri* 6 O-antigen with Z3206**

[0415] Locus pFOS cut and O-antigen cut (Z3206+)

[0416] Definition Ligation of inverted S.flexneri 6 O antigen cluster amplified with Z3206Nhe and wzzAscI cut with NheI and AscI into pCC1FOS with MCS cassette cut with NheI and AscI

[0417] Features Location/Qualifiers

[0418] CDS complement(370..396)

[0419] /label=wzz'

[0420] CDS 748..1752

[0421] /label=uge

[0422] CDS complement(1818..3011)

[0423] /label=ugd

[0424] CDS complement(3233..4639)

[0425] /label=gnd

[0426] CDS complement(4744..5577)

[0427] /label=wfbZ

[0428] CDS complement(5574..6443)

[0429] /label=wfbY

[0430] CDS complement(6460..7647)

[0431] /label=wzy

[0432] CDS complement(7703..8935)

[0433] /label=wzx

[0434] CDS complement(8932..9489)

[0435] /label=rmlC

[0436] CDS complement(9494..10372)

[0437] /label=rmlA

[0438] CDS complement(10430..11329)

[0439] /label=rmlD

[0440] CDS complement(11329..12414)

[0441] /label=rmlB

[0442] CDS complement(12787..13680)

[0443] /label=galF

[0444] CDS complement(13912..14907)

[0445] /label=Z3206

[0446] CDS complement(15065..15097)

[0447] /label='wcaM

[0448] CDS complement(15525..16184)

[0449] /label=cat

[0450] CDS 16403..16750

[0451] /label=redF

[0452] CDS 18145..18900

[0453] /label=repE

[0454] CDS 19479..20654

[0455] /label=parA

[0456] CDS 20654..21625

[0457] /label=parB

[0458] Length: 22887 bp

[0459] Type: DNA circular UNA

[0460] Sequence:

1 GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG

61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT

121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA ACTGTTGGGA

181 AGGGCGATCG GTGCGGGCCT CTTCGCTATT ACGCCAGCTG GCGAAAGGGG GATGTGCTGC

241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA AAACGACGGC

301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC GGGGATCCCA

361 CGTGGCGCGC CGCCATAGTT TAGCGATAAT CGCCATGAAT GCTATAGGAT AAATGATAAA

421 AATAATGAAT TATTACAAAG GGAACATAAG AGAATCAGGG CGAAAATCGC CATGATAACA

481 GGATGTTAGT CACTGCCAAA GAGATCGCGG GTGTAGACTT TGTCTGCCAC ATCCTTAAGC

541 TCTTCTGCCA TACGGTTGGA GATAATGACG TCGGCTTGTA TAGTCAGCTT AGAACAGCTT

601 TGATTGTAAA ATGAGCACGT TACCAGAAAA AACAGCCAAG TTTAGAACTG ATACCGATTA

661 TCTTTTTTTC TTGTCTAACG GTATAATTTA ACTTTCAGTT ATGCCAGATG AAGATTGGCT

721 ATATTCTAGC CTGAGCGAGG ATTATAAATG AAATTTCTGG TTACGGGAGC TGCTGGCTTT

781 ATCGGTTTCT ATCTAAGTAA ACGGCTTCTT GCAGCTGGTC ATCAGGTTGT AGGCATTGAC

841 AACTTAAATG ATTATTACGA TGTCAACCTC AAACAAGCAC GACTTGATTT ACTCAAGCAC

901 GACAACTTCA GTTTTTATAA AATTGACCTG GCCGATCGTG AGAAAATGGC GGCACTGTTT

961 GCAGACGAGC GGTTCGAACG CGTAATAAAC CTCGCTGCGC AAGCCGGTGT ACGTTACTCT

1021 CTTGAAAATC CCAATGCATA TGCAGATGCA AACCTGATTG GATTCCTGAA CATACTAGAA

1081 GGATGTCGCC ATAATAATGT TCAGCATCTA CTTTACGCTT CCTCCAGTTC TGTTTATGGC

1141 ATGAACCGCA AGATGCCTTT CTCTACAGAT GACTCTGTAG ATCATCCTGT TTCGCTTTAT

1201 GCAGCAACTA AAAAAGCGAA TGAACTCATG GCCCATACAT ATTCTCATTT GTATGGCTTA

1261 CCGACCACAG GGCTGCGTTT CTTTACGGTT TATGGTCCAT GGGGACGTCC GGATATGGCA

1321 TTATTTAAAT TCACTAAAGC CATGCTGGAA GGTAAAAGCA TTGATGTTTA CAACTTCGGC

1381 AAAATGAAGC GTGACTTTAC TTACATTGAT GATATTGCCG AAGCTATTAT TCGCTTACAG

1441 GATGTTATTC CAGAAAAAAA CCCACAGTGG GCTGTAGAAA CAGGCTCGCC TGCAACAAGT

1501 TCAGCACCAT ATCGTGTCTA TAACATTGGT AATAGTTCGC CTGTGGAGTT GATGGACTAT

1561 ATCAATGCGT TAGAAGAGGC TCTGGGTATT GAAGCCAACA AAAATATGAT GCCTCTCCAA

1621 CCCGGTGATG TACTGGAAAC CAGTGCTGAT ACAAAAGCAC TGTATGACGT AATAGGATTC

1681 AAACCTGAAA CGTCAGTTAA AGAAGGGGTA AAGAACTTTG TAGAATGGTA TCGTAACTTC

1741 TATAAAGTTT GATTTTACAA AACCATAAGA AAAGGCCCTA ATTTATTAGG GCCTTTTCTT

1801 AGAATGAAAC AAAATAATTA ATCATTGCCA AACAAGTCGC GCGTATAAAC TTTATCTGCT

1861 ACATCAGCCA GATCGGCAGA CATACGGTTA GAAATAATAA CATCAGCTTC TTGTTTGAAC

1921 GCATCCAGAT CACGTACCAC GCGCGACCGG AAAAAATCGT CCTCTTTCAT AGCTGGCTCA

1981 TAAACGATTA CAGGCACACC TTTCGCCTTG ATTCGCTTCA TAATACCCTG AATCGAGGAA

2041 GCACGAAAAT TGTCTGAACC ATTCTTCATA ATCAAACGAT AGACGCCAAC AACTTTCGGT

2101 TTACGTGCAA GGATAGAATC GGCAATAAAA TCTTTGCGCG TGCGGTTGGC GTCAACAATT

2161 GCCGAGATCA GGTTATTCGG CACAGACTGG TAATTTGCCA GTAACTGCTT AGTATCTTTC

2221 GGCAGACAAT AACCACCATA ACCGAATGAC GGGTTGTTGT AGTGATTACC GATACGCGGG

2281 TCAAGGCATA CGCCCTCAAT AATCTGGCGT GAATTAAGTC CCAGGCTTTC AGCATAACTA

2341 TCAAGTTCAT TGAAATACGC TACACGCATC GCCAGATAAG TGTTCGCAAA AAGTTTAATC

2401 GCCTCAGCCT CGGTTGAGTC AGTAAACAAT GTTGGTATGT CTTGCTTAAT GGCGCCTTCC

2461 TGTAATAACG CAGCAAAACG TTTAGCGCGT TCAGACTGCT CGCCAATCAC AATGCGTGAT

2521 GGGTGTAAGT TATCATAAAG TGCTTTACCT TCACGCAAAA ACTCAGGCGA AAAGATCACA

2581 TTTTCAATAC CAAAACGTTC TTTAATGGAC TCTGTAAAAC CAACAGGGAT AGTTGATTTT

2641 ATAATCATTA CCGCGTTGGG ATTAATTTCT GTCACATCAC GAATGACCGC TTCCACGCTT

2701 GAGGTATTAA AATAATTTGT TTTCGGATCA TAATCGGTAG GTGTGGCAAT AATAACGTAA

2761 TCGGCATTTT TATACGCGTC ATACTTATCT GTCGTAGCGC GGAAATTGAG ATCTTTAGTC

2821 GCCAGATACT CTTCAATCTC CTTATCAACA AGCGGTGACT GCCTCTTGTT AAGCATGTCC

2881 ACTTTGGCCT GAACGATATC CAGTGCAACC ACTTCGTGGT TTTGCGCAAT CAGAATACCA

2941 TTTGAAAGAC CAACATAACC TGTTCCTGAA ATTGTTATTT TCATTAGCTC TGACTTCTTC

3001 CGGTTAAACA TTTAGAGTGG TCATTAATCC ACCACCACAG CTTCATCTAC TGCGGGATTT

3061 TTAACGCTGA TGTAGATGAA CTGTCAAGGC AGCGATCCTG CTGTGCGGCG CTGTATTATA

3121 TCGCGTTTTT AACTATAAAT TATAAAAAAA GGCCCTAACC TGCCGCTTTG TATAATAAAA

3181 AAGCCCGGAG GGTTTCTCCG GGCCTTGCTT TGATTAATTG ATTTAAATCA GATTAATCCA

3241 GCCATTCGGT ATGGAACACA CCTTCTTTAT CAATGCGCTT ATAAGTATGC GCACCGAAAT

3301 AGTCACGCTG TGCCTGGATC AGGTTCGCAG GCAGAACAGC GGCGCGGTAG CTGTCGTAAT

3361 AGGCAACCGC AGCGGCGAAG GTCGGCACCG GGATACCGTT CTGTACTGCG TAAGCGACGA

3421 CATCGCGCAG CGCCTGCTGG TAGTCATCGG CAATTTGCTT GAAGTAAGGA GCCAGCAACA

3481 GGTTAGCGAT CTGCGGATTT TCGGCATAAG CATCGGTGAT TTTCTGCAGG AACTGCGCAC

3541 GGATGATGCA GCCAGCACGG AAAATCTTCG CGATTTCACC GTAGTTCAGA TCCCAGTTGT

3601 ACTCTTCAGA CGCAGCGCGT AGCTGAGAGA AGCCCTGAGC GTAAGAAACG ATTTTGCCCA

3661 GATACAGCGC ACGGCGAACT TTTTCGATGA ACTCAGCATT GTCGCCAGCT GGCTGCGCTT

3721 GCGGGCCAGA GAGAACTTTA GATGCGGCAA CACGCTGCTC TTTCAGAGAA GAGATATAAC

3781 GTGCAAACAC AGACTCGGTA ATCAGCGACA GCGGTTCGCC GAGATCCAGC GCGCTCTGGC

3841 TGGTCCATTT GCCCGTACCT TTGTTTGCTG CTTCATCCAG AATCACATCA ACCAGGTAGT

3901 TACCCTCTTC ATCTTTTTTG GTGAAGATAT CTTTGGTGAT GTCGATCAGG TAGCTGCTCA

3961 GTTCACCGTT ATTCCACTCG GTAAAGGTCT GCGCCAGTTC TTCGTTGGTG AGGTTCAAGC

4021 CACCTTTAAG CAGAGAATAG GCTTCAGCAA TCAGCTGCAT ATCACCGTAT TCAATACCGT

4081 TGTGAACCAT CTTCACATAA TGACCTGCAC CATCGGCACC AATATAGGTA ACGCACGGTT

4141 CGCCGTCTTC AGCCACAGCG GCGATTTTGG TCAGGATCGG CGCAATCAGT TCATAAGCTT

4201 CTTTCTGCCC ACCAGGCATA ATGGAAGGAC CTTTCAGCGC ACCTTCTTCA CCACCGGAAA

4261 CACCGGTACC GATAAAGTTA AAGCCTTCTG CAGAAAGCTC ACGGTTACGA CGAATGGTGT

4321 CATGGAAGAA GGTGTTACCA CCATCAATGA TGATGTCACC TTTATCGAGG TATGGCTTGA

4381 GGGAATCAAT AGCAGCATCC GTGCCAGCAC CTGCTTTCAC CATTAACAGG ATGCGACGAG

4441 GCGTTTCCAG AGATTCAACA AATTCTTTCA CCGTATAGTA AGGAACCAGT TTCTTGCCTG

4501 GATTTTCGGT AATCACTTCT TCGGTCTTTT CACGGGAACG GTTGAAAATA GAGACGGTAT

4561 AACCACGGCT TTCGATATTG AGCGCAAGGT TGCGCCCCAT CACTGCCATA CCGACGACGC

4621 CGATCTGTTG CTTTGACATT GTTTACTCCT GTCAGGATAC CGCTGGGTGG TATGCGGGTT

4681 ATGCTTAATT ATAGAATATG CCTAATAAAA ATAAATCCAT AACACTTAAT CAGAAAATTA

4741 TTATTATCGA TTCCTAACGA TTGAATACAT CAGCTCCTTT AATTTAGATG GCATTATACG

4801 AAAAAATGTT CTCAACATAG CATTACTTAT TAATTCATTT TTTCGAATAA AACCAATTTT

4861 ATATTGATAA TACAATACTT TATACTCGTA CAATAAATAT GACAATCCAC GTCGAGCCAT

4921 AAGATTACGA CCAGTTCGCA TTTTTAATAA AATATCTGGA AGATTTGCAA ATCTTGCATT

4981 ATGTACAATT AATAGGCTCC ACAATGCAAA ATCTTGAGAT TTTCTGAATG GAGGATAACC

5041 ACCAACAGCT AATACTGTAT TCTTTCTAAA AATTACAGAA GGATGGCTAA CTGCGCTTCG

5101 TTTCCTCGCG AATTTAACTA TTTCTCTATG TTCGAGAGGC ACTTTGCGTG TTGAAATAAA

5161 CTCCTCAGTA ACAGTTTCAA TTTCATCAAT AAAACTGCCA CATACATCTA TTTCTGAATT

5221 ATTAATCATA AAAGAAATTT GTTTCTCAAA CCGATGAGGC AAAGAAATAT CATCAGCATC

5281 CATTCTTGCC ACTAACTCAT TCCTACAAGC CTTTAATCCT TCATTTAAGG CATTAGCCAA

5341 TCCAACATTT CTAGGTAAAG GTACAAATGT TACTATTTTA TTGCCAACAT CATCAATGAA

5401 TGAATTTATA ATATCGATGT GTGTTTGATG GAGTTCTCCA TCTGCAACAA TTACTATTTG

5461 ATCTGGCTTA AGTGTTTGAT CGTGAAAAAT AGAGCGTAGA GCCACCTCAA AAAATTGCGG

5521 TAGATCATTT TTATAAATGC TAATTAAAAC TGAGAATTTT TCTAATCTAT GATTCATTTC

5581 ATTTTACCAC TTCGACCCAT TAAACCGTCA TTAATGCCTT TTAAAAAAAA ATATAACCTT

5641 TTATTACCAT TTGGAAGGAA AATAGGATAT AAAAAAACCT TTCCAATTAA TTTAACCAGA

5701 CTAGAAATTT TCCAGTAGAT GGGTACATAA TTTTTATTTA ATAAAAGAAA GATATTTCGA

5761 GTAGCATAAT AATGACGAAA TGGGCTTGGC AAACCGACAG AAAGAATATT TAAGATCTTA

5821 AATCGCCCAT CTCCAAGTCT ATGTGCAAGT AACGCATTTT TATTCCTAAT TACTTTAAAC

5881 CCAGCAGCTC TTAATCTCCA ACAATATTCA TGGTCTACCG CATCGATAAA AAGCTCATCT

5941 TTCATTCCTC CAACAATCAA CCAACTATTT TTTGGTATTA GACTGCCAGA ACTTAATGTA

6001 CTATCTACCT CATAATAAAC TTCTGTAAGT GGTTTCCCTT TTTTTACCCT TGCTTTATTT

6061 AATTCACCAG TTACTTTATC AAAATCTTGT GAACCAACTA AACCAACATT GACATTTTGT

6121 TTAAGCAATT TTTTGTAACA AGTAAGTAAC TGCTCTACCA TCTTAGGATC AGGAATACTA

6181 TCCTGATCCA TTTGCAATAT AAAATCAGCG CCATTTTCAA AAGCCCATTT CATTCCTATA

6241 CTTTGGGCTT CTGCTATGCC TAAATTATCA TTGAAATTGA ATATTTTTAC ATCGCCTGAA

6301 GAATTTTCAG CATATTTATA ACCATTTGTA GAGTTATTGC AAACGACAAC TTTAGTAACT

6361 TGTCTCAACA ATAATTCAAC CGCATTTTTT AAATCATTAT GTTCTGGGTT GTAAGCAACC

6421 AAAACGGCAT ATACAGTGTC CATCTTCACC TTAAAACCTT CATTTAGCTT TCATCTTTTT

6481 TAGAACATTA CTTAATGTCA CTAATACAAT TATTACAGCA ACATGGTTAG AGTCTAAAAT

6541 ATAAGGATTA GTAATTGCAT AAGAAACATA TAGAAAATAT AGCACACACA ACTCACTGTA

6601 TTTTATGATT TTAATCGTGA GAAGGAGATT AATTAATAAA AACAAAGTAA ATAAAATAAC

6661 GCCAAGTTGA TTTAAAAAAT AAACTGACTG CAATTCATAA TATATATATG CACTATAATC

6721 ACGGATAGGA GTTTGAATTT TGATGACATT ACCCAAACCA GAACCTATAA CAAAATTTGA

6781 TACAGACTCT GTAAGATCAT TAATTAATAC AGTAAACTGA TCCCATCTAA CTCCTAAAGA

6841 AGAATCAGCT CCATTTGATT TCATGATTAT CAACTCAATT GAATATGTAA TAAAAAAAGG

6901 GAGAATCACA GTAAGAAAAA CCCCAAAAAT AATTTTCCTT AATTTAGCGT ATCGTGAGTT

6961 AGATTTAGAA CATAGTATAA TATACATAAA AAACAAGCAT ATCGAAACAA AATATGCAAA

7021 ATTACCAGCC ACTATAGTAC CTATAGCCAG AATAACGGTT ATTGTATTTT TGAATCGATA

7081 ATAGAAATAA TCTTTTATGA CTATATGCAA CATAAAGGCA AATGGAATGA GAGCATTTCC

7141 TTTAATTTGA ACTCTATAGA AACCACTTCC ATATGTATAA ACATCACCAT AATCATTCTC

7201 CAAAAAATAA TGTCTTAGTG CTGAATAATC ACCAATACCA TATGTTTTTG TCATATAAAT

7261 ACTAATGATG GATATAATAA CCGCCTGTAA TACCATTAAA TATAAAAATA TTTTAACAAT

7321 CGAGATGGTT CCATAAGAGC AGAAATAAGC ACATAATATA AATAATATGA TAATATAAAA

7381 CCTAATTATT ATCGCTATAT CGTTACCCTT GATATAGGAA TAAATAAAAT TTATAAAAAG

7441 AGCTAATAGA AATATTAAAA TAACAGGATA GTGATATATT CCGTTTGCAA TTTTCTTTGT

7501 AAATGACATG ATACAAAGAC ATAAAAACCC CTCCATAATC CAACTATATT GAATAAATGG

7561 AAAGCTACGT GTAAGGAAAA ATATAAACCC AAAAAACAAA AGAACACTTA AACTTTTGTC

7621 TTTTGAGTTA TAAAAATCAG AAGTCATGTT TGCACTCTAA TTAGATGGGC TTGAGGAAGT

7681 AATCCCTAAA ATCAATTCGC TATTAATATT TCGTATCAAT TAATAATAAT ATCAAAAAAT

7741 CTAACGATGT TCTTACAGAC CATGCTATTG CGGCTCCAAC AATTCCCCAA TGATAAATAA

7801 AAATATATAA TATGCATAAA TATGGGATAA CTTCGAGCAA ATGAATAATA GCTGTAATTT

7861 TTGATCTTCC ACTAGCCTGA ACTGAAACAA ATGGGATTTG TGCAATGCAA TTAAAAAAGA

7921 AACCTATTGC AAGAATTTTT AATACTATAC CTGGCGTCCC ATGATATGTA GGTCCCATCC

7981 AAGCGGACAT TATAAAATCT GATAAAATAA TTATCAACAT TACAATTGGA AGTATACCAA

8041 TAACCATTAT AAAATATGAT AATATTTTAG TTTGCTTTAC CGATTGCAAT TCTGAACTTA

8101 ATCTTGGAAA AATAGCTCTG GACAACGCAC TTGGTAATAT CGTTAAGCGT TGTATACCTT

8161 CAGACGGAGC AGTATAAAAA GAAACTTTAT CAGCCCCCAC AATGTGTGAA AGAATAAAAC

8221 GATCCATATA TGTCATAATA GGGCTAATAA TATTGCTAAC TGTTATCCAG CTTCCAAAGC

8281 CGATTAATCT TTTAACTGTT ACAATTTTTA CAGACAGCCC AGATGATATT ATTAGTTTTC

8341 GACTAAATAT AAAGGTCACT ATAAGTGATA AGACTCTTGC CATAACTAAA CCATATATAG

8401 CACTTAGTAA TCCTCCATGA AAAAAACAGA AAATCACTGG TAATCCAGCC ACAAAAGAGT

8461 TGTTAATTGA TTTTATTAAA TTTACTTTTC TGAACTTTTC CATCCCCTCA AAAATCCCCA

8521 ACCAGACTTG GTTTAACAAG TATAAGGGTA TGGTAGCTGA AATAATATAT ATTGCTTTGA

8581 CAGATTCTAC AACATGATTC GCGTTAATGT TTAATAATTT AACAATTACA TTGCTACTCA

8641 AAAATAGTAC ACTACCGCCA ATCAAGCCCA ATATAGTTAG AATTACCGTT GAAGTTGAAA

8701 TGATCGCTCT TAATTCTTTA TGAACATTTT TATATATTGA TACTTCTCTT ATAACAGCTC

8761 TGGTCAATCC AGCATCAAAA ATACTTGCAT ATCCAACTAA GGCAATAGCT AACGTAAAAA

8821 GGCCAAATTG CTCGGTCCCT AGAATTCTAG ACAGTATACC TAACGCAGGA ATTGCTATTA

8881 ATGATGGTAT AATATACCCA CTTATATTCC ATAAAGTATT CTTTACAATA CTCACAAAAA

8941 TAATTCCTTC ATGTTATGCA ATTCTTTAGC CCTTGCATCT TTAATCGATA AAATATAATT

9001 ATTATGTTCT ATCGTCGGCC ATTTTATGCT CAGAATAGGA TCATTCCATA CAATCCCTCT

9061 ATCACTATCA GGATGATAAT AGTTCGTCGT TTTATATAAA AATTCCGCAG TCTCGCTCAG

9121 CACCAAAAAA CCATGTGCAA ATCCCTCAGG GATCCACAAT TGCCGCTTAT TCTCAGCAGA

9181 TAAATTCACC CCAACCCATT TACCAAAGGT AGGCGACGAT TTACGAATAT CAACAGCTAC

9241 ATCAAAAACC TCACCAACAA CGCAACGTAC CAGTTTCCCT TGCGCATAAG GTTCTAACTG

9301 ATAATGCAGC CCGCGTAAAA CACCTTTACT AGACTTCGAA TGGTTATCCT GAACAAATTC

9361 AACCTTACGT CCTACAGCTT CTTCGAAAAC TTTCTGATTA AAGCTTTCCA TAAAGAAACC

9421 ACGCTCATCA CCAAAAACTT TCGGCTCGAA AATTAACACA TCAGGAATTT CTGTTTTAAT

9481 TACGTTCATT TTATTAATAA CCTTTAATCA TTTTCAGCAG ATACTGTCCA TAAGCATTTT

9541 TTTTCAGCGG CTCCGCTAAT GCTTTCACCT GTTCAGCATC AATAAACCCT TTACGGTAAG

9601 CAATTTCTTC TGGGCAGGAA ACCTTTAGTC CCTGGCGCTC TTCAATGGTG GCAATGAAGT

9661 TGCTTGCTTC AATAAGACTC TGATGTGTCC CCGTATCCAG CCATGCATAA CCACGCCCCA

9721 TCATGGCAAC GGATAAACGC CCCTGTTCCA TATAAATACG GTTAATATCG GTAATTTCCA

9781 GTTCACCACG GGCAGAAGGC TTAAGGTTTT TCGCCATTTC GACAACGTCG TTATCATAGA

9841 AATAAAGCCC GGTTACCGCA TAATTACTTT TTGGTTGTAG CGGTTTTTCT TCCAGGCTTA

9901 TTGCCGTACC GTTTTTATCA AACTCAACGA CGCCGTAGCG TTCAGGATCA TTAACGTGAT

9961 AGGCAAATAC CGTTGCACCA CTTTCTTTGT TAACAGCGAC ATCCATTAAC TTCGGCAGAT

10021 CATGACCGTA GAAGATATTA TCACCAAGAA CCAAAGCACA ATCATCACCA CCGATAAACT

10081 CTTCACCGAT AATAAACGCC TGCGCAAGCC CATCTGGAGT CGGTTGCACT TTGTACTGAA

10141 GATTTAGCCC CCACTGGCTA CCGTCACCTA GCAGTTGTTG AAAACGAGGA GTATCCTGTG

10201 GCGTACTAAT AATCAGAATA TCGCGAATAC CCGCCAACAT CAGTGTAGAG AGCGGGTAAT

10261 AGATCATCGG CTTATCATAA ATAGGTAATA GCTGTTTACT GACAGCCATA GTCACAGGAT

10321 AAAGACGTGT ACCAGAACCA CCCGCTAAAA TAATACCTTT ACGCGTTTTC ATTTCATCAT

10381 TCCTTTTAAT TCATCTTGCT CCACCATCAC GAACAAGATG CAAAAACTAT TAAATTGCTG

10441 TAGTCGTAAT TAATTCGTTG AGCATTCGTT TCACACCAAC CTGCCAGTCA GGCAAGACAA

10501 GCGCAAAGTT CTGCTGAAAT TTTTCTGTAT TAAGGCGAGA GTTATGTGGA CGACGAGCTG

10561 GTGTAGGATA GGCTGTTGTT GGTACTGCGT TGAGCTTGTT GAGTGCAAGG GGAATACCTG

10621 CTTTGCGCGC CTCTTCAAAA ACCAGCGCAG CATAATCGTG CCAGGTTGTG GTACCACTGG

10681 CTACCAGATG GTACAAACCT GCGACTTCCG GTTTATTCAG TGCCACACGA ATAGCATGTG

10741 CCGTACAATC AGCCAGCAGC TCAGCACCTG TTGGCGCACC AAATTGATCA TTTATCACAG

10801 CCAGTTCTTC GCGCTCTTTT GCCAGACGCA ACATCGTTTT GGCGAAGTTA TTTCCTTTAG

10861 CTGCGTATAC CCAGCTGGTA CGGAAAATAA GATGCTTCGC GCAATGTTCC TGTAACGCTT

10921 TTTCTCCGGC TAACTTGGTT TCACCGTAAA CATTTAGCGG TGCGGTTGCA TCCGTCTCCA

10981 GCCATGGCGT GTCGCCATTT CCAGGGAATA CGTAGTCAGT TGAGTAATGA ATTACCCAAG

11041 CCCCAACTTC ATTAGCCTCT TTTGCAATTG ATTCAACACT AGTCGCATTG AGTAATTGTG

11101 CAAATTCGGG TTCTGACTCA GCCTTATCTA CTGCGGTGTG AGCCGCAGCA TTAACAATAA

11161 CATCAGGTCG AATTCTTTTG ACTGTTTCAG CTACACCTTC AGGATTACTA AAATCACCAC

11221 AATAATCAGT GGAGTGAACA TCAAGAGCAA TCAAATTACC CAAAGGTGCC AGAGCACGCT

11281 GTAGTTCCCA ACCTACCTGC CCTGTTTTGC CGAAAAGGAG GATATTCATT ACTGGCGGCC

11341 CTCATAGTTC TGTTCAATCC ACGATTGATA AGCACCACTT TTCACATTAT CAACCCATTT

11401 TGTATTGGAC AGGTACCATT CCAATGTCTT CCGAATCCCG CTCTCAAACG TTTCCTGCGG

11461 TTTCCAGCCC AATTCGCGGC TAATCTTCTC TGCATCAATC GCATAACGGC GATCGTGTCC

11521 CGGGCGATCG GCAACATAAG TAATTTGCTC GCGGTAAGAT TTCTCTTTCG GTACAATCTC

11581 ATCCAGCAAA TCACAAATAG TGAGCACTAC ATCGATGTTT TTCTTTTCGT TGTGTCCACC

11641 AATGTTATAA GTTTCACCCG CTTTACCTTC GGTTACGACG GTATATAACG CACGCGCATG

11701 ATCTTCAACA TACAGCCAGT CACGAATTTG ATCCCCTTTG CCATAAATAG GTAATGCCTT

11761 ACCTTCCAGA GCATTCAGAA TAACCAATGG AATCAATTTT TCCGGGAAAT GATAAGGACC

11821 ATAATTATTA GAGCAATTAG TCACAATGGT TGGTAAACCA TAGGTACGTT TCCACGCGCG

11881 GACTAAATGA TCGCTGGATG CTTTTGAAGC GGAATAAGGG CTGCTTGGCG CGTAAGCTGT

11941 TGTCTCTGTA AATAAGGGTA ATTCTTCTGT ATTATTTACC TCGTCAGGAT GAGGCAAATC

12001 ACCATAGACT TCGTCAGTAG AAATATGATG AAAACGGAAT CTAGTTTTCT TGTCGCTATC

12061 AAGAGCAGAC CAATAATTGC GAGCGGCTTC CAAAAGGACA TATGTACCAA CAATATTGGT

12121 TTCAATAAAT GCCGCAGGAC CTGTAATTGA ACGGTCAACA TGGCTTTCAG CAGCCAGGTG

12181 CATCACTGCA TCTGGCTGAT GCTGAGCAAA AATCCGTGCC ATTGCAGCTG CATCGCAAAT

12241 ATCCGCATGT TCAAAAACAT AGCGTTCAGA ATCAGAAACA TCAGCAAGTG ATTCCAGGTT

12301 TCCGGCGTAC GTTAATTTAT CGACATTAAC AACACTATCC TGCGTATTAT TTATAATGTG

12361 ACGAACTACA GCAAAACCAA TAAATCCTGC GCCACCAGTA ACAAGTATTT TCACCTAATT

12421 TATTCCATAT TGCTTCAGAG CATGCTGTGA AATAAGCGGC TCTCAGTTTG ATTAATAGAA

12481 GTATTAATGC ACGCTACCGC CCCTGGCTTT ACAGCTACCA GAGCACTGCA TGCATGCCTA

12541 CGATGTGACG AGCGTTACCC ACTCGCGCTA AACCCGAAAA ATTCAAAAGC TAATTGTCTT

12601 ACCAATCCGC TCTGGAAACA AGGAAAATCC TGGAAAACTT TGACTAAAAT CCTATTGCTA

12661 ACTCGTTGTT ATTCTGATTG TTTATATAAA ACAACGGCAG GAATATTCGC AACAAATTAC

12721 TTTCACCACG AATCTTCACT GCCGTTATAA TTTTCTTATC AACCGTTACA TCCGGTCAGA

12781 TTTTCATTAT TCGCTTAACA GCTTCTCAAT ACCTTTACGG AACTTCGCCC CTTCTTTCAG

12841 GTTGCGCAGC CCATACTTCA CAAACGCCTG CATATAGCCC ATTTTTTTAC CGCAGTCGTA

12901 GCTGTCGCCG GTCATCAGCA TTGCATCAAC GGACTGTTTT TTCGCCAGCT CGGCAATGGC

12961 ATCAGTCAGC TGAATACGTC CCCATGCACC AGGCTGAGTA CGTTCAAGTT CCGGCCAAAT

13021 ATCGGCAGAA AGCACATAGC GACCAACGGC CATGATGTCT GAGTCCAGCG TCTGCGGCTG

13081 ATCCGGTTTT TCGATAAATT CAACAATGCG GCTGACTTTA CCTTCGCGAT CCAGCGGTTC

13141 TTTGGTCTGG ATGACGGAGT ATTCAGAGAG GTCACCCGGC ATACGTTTTG CCAGCACCTG

13201 GCTACGGCCC GTTTCATTGA AGCGCGCAAT CATGGCAGCA AGGTTGTAGC GTAGCGGGTC

13261 GGCGCTGGCG TCGTCGATCA CAACGTCTGG CAGCACCACG ACAAATGGAT TGTCACCAAT

13321 GGCGGGTCGT GCACACAAAA TGGAGTGACC TAAACCTAAA GGTTCGCCCT GACGCACGTT

13381 CATAATAGTC ACGCCCGGCG GGCAGATAGA TTGCACTTCC GCCAGTAGTT GACGCTTCAC

13441 GCGCTGCTCA AGGAGAGATT CTAATTCATA AGAGGTGTCG AAGTGGTTTT CGACCGCGTT

13501 CTTGGACGCA TGAGTTACCA GGAGGATTTC TTTGATCCCT GCAGCCACAA TCTCGTCAAC

13561 AATGTACTGA ATCATTGGCT TGTCGACGAT CGGTAGCATC TCTTTGGGTA TCGCCTTAGT

13621 GGCAGGCAAC ATATGCATCC CAAGACCCGC TACCGGTATA ACTGCTTTTA AATTCGTCAT

13681 TATTTTCCTA CCTCTAAGGG GCTGATAGTG CGTAAATTAT TGTCATAGGT TAGCCAAACG

13741 GTATGGCTAT ATACCAAGCA TAACTTTGAT TAAACCTTAC GATAACACTA CACACCATCA

13801 GCATCTGGGT TACTCGGATT ACTCGGAAAT CCACATACTG ATAATTTAAT CAGTACCTCT

13861 TTCCGAATAA TCGTAGTCCA ACCTGGTCCT TTTTTCTCTG ACTCGTCTGC ATTACTCAGA

13921 AACAAACGTT ATGTCGTCTT TTTTGGCATG GACGAATTCA TACTGCAGAG TTCGATCCAG

13981 ACCTTGCGAC AGCGTATACG GTGCAACAAA ACCTGAAGAA TGCACTTTCG TTGCGTCAAA

14041 CTGTGTTGTT GCGCAGAATT TTTTCACGCG CACAGAGCTG ACAGCGTATT TTTTGCCCGT

14101 AATTTTGCTC AGGATATCAA AGCAATATCC ACCCAGCATT CCTAGTGGGT AAGGCAAGTG

14161 CATAGAAGGG ATCTTTTTGT TCAGGCTTTG TTCAACTTCA GCAACCAACT GGTTCATGTT

14221 CAGGTCTGGC TTATCAACAT AGTTATAAAC CTCATAACCT GCGGCAACAT TCTTCAGTTT

14281 GTACTTGATA AACTCAACAA TGTTTCCAAC ATAAGCCATG GACTTATAGT TAGTCCCTGC

14341 GCCCACCATC ATAAACTTGC CGCCAGCGAT CTGTTTCAGC AAGTTATAGA CGTTACCGCG

14401 GTTGCGTTCA CCGAAGATAA CGGTAGGACG GATGATGGTT AATGAACGTT CTGTTGGTGC

14461 TTTGTTATAC CATTCACGCA GCACTTCCTC TGCCTGCCAC TTACTTTTGC CGTAGTGGTT

14521 GAAAGGGTCG TGTGGATGGT TTTCGTCAGG GTTGTGTTTG TTCAAACCAT AAACAGCAAC

14581 GGAACTGGTA AAGATGATAT TTTTAACGCC ATTTTTTTCC ATGGCCGCCA GCACATTGCG

14641 GGTACCCTGA ACGTTGACAT CATAATAGAG AGAAGTAGGG CTGACGTCAT CGCGGTGTTC

14701 CGCTGCCAGT AGTACAACAG TGTCAAAACC GGCTAACGCC TGGTCGAGTG CCTGTTGATC

14761 ACGAACATCA CCAATCTGTG TGATTTCTGG ATAAAAGTGG CTCTGCCGTT TGTCCAGGTT

14821 CTTGATATTA AAGTCAGCAA TTGCCGTTTC AAGTAGTCGG GTTCCTACGA ATCCGGAAGC

14881 TCCTATGAGC AAAACGTTAT TGTTCATAAA TCACTTTAGT CTGGTTGTTA CGTAAGAAAC

14941 ACAAGATAAA GATGAGTACC TTCCCTGAGT AGTCAATGCT GCCCAGCCCC AGCTTTAACA

15001 GTTAGTGTGA GGATTATAAT CTTTTAGAAC ATTATATCCA GTAAGTTTAT GAATGGTCGC

15061 AAATCTACTC TCTCCGTTCC GGCAATCTAA AGTTAATGCT AGCGACGTCG TGGGATCCTC

15121 TAGAGTCGAC CTGCAGGCAT GCAAGCTTGA GTATTCTATA GTCTCACCTA AATAGCTTGG

15181 CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA

15241 ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA

15301 CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC

15361 ATTAATGAAT CGGCCAACGC GAACCCCTTG CGGCCGCCCG GGCCGTCGAC CAATTCTCAT

15421 GTTTGACAGC TTATCATCGA ATTTCTGCCA TTCATCCGCT TATTATCACT TATTCAGGCG

15481 TAGCAACCAG GCGTTTAAGG GCACCAATAA CTGCCTTAAA AAAATTACGC CCCGCCCTGC

15541 CACTCATCGC AGTACTGTTG TAATTCATTA AGCATTCTGC CGACATGGAA GCCATCACAA

15601 ACGGCATGAT GAACCTGAAT CGCCAGCGGC ATCAGCACCT TGTCGCCTTG CGTATAATAT

15661 TTGCCCATGG TGAAAACGGG GGCGAAGAAG TTGTCCATAT TGGCCACGTT TAAATCAAAA

15721 CTGGTGAAAC TCACCCAGGG ATTGGCTGAG ACGAAAAACA TATTCTCAAT AAACCCTTTA

15781 GGGAAATAGG CCAGGTTTTC ACCGTAACAC GCCACATCTT GCGAATATAT GTGTAGAAAC

15841 TGCCGGAAAT CGTCGTGGTA TTCACTCCAG AGCGATGAAA ACGTTTCAGT TTGCTCATGG

15901 AAAACGGTGT AACAAGGGTG AACACTATCC CATATCACCA GCTCACCGTC TTTCATTGCC

15961 ATACGAAATT CCGGATGAGC ATTCATCAGG CGGGCAAGAA TGTGAATAAA GGCCGGATAA

16021 AACTTGTGCT TATTTTTCTT TACGGTCTTT AAAAAGGCCG TAATATCCAG CTGAACGGTC

16081 TGGTTATAGG TACATTGAGC AACTGACTGA AATGCCTCAA AATGTTCTTT ACGATGCCAT

16141 TGGGATATAT CAACGGTGGT ATATCCAGTG ATTTTTTTCT CCATTTTAGC TTCCTTAGCT

16201 CCTGAAAATC TCGATAACTC AAAAAATACG CCCGGTAGTG ATCTTATTTC ATTATGGTGA

16261 AAGTTGGAAC CTCTTACGTG CCGATCAACG TCTCATTTTC GCCAAAAGTT GGCCCAGGGC

16321 TTCCCGGTAT CAACAGGGAC ACCAGGATTT ATTTATTCTG CGAAGTGATC TTCCGTCACA

16381 GGTATTTATT CGCGATAAGC TCATGGAGCG GCGTAACCGT CGCACAGGAA GGACAGAGAA

16441 AGCGCGGATC TGGGAAGTGA CGGACAGAAC GGTCAGGACC TGGATTGGGG AGGCGGTTGC

16501 CGCCGCTGCT GCTGACGGTG TGACGTTCTC TGTTCCGGTC ACACCACATA CGTTCCGCCA

16561 TTCCTATGCG ATGCACATGC TGTATGCCGG TATACCGCTG AAAGTTCTGC AAAGCCTGAT

16621 GGGACATAAG TCCATCAGTT CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC

16681 TGCCCGGCAC CGGGTGCAGT TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA

16741 ATTATCCTGA GAATAAATGC CTTGGCCTTT ATATGGAAAT GTGGAACTGA GTGGATATGC

16801 TGTTTTTGTC TGTTAAACAG AGAAGCTGGC TGTTATCCAC TGAGAAGCGA ACGAAACAGT

16861 CGGGAAAATC TCCCATTATC GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG

16921 TTTATAGGAA GTAGTGTTCT GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA

16981 TGCCTTCAGG AACAATAGAA ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT

17041 CAGCAATGGA CAGAACAACC TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG

17101 CCAGTAGTGC TCGCCGCAGT CGAGCGACAG GGCGAAGCCC TCGGCTGGTT GCCCTCGCCG

17161 CTGGGCTGGC GGCCGTCTAT GGCCCTGCAA ACGCGCCAGA AACGCCGTCG AAGCCGTGTG

17221 CGAGACACCG CGGCCGGCCG CCGGCGTTGT GGATACCTCG CGGAAAACTT GGCCCTCACT

17281 GACAGATGAG GGGCGGACGT TGACACTTGA GGGGCCGACT CACCCGGCGC GGCGTTGACA

17341 GATGAGGGGC AGGCTCGATT TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC

17401 GAAAACGCCT GATTTTACGC GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT

17461 GCCCTGCGGT ATTGACACTT GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG

17521 ACACTTGAGG GGCAGAGTGC TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG

17581 TCCACAGGCA GAAAATCCAG CATTTGCAAG GGTTTCCGCC CGTTTTTCGG CCACCGCTAA

17641 CCTGTCTTTT AACCTGCTTT TAAACCAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC

17701 GCCCTGTGCG CGTGACCGCG CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCCTCCCG

17761 GTCGAGTGAG CGAGGAAGCA CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC

17821 CTGAAAAAAC TTCCCTTGGG GTTATCCACT TATCCACGGG GATATTTTTA TAATTATTTT

17881 TTTTATAGTT TTTAGATCTT CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT

17941 CTAGAGAAGG TGTTGTGACA AATTGCCCTT TCAGTGTGAC AAATCACCCT CAAATGACAG

18001 TCCTGTCTGT GACAAATTGC CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT

18061 TTTCACAAAG TTATCCCTGC TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAAACT

18121 TGTCACACTT CACATGGATC TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT

18181 AAAAATAGCC CGCGAATCGT CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC

18241 CGGGATCAAA AACGTATGCT GTATCTGTTC GTTGACCAGA TCAGAAAATC TGATGGCACC

18301 CTACAGGAAC ATGACGGTAT CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA

18361 TTGACCTCTG CGGAAGCCAG TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG

18421 GAAGTGGTTT TTTATCGCCC TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT

18481 CCTTGGTTTA TCAAACGTGC GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC

18541 CCATATCTCA TTCCCTTCTT TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT

18601 GAAACAAAAG AAATCACCAA TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT

18661 CGTAAGCCGG ATGGCTCAGG CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC

18721 CAGCTGCCTC AAAGTTACCA GCGTATGCCT GACTTCCGCC GCCGCTTCCT GCAGGTCTGT

18781 GTTAATGAGA TCAACAGCAG AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC

18841 CGCCAGACGA CTCATATCGT ATTTTCCTTC CGCGATATCA CTTCCATGAC GACAGGATAG

18901 TCTGAGGGTT ATCTGTCACA GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA

18961 GGGTAATTTG TCACAGTTTT GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT

19021 GAACTGTAAT TTTTAAGGAA GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCCTTCT

19081 CTTTCCCTTC GTCATGTGAC CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG

19141 ATTATCACAG TTTATTACTC TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT

19201 TCCCACGGTG GATATTTCTT CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT

19261 CTTTGCTTCC TCGCCAGTTC GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT

19321 GATAATAAGT GACTGAGGTA TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT

19381 TAAACAACTT TGCGGTTTTT TGATGACTTT GCGATTTTGT TGTTGCTTTG CAGTAAATTG

19441 CAAGATTTAA TAAAAAAACG CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAACTCATGG

19501 AAACACTTAA CCAGTGCATA AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC

19561 AGTTTAATGA TGACAGCCCG GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG

19621 CAGCGGATTT AGTTGGGGTT TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC

19681 TACCGCACCC GGATATGGAA ATTCGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG

19741 AACAAATTAA TCATATGCGT GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT

19801 TTCCACCGGT GATCGGGGTT GCTGCCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG

19861 TTCATCTTGC TCAGGATCTG GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG

19921 ACCCCCAGGG AACAGCCTCA ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG

19981 AAGACACTCT CCTGCCTTTC TATCTTGGGG AAAAGGACGA TGTCACTTAT GCAATAAAGC

20041 CCACTTGCTG GCCGGGGCTT GACATTATTC CTTCCTGTCT GGCTCTGCAC CGTATTGAAA

20101 CTGAGTTAAT GGGCAAATTT GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC

20161 GACTGGCCAT TGAAACTGTT GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA

20221 ACCTGGGTAT CGGCACGATT AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC

20281 CTGCTGAGTT GTTTGACTAC ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC

20341 TCAAGAACGT TGATCTTAAA GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA

20401 GCAATAGTAA TGGCTCTCAG TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA

20461 GCATGGTTCT AAAAAATGTT GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA

20521 TGAGAACTGT TTTTGAACAG GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG

20581 CTCTTTCTAT TTGGGAACCT GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT

20641 GGGAGATTAG ATAATGAAGC GTGCGCCTGT TATTCCAAAA CATACGCTCA ATACTCAACC

20701 GGTTGAAGAT ACTTCGTTAT CGACACCAGC TGCCCCGATG GTGGATTCGT TAATTGCGCG

20761 CGTAGGAGTA ATGGCTCGCG GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA

20821 GTTTACTCTT GAAGTGCTCC GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG

20881 TAATGAACGT GACCAGGAGC TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT

20941 TCTACTGACT GGTCAACAGA CACCGGCGTT CGGTCGAAGA GTATCTGGTG TCATAGAAAT

21001 TGCCGATGGG AGTCGCCGTC GTAAAGCTGC TGCACTTACC GAAAGTGATT ATCGTGTTCT

21061 GGTTGGCGAG CTGGATGATG AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG

21121 CCCAACAAGT GCTTATGAAC GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC

21181 TGGAAATATT TCTGCGCTGG CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG

21241 TATCAACACC GCCAAATTGC CTAAATCAGT TGTTGCTCTT TTTTCTCACC CCGGTGAACT

21301 ATCTGCCCGG TCAGGTGATG CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA

21361 GCAGCAGGCA TCTAACCTTC ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA

21421 AGTTATCACT CTTTTAACTT CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAG

21481 CTCACGACAT CAGTTTGCTC CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT

21541 TAACCTGGAC AGGTCTCGTG TTCCAACTGA GTGTATAGAG AAAATTGAGG CCATTCTTAA

21601 GGAACTTGAA AAGCCAGCAC CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT

21661 ACTTAATGTC CTTTGTTACA GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC

21721 CACTGTTCCA CTTGTATCGT CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT

21781 CGTCGGTCTG ATTATTAGTC TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA

21841 GTCTGGGACC ACGGTCCCAC TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC

21901 CACTCGTATC GTCGGTCTGA TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC

21961 TGATTATTAG TCTGGGACCA CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA

22021 CCACGGTCCC ACTCGTATCG TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA

22081 TCGTCGGTCT GATTATTAGT CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC

22141 GGTCTGGGAC CACGGTCCCA CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT

22201 CCATCAATGC CTGTCAAGGG CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA

22261 CCTCGGTGTG CGGTTGTATG CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT ATCCACAACA

22321 TTTTGCGCAC GGTTATGTGG ACAAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC

22381 CGGGCTGCAT CCGATGCAAG TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT

22441 TGCCCCGTAT TCAGTGTCGC TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT

22501 GCAGATCAAT TAATACGATA CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT

22561 CCACGCACGT TGTGATATGT AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA

22621 TCCGACAGGT TACGGGGCGG CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT

22681 TTAAGGCGTT TCCGTTCTTC TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA

22741 AAAGAAAGGA AACGACAGGT GCTGAAAGCG AGCTTTTTGG CCTCTGTCGT TTCCTTTCTC

22801 TGTTTTTGTC CGTGGAATGA ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG

22861 CATACATTAT ACGAAGTTAT ATTCGAT

Claims

1. A recombinant prokaryotic biosynthetic system comprising:

(i) a nucleic acid that encodes an epimerase that synthesizes N-acetylgalactosamine on undecaprenyl pyrophosphate, wherein said nucleic acid is at least 95% identical to SEQ ID NO: 1;

(ii) at least one nucleic acid that encodes a glycosyltransferase that is capable of assembling an oligo- or polysaccharide on a lipid carrier;

(iii) a nucleic acid that encodes an oligosaccharyl transferase from *Campylobacter jejuni*; and

(iv) a nucleic acid that encodes a protein comprising one or more of an introduced consensus sequence, D/E – X–N– Z – S/T, wherein X and Z can be any natural amino acid except proline,

wherein the oligosaccharide assembled by said glycosyltransferases can be transferred from the lipid carrier to the protein by the oligosaccharyl transferase.

2. The recombinant prokaryotic biosynthetic system of claim 1, wherein said nucleic acid encoding the epimerase consists of SEQ ID NO. 1.

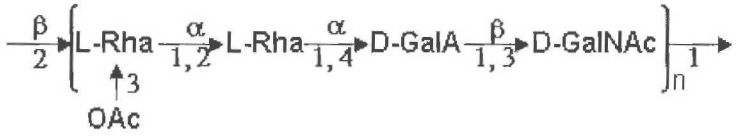
3. The recombinant prokaryotic biosynthetic system of claim 1, wherein said protein is *P. aeruginosa* exoprotein.

4. The recombinant prokaryotic biosynthetic system of claim 1, wherein said oligo- or polysaccharide is from a Gram-negative bacterium.

5. The recombinant prokaryotic biosynthetic system of claim 4, wherein said oligo- or polysaccharide is from *Shigella flexneri*.

6. The recombinant prokaryotic biosynthetic system of claim 5, wherein said oligo- or polysaccharide is from *Shigella flexneri* 6.

7. The recombinant prokaryotic biosynthetic system of claim 1, wherein said oligo- or polysaccharide comprises a structure:



8. The recombinant prokaryotic biosynthetic system of claim 4, wherein said oligo- or polysaccharide is from *E. coli*.

9. The recombinant prokaryotic biosynthetic system of claim 8, wherein said oligo- or polysaccharide is from *E. coli* O157.

10. The recombinant prokaryotic biosynthetic system of claim 1, wherein said oligo- or polysaccharide comprises a structure, α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc.

11. A method of producing an N-linked glycosylated protein, comprising:

a.) introducing into a host organism:

i.) a nucleic acid that encodes an epimerase that synthesizes N-acetylgalactosamine on undecaprenyl pyrophosphate, wherein said nucleic acid is at least 95% identical to SEQ ID NO: 1;

ii.) at least one nucleic acid that encodes a glycosyltransferase that is capable of assembling an oligo- or polysaccharide on a lipid carrier;

iii.) a nucleic acid that encodes an oligosaccharyl transferase from *Campylobacter jejuni*; and

iv.) a nucleic acid that encodes a protein comprising one or more of an introduced consensus sequence, D/E – X–N– Z – S/T, wherein X and Z can be any natural amino acid except proline; and

b.) culturing said host organism until at least one N-glycosylated protein is produced.

12. The method of claim 11, wherein said nucleic acid encoding the epimerase consists of SEQ ID NO. 1.

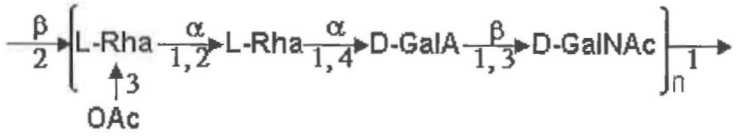
13. The method of claim 11, wherein said protein is *P. aeruginosa* exoprotein.

14. The method of claim 11, wherein said oligo- or polysaccharide is from a Gram-negative bacterium.

15. The method of claim 14, wherein said oligo- or polysaccharide is from *Shigella flexneri*.

16. The method of claim 15, wherein said oligo- or polysaccharide is from *Shigella flexneri* 6.

17. The method of claim 11, wherein said oligo- or polysaccharide comprises a structure:



18. The method of claim 14, wherein said oligo- or polysaccharide is from *E. coli*.

19. The method of claim 18, wherein said oligo- or polysaccharide is from *E. coli* O157.

20. The method of claim 11, wherein said oligo- or polysaccharide comprises a structure, α-D-PerNAc-α-L-Fuc-β-D-Glc-α-D-GalNAc.