# GLYCOSYLATED PROTEIN EXPRESSION IN PROKARYOTES

[0001] This application claims priority benefit of U.S. Provisional Patent Application Serial No. 61/018,772, filed January 3, 2008, which is hereby incorporated by reference in its entirety.

## FIELD OF INVENTION

[0002] The present invention relates to glycosylated protein expression in prokaryotes.

## BACKGROUND OF THE INVENTION

### Glycotherapeutics

[0003] Protein-based therapeutics currently represent one in every four new drugs approved by the FDA (Walsh, G., "Biopharmaceutical Benchmarks," *Nat Biotechnol* 18:831-3 (2000); Walsh, G, "Biopharmaceutical Benchmarks," *Nat Biotechnol* 21:865-70 (2003); and Walsh, G, "Biopharmaceutical Benchmarks," *Nat Biotechnol* 24:769-76 (2006)).

[0004] While several protein therapeutics can be produced using a prokaryotic expression system such as *E. coli* (e.g., insulin), the vast majority of therapeutic proteins require additional post-translational modifications, thought to be absent in prokaryotes, to attain their full biological function. In particular, *N*-linked protein glycosylation is predicted to affect more than half of all eukaryotic protein species (Apweiler et al., "On the Frequency of Protein Glycosylation, as Deduced From Analysis of the SWISS-PROT Database," *Biochim Biophys Acta* 1473:4-8 (1999)) and is often essential for proper folding, pharmacokinetic stability, tissue targeting and efficacy for a large number of proteins (Helenius et al., "Intracellular Functions of N-linked Glycans," *Science* 291:2364-9 (2001)). Since most bacteria do not glycosylate their own proteins, expression of most therapeutically relevant glycoproteins, including antibodies, is

relegated to mammalian cells. However, mammalian cell culture suffers from a number of drawbacks including: (i) extremely high manufacturing costs and low volumetric productivity of eukaryotic hosts, such as CHO cells, relative to bacteria; (ii) retroviral contamination; (iii) the relatively long time required to generate stable cell lines; (iv) relative inability to rapidly generate stable, "high-producing" eukaryotic cell lines via genetic modification; and (v) high product variability created by glycoform heterogeneity that arises when using host cells, such as CHO, that have endogenous non-human glycosylation pathways (Choi et al., "Use of Combinatorial Genetic Libraries to Humanize N-linked Glycosylation in the Yeast Pichia pastoris," *Proc Natl Acad Sci U S A* 100:5022-7 (2003)). Expression in *E. coli*, on the other hand, does not suffer from these limitations.

**Expression of a glycosylated therapeutic proteins in *E. coli***

**[0005]** Many therapeutic recombinant proteins are currently expressed using *E. coli* as a host organism. One of the best examples is human insulin, which was first produced in *E. coli* by Eli Lilly in 1982. Since that time, a vast number of human therapeutic proteins have been approved in the U.S. and Europe that rely on *E. coli* expression, including human growth hormone (hGH), granulocyte macrophage colony stimulating factor (GM-CSF), insulin-like growth factor (IGF-1, IGFBP-3), keratinocyte growth factor, interferons (IFN-α, IFN-β1b, IFN-γ1b), interleukins (IL-1, IL-2, IL-11), tissue necrosis factor (TNF-α), and tissue plasminogen activator (tPA). However, almost all glycoproteins are produced in mammalian cells. When a protein that is normally glycosylated is expressed in *E. coli*, the lack of glycosylation in that host can yield proteins with impaired function. For instance, aglycosylated human monoclonal antibodies (mAbs) (e.g., anti-tissue factor IgG1) can be expressed in soluble form and at high levels in *E. coli* (Simmons et al., "Expression of Full-length Immunoglobulins in Escherichia coli: Rapid and Efficient Production of Aglycosylated Antibodies," *J Immunol Methods* 263:133-47 (2002)). However, while *E. coli*-derived mAbs retained tight binding to their cognate antigen and neonatal receptor and exhibited a circulating half-life comparable to mammalian cell-derived antibodies, they were incapable of binding to C1q and the FcγRI receptor due to the absence of *N*-glycan.

**Eukaryotic and prokaryotic *N*-linked protein glycosylation**

**[0006]** *N*-linked protein glycosylation is an essential and conserved process occurring in the endoplasmic reticulum (ER) of eukaryotic organisms (Burda et al., "The Dolichol Pathway of N-linked Glycosylation," *Biochim Biophys Acta* 1426:239-57 (1999)). It is important for protein folding, oligomerization, quality control, sorting, and transport of secretory and membrane proteins (Helenius et al., "Intracellular Functions of N-linked Glycans," *Science* 291:2364-9 (2001)). The eukaryotic *N*-linked protein glycosylation pathway (Figure 1) can be divided into two different processes: (i) the assembly of the lipid-linked oligosaccharide at the membrane of the endoplasmic reticulum and (ii) the transfer of the oligosaccharide from the lipid anchor dolichyl pyrophosphate to selected asparagine residues of nascent polypeptides. The characteristics of *N*-linked protein glycosylation, namely (i) the use of dolichyl pyrophosphate (Dol-PP) as carrier for oligosaccharide assembly, (ii) the transfer of only the completely assembled $Glc_3Man_9GlcNAc_2$ oligosaccharide, and (iii) the recognition of asparagine residues characterized by the sequence N-X-S/T where N is asparagine, X is any amino acid except proline, and S/T is serine/threonine (Gavel et al., "Sequence Differences Between Glycosylated and Non-glycosylated Asn-X-Thr/Ser Acceptor Sites: Implications for Protein Engineering," *Protein Eng* 3:433-42 (1990)) are highly conserved in eukaryotes. The oligosaccharyltransferase (OST) catalyzes the transfer of the oligosaccharide from the lipid donor dolichylpyrophosphate to the acceptor protein. In yeast, eight different membrane proteins have been identified that constitute the complex *in vivo* (Kelleher et al., "An Evolving View of the Eukaryotic Oligosaccharyltransferase," *Glycobiology* 16:47R-62R (2006)). STT3 is thought to represent the catalytic subunit of the OST (Nilsson et al., "Photocross-linking of Nascent Chains to the STT3 Subunit of the Oligosaccharyltransferase Complex," *J Cell Biol* 161:715-25 (2003) and Yan et al., "Studies on the Function of Oligosaccharyl Transferase Subunits. Stt3p is Directly Involved in the Glycosylation Process," *J Biol Chem* 277:47692-700 (2002)). It is the most conserved subunit in the OST complex (Burda et al., "The Dolichol Pathway of N-linked Glycosylation," *Biochim Biophys Acta* 1426:239-57 (1999)).

**[0007]** Conversely, the lack of glycosylation pathways in bacteria has greatly restricted the utility of prokaryotic expression hosts for making therapeutic proteins, especially since by certain estimates "more than half of all proteins in nature will eventually be found to be glycoproteins" (Apweiler et al., "On the Frequency of Protein Glycosylation, as Deduced From Analysis of the SWISS-PROT Database," *Biochim Biophys Acta* 1473:4-8 (1999)). Recently, however, it was discovered that the genome of a pathogenic bacterium, *C. jejuni*, encodes a pathway for *N*-linked protein glycosylation (Szymanski et al., "Protein Glycosylation in Bacterial Mucosal Pathogens," *Nat Rev Microbiol* 3:225-37 (2005)). The genes for this pathway, first identified in 1999 by Szymanski and coworkers (Szymanski et al., "Evidence for a System of General Protein Glycosylation in Campylobacter jejuni," *Mol Microbiol* 32:1022-30 (1999)), comprise a 17-kb locus named *pgl* for protein glycosylation. Following discovery of the *pgl* locus, in 2002 Linton *et al.* identified two *C. jejuni* glycoproteins, PEB3 and CgpA, and showed that *C. jejuni*-derived glycoproteins such as these bind to the *N*-acetyl galactosamine (GalNAc)-specific lectin soybean agglutinin (SBA) (Linton et al., "Identification of N-acetylgalactosamine-containing Glycoproteins PEB3 and CgpA in Campylobacter jejuni," *Mol Microbiol* 43:497-508 (2002)). Shortly thereafter, Young *et al.* identified more than 30 potential *C. jejuni* glycoproteins, including PEB3 and CgbA, and used mass spectrometry and NMR to reveal that the *N*-linked glycan was a heptasaccharide with the structure GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,*N*-Asn ($GalNAc_5GlcBac$, where Bac is bacillosamine or 2,4-diacetamido-2,4,6-trideoxyglucose) (Young et al., "Structure of the N-linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, Campylobacter jejuni," *J Biol Chem* 277:42530-9 (2002)) (Figure 2). The branched heptasaccharide is synthesized by sequential addition of nucleotide-activated sugars on a lipid carrier undecaprenylpyrophosphate on the cytoplasmic side of the inner membrane (Feldman et al., "Engineering N-linked Protein Glycosylation with Diverse O Antigen Lipopolysaccharide Structures in Escherichia coli," *Proc Natl Acad Sci U S A* 102:3016-21 (2005)) and, once assembled, is flipped across the membrane by the putative ATP-binding cassette (ABC) transporter WlaB (Alaimo et al., "Two Distinct But

Interchangeable Mechanisms for Flipping of Lipid-linked Oligosaccharides," *Embo J* 25:967-76 (2006) and Kelly et al., "Biosynthesis of the N-linked Glycan in Campylobacter jejuni and Addition Onto Protein Through Block Transfer," *J Bacteriol* 188:2427-34 (2006)). Next, transfer of the heptasaccharide to substrate proteins in the periplasm is catalyzed by an OST named PglB, a single, integral membrane protein with significant sequence similarity to the catalytic subunit of the eukaryotic OST STT3 (Young et al., "Structure of the N-linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, Campylobacter jejuni," *J Biol Chem* 277:42530-9 (2002)). PglB attaches the heptasaccharide to asparagine in the motif D/E-$X_1$-N-$X_2$-S/T (where D/E is aspartic acid/glutamic acid, $X_1$ and $X_2$ are any amino acids except proline, N is asparagine, and S/T is serine/threonine), a sequon similar to that used in the eukaryotic glycosylation process (N-X-S/T) (Kowarik et al., "Definition of the Bacterial N-glycosylation Site Consensus Sequence," *Embo J* 25:1957-66 (2006)).

**Glycoengineering of microorganisms**

**[0008]** A major problem encountered when expressing therapeutic glycoproteins in mammalian, yeast, or even bacterial host cells is the addition of non-human glycans. For instance, yeast, one of the two most frequently used systems for the production of therapeutic glycoproteins, transfer highly immunogenic mannan-type *N*-glycans (containing up to one hundred mannose residues) to recombinant glycoproteins. Mammalian expression systems can also modify therapeutic proteins with non-human sugar residues, such as the *N*-glycosylneuraminic acid (Neu5Gc) form of sialic acid (produced in CHO cells and in milk) or the terminal α(1,3)-galactose (Gal) (produced in murine cells). Repeated administration of therapeutic proteins carrying non-human sugars can elicit adverse reactions, including an immune response in humans.

**[0009]** As an alternative to using native glycosylation systems for producing therapeutic glycoproteins, the availability of glyco-engineered expression systems could open the door to customizing the glycosylation of a therapeutic protein and could lead to the development of improved therapeutic glycoproteins. Such a system would have the potential to eliminate undesirable glycans and perform human glycosylation to a high degree of homogeneity. To date, only the yeast *Pichia pastoris* has been glyco-

engineered to provide an expression system with the capacity to control and optimize glycosylation for specific therapeutic functions (Gerngross, T. U., "Advances in the Production of Human Therapeutic Proteins in Yeasts and Filamentous fungi," *Nat Biotechnol* 22:1409-14 (2004); Hamilton et al., "Glycosylation Engineering in Yeast: The Advent of Fully Humanized Yeast," *Curr Opin Biotechnol* 18:387-92 (2007); and Wildt et al., "The Humanization of N-glycosylation Pathways in Yeast," *Nat Rev Microbiol* 3:119-28 (2005)).

**[0010]** For example, a panel of glyco-engineered *P. pastoris* strains was used to produce various glycoforms of the monoclonal antibody Rituxan (an anti-CD20 IgG1 antibody) (Li et al., "Optimization of Humanized IgGs in Glycoengineered Pichia pastoris," *Nat Biotechnol* 24:210-5 (2006)). Although these antibodies share identical amino acid sequences to commercial Rituxan, specific glycoforms displayed ~100-fold higher binding affinity to relevant FcγRIII receptors and exhibited improved *in vitro* human B-cell depletion (Li et al., "Optimization of Humanized IgGs in Glycoengineered Pichia pastoris," *Nat Biotechnol* 24:210-5 (2006)). The tremendous success and potential of glyco-engineered *P. pastoris* is not without some drawbacks. For instance, in yeast and all other eukaryotes *N*-linked glycosylation is essential for viability (Herscovics et al., "Glycoprotein Biosynthesis in Yeast," *FASEB J* 7:540-50 (1993) and Zufferey et al., "STT3, a Highly Conserved Protein Required for Yeast Oligosaccharyl Transferase Activity In Vivo," *EMBO J* 14:4949-60 (1995)). Thus, the systematic elimination and re-engineering by Gerngross and coworkers of many of the unwanted yeast *N*-glycosylation reactions (Choi et al., "Use of Combinatorial Genetic Libraries to Humanize N-linked Glycosylation in the Yeast Pichia pastoris," *Proc Natl Acad Sci U S A* 100:5022-7 (2003)) has resulted in strains that are "sick" compared to their wild-type progenitor. This can be worsened during high-level glycoprotein expression due to the large metabolic burden placed on the yeast glycosylation system. As a result, the cell yield that can be obtained during large-scale fermentation is limited. Furthermore, elimination of the mannan-type *N*-glycans is only half of the glycosylation story in yeast. This is because yeast also perform *O*-linked glycosylation whereby *O*-glycans are linked to Ser or Thr residues in glycoproteins (Gentzsch et al., "The PMT Gene Family: Protein O-

glycosylation in Saccharomyces cerevisiae is Vital," *EMBO J* 15:5752-9 (1996)). As with *N*-linked glycosylation, *O*-glycosylation is essential for viability (Gentzsch et al., "The PMT Gene Family: Protein O-glycosylation in Saccharomyces cerevisiae is Vital," *EMBO J* 15:5752-9 (1996)) and thus cannot be genetically deleted from glyco-engineered yeast. Since there are differences between the *O*-glycosylation machinery of yeast and humans, the possible addition of *O*-glycans by glyco-engineered yeast strains has the potential to provoke adverse reactions including an immune response.

**[0011]** Recently, Aebi and his coworkers transferred the *C. jejuni* glycosylation locus into *E. coli* and conferred upon these cells the extraordinary ability to post-translationally modify proteins with *N*-glycans (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002)). However, despite the functional similarity shared by the prokaryotic and eukaryotic glycosylation mechanisms, the oligosaccharide chain attached by the prokaryotic glycosylation machinery ($GalNAc_5GlcBac$) is structurally distinct from that attached by eukaryotic glycosylation pathways (Szymanski et al., "Protein Glycosylation in Bacterial Mucosal Pathogens," *Nat Rev Microbiol* 3:225-37 (2005); Young et al., "Structure of the N-linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, Campylobacter jejuni," *J Biol Chem* 277:42530-9 (2002); and Weerapana et al., "Asparagine-linked Protein Glycosylation: From Eukaryotic to Prokaryotic Systems," *Glycobiology* 16:91R-101R (2006)). Numerous attempts (without success) have been made to reprogram *E. coli* with a eukaryotic *N*-glycosylation pathway to express *N*-linked glycoproteins with structurally homogeneous human-like glycans.

**[0012]** The present invention is directed to overcoming the deficiencies in the art.

## SUMMARY OF THE INVENTION

**[0013]** A first aspect of the present invention relates to a prokaryotic host cell comprising eukaryotic glycosyltransferase activity, where the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity.

**[0014]** One aspect of the present invention is directed to a glycoprotein conjugate comprising a protein and at least one peptide comprising a D-$X_1$-N-$X_2$-T motif fused to the protein, where D is aspartic acid, $X_1$ and $X_2$ are any amino acid other than proline, N is asparagine, and T is threonine.

**[0015]** Another aspect of the present invention is directed to a method of producing a glycosylated protein. This method comprises providing a prokaryotic host cell comprising eukaryotic glycosyltransferase activity, where the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity. The prokaryotic host cell is then cultured under conditions effective to produce a glycosylated protein.

**[0016]** A further aspect of the present invention pertains to a method for screening bacteria or bacteriophages. This method involves expressing one or more glycans on the surface of a bacteria and attaching a label on the one or more glycans on the surface of the bacteria or on the surface of a bacteriophage derived from the bacteria. The label is then analyzed in a high-throughput format.

**[0017]** Another aspect of the present invention relates to a glycosylated antibody comprising an Fv portion which recognizes and binds to a native antigen and an Fc portion which is glycosylated at a conserved asparagine residue.

**[0018]** One aspect of the present invention relates to a reprogrammed prokaryotic host with a *N*-glycosylation pathway to express *N*-linked glycoproteins with structurally homogeneous human-like glycans. Prokaryotic host cells can comprise glycosyltransferase activities in the form of a dolichyl-linked UDP-GlcNAc transferase and a mannosyltransferase. In some embodiments, the UDP-GlcNAc transferase comprises *alg13* and *alg14* gene activity. In other embodiments, the mannosyltransferase comprise *alg1* and *alg2* gene activity. In additional embodiments, the prokaryotic host cell comprises a flippase activity including *pglK* and *rft1*. In further embodiments, the prokaryotic host cell comprises at least one oligosaccharyl transferase activity, such as *pglB* and *STT3*.

**[0019]** In preferred aspects, the present invention commercializes technologies for the design, discovery, and development of glycoprotein diagnostics and therapeutics.

Specifically, the present invention provides for the development of a low-cost strategy for efficient production of authentic human glycoproteins in microbial cells with the potential to revolutionize the enterprise surrounding the manufacturing of therapeutic proteins. In various aspects, the glyco-engineered bacteria of the invention are capable of stereospecific production of *N*-linked glycoproteins. In one embodiment, bacteria have been genetically engineered with a collection of genes encoding a novel glycosylation pathway that is capable of efficiently glycosylating target proteins at specific asparagine acceptor sites (e.g., *N*-linked glycosylation). Using these specially engineered cell lines, virtually any recombinant protein-of-interest can be expressed and glycosylated, thus, production of numerous authentic human glycoproteins is possible.

**[0020]** Further, the invention provides proprietary platform technologies for engineering permutations of sugar structures, thereby enabling for the first time "bacterial glycoprotein engineering." One expectation of glycoengineering - the intentional manipulation of protein-associated carbohydrates to alter pharmacokinetic properties of proteins – is to elucidate the role of glycosylation in biological phenomena. Accordingly, in various aspects, the invention provides biotechnological synthesis of novel glycoconjugates and immunostimulating agents for research, industrial, and therapeutic applications.

**[0021]** The major advantage of *E. coli* as a host for glycoprotein expression is that, unlike yeast and all other eukaryotes, there are no native glycosylation systems. Thus, the addition (or subsequent removal) of glycosylation-related genes should have little to no bearing on the viability of glyco-engineered *E. coli* cells. Furthermore, the potential for non-human glycan attachment to target proteins by endogenous glycosylation reactions is eliminated in these cells.

**[0022]** Accordingly, in various *embodiments*, an alternative for glycoprotein expression is disclosed where a prokaryotic host cell is used to produce *N*-linked glycoproteins, which provides an attractive solution for circumventing the significant hurdles associated with eukaryotic cell culture. The use of bacteria as a production vehicle is expected to yield structurally homogeneous human-like *N*-glycans while at the

same time dramatically lowering the cost and time associated with protein drug development and manufacturing.

**[0023]** Other key advantages include: (i) the massive volume of data surrounding the genetic manipulation of bacteria; (ii) the established track record of using bacteria for protein production -- 30% of protein therapeutics approved by the FDA since 2003 are produced in *E. coli* bacteria; and (iii) the existing infrastructure within numerous companies for bacterial production of protein drugs.

**[0024]** In comparison to various eukaryotic protein expression systems, the process employed using the methods and composition of the invention provides a scalable, cost-effective, optimal recombinant glycoprotein expression, free of human pathogens, free of immunogenic *N*- and *O*-linked glycosylation reactions, capable of rapid cloning and fast growth rate, fast doubling time (~20 minutes), high growth (high OD), high titer and protein yields (in the range of 50% of the total soluble protein (TSP)), ease of product purification from the periplasm or supernatant, genetically tractable, thoroughly studied, compatible with the extensive collection of expression optimization methods (e.g., promoter engineering, mRNA stabilization methods, chaperone co-expression, protease depletion, etc.).

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0025]** Figure 1 illustrates a biosynthetic scheme of the lipid-linked oligosaccharide and the transfer to protein at the membrane of the endoplasmic reticulum in *S. cerevisiae*. The loci required for the individual reactions are indicated. The origin of the mannose residues either from GDP-mannose directly (light shading) or from dolichylphosphomannose (dark shading) is indicated. See Burda et al., "The Dolichol Pathway of N-linked Glycosylation," *Biochim Biophys Acta* 1426:239-57 (1999), which is hereby incorporated by reference in its entirety.

**[0026]** Figure 2 illustrates the biosynthesis science of *N*-linked glycoproteins in bacteria. In *C. jejuni*, *N*-linked glycosylation proceeds through the sequential addition of nucleotide-activated sugars onto a lipid carrier, resulting in the formation of a branched

heptasaccharide. This glycan is then flipped across the inner membrane by PglK (formerly WlaB) and the OTase PglB then catalyzes the transfer of the glycan to an asparagine side chain. Bac is 2,4-diacetamido-2,4,6-trideoxyglucose; GalNAc is *N*-acetylgalactosamine; HexNAc is *N*-acetylhexosamine; Glc is glucose. See Szymanski et al., "Protein Glycosylation in Bacterial Mucosal Pathogens," *Nat Rev Microbiol* 3:225-37 (2005) , which is hereby incorporated by reference in its entirety.

**[0027]** Figures 3A-B are photos of Western blots of glycosylated PEB3 in glyco-engineered *E. coli*. The *C. jejuni* glycosylation substrate PEB3, carrying a C-terminal 6x his tag, was expressed and purified from the periplasm of *E. coli* cells that co-expressed either the complete set of *pgl* genes from pACYC184-*pgl* (pgl+) or a modified *pgl* gene cluster that lacked the *pglB* gene encoding the essential OTase (pgl-). Purified PEB3 was detected in both pgl+ and pgl- cells, as evidenced by Western blotting using an anti-polyhistidine antibody (Figure 3A). However, PEB3 was only glycosylated in pgl+ cells based on binding to the GalNAc-specific lectin SBA, whereas PEP3 from pgl- cells was aglycosylated (Figure 3B). Purified PEB3 was serially-diluted as indicated.

**[0028]** Figures 4A-D show the results of the glycosylation of *E. coli* maltose binding protein (MBP). Figure 4A shows a peptide glycosylation tag. Figure 4B shows an anti-His Western blot of (left-to-right) MBP with a C-terminal GlycTag (GT), the *C. jejuni* glycoprotein cjAcrA, MBP with an N-terminal GT, MBP C-terminal GT without a secretion signal peptide, and MBP & GFP each with a C-terminal GT and a Tat-specific (ssTorA) signal peptide. Proteins were Ni-purified from glyco-engineered *E. coli* (pgl+, except lane 2) and immunoblotted with anti-HIS serum. Figure 4C shows a Western blot against the bacterial heptasaccharide using anti-Hept serum. Figure 4D shows at least three discrete bands characteristic of multiple *N*-glycans for MBP C-terminal GT (left) and MBP N-terminal GT (right).

**[0029]** Figures 5A-C depict the results of glycosylated IgG M18.1 in pgl+ *E. coli*. Figure 5A shows the glycosylation at Asn297 in $C_H2$ results in a conformational shift in the Fc region of the IgG that endows binding to the appropriate receptor molecules to elicit effector function. Western blot analysis of IgG M18.1 purified from pgl- (Figure 5B) and pgl+ (Figure 5C) *E. coli* using Protein-A-G resin columns (Pierce). Samples

were run in non-reducing 12% SDS gels and immunoblotted with anti-human IgG and hR6P antiserum.

**[0030]** Figure 6A shows a schematic of glycoprotein surface display and Western blot analysis confirming glycosylation of CjaA in pgl+ and pgl- cells via glycoprotein-specific antiserum (hR6P). Figure 6B shows the transfer of heptasaccharide to the outer surface via WaaL-mediated ligation to lipid A. Figure 6C shows the quantification of SBA-Alexa Fluor labeling using flow cytometry.

**[0031]** Figure 7 is a schematic example of the glycophage system in accordance with the present invention. Plasmids or phagemids encoding the proteins for lipid linked oligosaccharide synthesis, for the oligosaccharide transfer (OTase), and for the acceptor scFv-g3p fusion protein are shown. The oligosaccharide is assembled on a lipid carrier, bactoprenylpyrophosphate, at the cytoplasmic site (Cyt) of the plasma membrane (catalyzed by individual glycosyltransferases). The oligosaccharide is then translocated across the inner membrane (IM) to the periplasmic space (Per) and transferred to specific asparagine residues of the acceptor protein by the oligosaccharyltransferase. After infection with helper phage VCSM13, phages that display the glycosylated acceptor protein (glycophage) are bound to immobilized soybean agglutinin (SBA) and eluted with galactose. Glycophages which have been eluted are used to infect *E. coli* (F+) cells selected for the antibiotic resistance present on the phagemid. The glyco-phenotype of the phage can be connected to the genotype of any of the required steps according to the presence of the ori M13 and the subsequent packaging of the phagemid into phage particles. Dhfr is dihydrofolate reductase; bla is β-lactamase; cat is chloramphenicol acetyltransferase.

**[0032]** Figures 8A-D represent time-dependent expression and glycosylation of AcrA-g3p in pgl+ (Figures 8A and 8C) and pglmut (Figure 8B and 8D) cells visualized by immunodetection. Whole cell lysates were prepared from either non-induced cells (lane 1) or from cells induced with 50 mM arabinose for 1 h (lane 2), 3 h (lane 3), 5 h (lane 4), and 16 h (lane 5). Proteins were separated by 10% SDS-PAGE and transferred to a nitrocellulose membrane. AcrA-g3p and glycosylated AcrA-g3p (glyco-AcrA) were

visualized with AcrA-specific antibodies (Figures 8A-8B) or with R12 antiserum (Figure 8C-8D). MW markers are indicated on the right.

**[0033]** Figure 9A shows quantification of glycophage enrichment by SBA biopanning. Phages produced from either glycosylation-competent (pgl, black bars) or glycosylation-incompetent (pglmut, grey bars) cells were applied to SBA-column purification. The values that represent the total amount of colony forming units (cfu) present within each fraction of the SBA panning procedure, as determined after infection of TG1 cells, are the means of at least three independent experiments. The amount of phages applied to SBA panning and the resulting cfus after *E. coli* infection varied by less than 6%. Fraction 1, cfu applied to the SBA column; fraction 2, SBA flow-through; fractions 3 and 4, PBS washing steps; fraction 5, 6, and 7, washing steps with 30 mM galactose in PBS; fraction 8, 9, and 10, elution steps with 300 mM galactose in PBS. Figure 9B is a photograph of immunodetection of AcrA-g3p and glycosylated AcrA-g3p (glyco-AcrA-g3p) displayed on phages. Phages were produced from pgl+ (panels a, c) or pglmut cells (panels b, d) and applied to SBA panning. The presence of AcrA and glyco-AcrA was visualized with anti-AcrA (panels a, b) or with R12 antiserum (panels c, d). Lane 1, raw phage preparation; lane 2, SBA flow-through; lanes 3 and 4, wash fractions with PBS; lanes 5 to 7, wash fractions with 30 mM galactose in PBS; lanes 8 to 10, elution fractions with 300 mM galactose in PBS. In lanes 1 to 4, $1x10^8$ phages were applied to SDS-PAGE. In lanes 5 to 10, $3.5 \times 10^7$, $1.2 \times 10^4$, $4.0x10^3$, $1.3x\ 10^6$, $2.5x10^6$, $1.2x10^6$ phages prepared from pgl+ (panels a, c) or $1.5x10^6$, $3.5x10^3$, $3.0x10^3$, $4.5x10^3$, $0.5x10^4$, $1.5x10^3$ phages prepared from pglmut cells (panels b, d) were used, respectively. MW markers are indicated on the right. The amount of phages obtained by SBA panning and applied to SDS-PAGE varied by less than ± 6%.

**[0034]** Figures 10A-10B are schematic drawings depicting glycan engineering in *E. coli*. Figure 10A shows the evolutionary trajectory from bacterial to mammalian glycoforms. Figure 10B shows the pathway for biosynthesis and transfer of $Man_3GlcNAc_2$ core glycoform to bacterial substrate proteins.

**[0035]** Figures 11A-B are photographs of Western blots, depicting expression of Alg13/14 in *E. coli*. Figure 11A shows a Western blot analysis of the soluble

cytoplasmic fraction from wt *E. coli* cells probed with anti-his antibody to detect Alg13-his. Figure 11B shows a Western blot analysis of different fractions isolated from wt and Δ*dnaJ* cells probed with anti-FLAG antibody to detect Alg14-FLAG. Samples were collected at 0, 1, 2, and 3 hours post induction (hpi) for Alg13 and at 3 hpi for Alg14. Samples were prepared by centrifugation of lysed cells at 20,000xg for 20 min and collecting the supernatant as the soluble fraction and the pellet as the insoluble fraction (insol). For Alg14, the soluble fraction was further spun at 100,000xg for 1 hr and the supernatant and pellet were collected as the soluble (sol) and membrane (mem) fractions, respectively.

**[0036]** Figure 12 is a photograph of a Western blot depicting expression of Alg1 and Alg2 in *E. coli*. These Western blots are of membrane fractions from Δ*dnaJ* cells harvested 3, 4, and 5 hpi. Blots were probed with anti-his antibodies.

**[0037]** Figure 13 is an alignment between the wild-type and codon optimized nucleotide sequences for Alg1.

**[0038]** Figure 14 is an alignment between the wild-type and codon optimized nucleotide sequences for Alg2.

**[0039]** Figure 15 is an alignment between the wild-type and codon optimized nucleotide sequences for Alg13.

**[0040]** Figure 16 is an alignment between the wild-type and codon optimized nucleotide sequences for Alg14.

**[0041]** Figure 17 is an alignment between the wild-type and codon optimized nucleotide sequences for Rft1.

**[0042]** Figure 18 is an alignment between the wild-type and codon optimized nucleotide sequences for Sttc3.

## DETAILED DESCRIPTION OF THE INVENTION

### Definitions

**[0043]** The following definitions of terms and methods are provided to better describe the present disclosure and to guide those of ordinary skill in the art in the practice of the present disclosure.

**[0044]** Unless explained otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. The materials, methods, and examples are illustrative only and not intended to be limiting. Other features of the disclosure are apparent from the following detailed description and the claims.

**[0045]** As used herein, "comprising" means "including" and the singular forms "a" or "an" or "the" include plural references unless the context clearly dictates otherwise. For example, reference to "comprising a cell" includes one or a plurality of such cells. The term "or" refers to a single element of stated alternative elements or a combination of two or more elements, unless the context clearly indicates otherwise.

**[0046]** The term "human-like" with respect to a glycoproteins refers to proteins having attached N-acetylglucosamine (GlcNAc) residue linked to the amide nitrogen of an asparagine residue (*N*-linked) in the protein, that is similar or even identical to those produced in humans.

**[0047]** "*N*-glycans" or "*N*-linked glycans" refer to *N*-linked oligosaccharide structures. The *N*-glycans can be attached to proteins or synthetic glycoprotein intermediates, which can be manipulated further *in vitro* or *in vivo*. The predominant sugars found on glycoproteins are glucose, galactose, mannose, fucose, N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), and sialic acid (e.g., N-acetyl-neuraminic acid (NeuAc)).

**[0048]** Unless otherwise indicated, and as an example for all sequences described herein under the general format "SEQ ID NO:", "nucleic acid comprising SEQ ID NO:1"

refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:1, or (ii) a sequence complementary to SEQ ID NO:1. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

**[0049]** An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., RNA, DNA, or a mixed polymer) or glycoprotein is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases and genomic sequences with which it is naturally associated. The term embraces a nucleic acid, polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

**[0050]** However, "isolated" does not necessarily require that the nucleic acid, polynucleotide or glycoprotein so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed "isolated" if a heterologous sequence is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. In this context, a heterologous sequence is a sequence that is not naturally adjacent to the endogenous nucleic acid sequence, whether or not the heterologous sequence is itself endogenous (originating from the same host cell or progeny thereof) or exogenous (originating from a different host cell or progeny thereof). By way of example, a promoter sequence can be substituted (e.g., by homologous recombination) for the native promoter of a gene in the genome of a host cell, such that this gene has an altered expression pattern. This gene would now become "isolated" because it is separated from at least some of the sequences that naturally flank it.

**[0051]** A nucleic acid is also considered "isolated" if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered "isolated" if it contains an insertion, deletion, or a point mutation introduced artificially, e.g., by human intervention. An "isolated nucleic acid" also includes a nucleic acid integrated into a host cell chromosome at a heterologous site and a nucleic acid construct present as an episome. Moreover, an "isolated nucleic acid" can be substantially free of other cellular material or substantially free of culture medium when produced by recombinant techniques or substantially free of chemical precursors or other chemicals when chemically synthesized.

**Glycosylation Engineering**

**[0052]** A first aspect of the present invention relates to a prokaryotic host cell comprising eukaryotic glycosyltransferase activity, where the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity.

**[0053]** The prokaryotic host cell of the present invention has eukaryotic dolichyl-linked UDP-GlcNAc transferase activity which may comprise Alg13 activity and Alg14 activity. The Alg13 activity and Alg14 activity is achieved with either wild-type nucleotide sequences or codon optimized sequences. As shown in Figure 10B, these enzymes serve to add GlcNAc unit to bactoprenol. The alg13 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 1:

atgggtattattgaagaaaaggctcttttgttacgtgtggggcaacggtgccatttccaaagctcgtctcatgtgtgctaagcgacg
aattctgccaagaattgattcaatatggattcgtacgtctaatcattcagtttgggagaaactacagttctgaatttgagcatttagtgc
aagaacgcgggggccaaagagaaagccaaaaaattccaattgaccagtttggctgtggcgacaccgcaagacagtatgtcctg
atgaacgggaaattaaaggtgatcgggtttgacttttcgaccaagatgcaaagtattatacgtgattattcagatttggtcatatcaca
cgctggaacgggctctatactagattctctacggttgaataaaccgttgatagtttgcgtaaacgattctttgatggataaccaccag
cagcagatagcagacaagtttgtagagttgggctacgtatggtcttgtgcacccactgaaacaggtttgatagctggtttacgtgc
atctcaaacagagaaactcaaaccattcccagtttctcataacccgtcatttgagcgattgctagttgaaactatatacagctag.

The alg13 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 2 as follows:

ATGGGTATCATCGAAGAAAAAGCTCTGTTCGTTACCTGCGGTGCTACCGTTCC
GTTCCCGAAACTGGTTTCTTGCGTTCTGTCTGACGAATTCTGCCAGGAACTGA
TCCAGTACGGTTTCGTTCGTCTGATCATCCAGTTCGGTCGTAACTACTCTTCTG
AATTCGAACACCTGGTTCAGGAACGTGGTGGTCAGCGTGAATCTCAGAAAAT
CCCGATCGACCAGTTCGGTTGCGGTGACACCGCTCGTCAGTACGTTCTGATGA
ACGGTAAACTGAAAGTTATCGGTTTCGACTTCTCTACCAAAATGCAGTCTATC
ATCCGTGACTACTCTGACCTGGTTATCTCTCACGCTGGTACCGGTTCTATCCT
GGACTCTCTGCGTCTGAACAAACCGCTGATCGTTTGCGTTAACGACTCTCTGA
TGGACAACCACCAGCAGCAGATCGCTGACAAATTCGTTGAACTGGGTTACGT
TTGGTCTTGCGCTCCGACCGAAACCGGTCTGATCGCTGGTCTGCGTGCTTCTC
AGACCGAAAAACTGAAACCGTTCCCGGTTTCTCACAACCCGTCTTTCGAACG
TCTGCTGGTTGAAACCATCTACTCTTAA

The alg14 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 3 as follows:

atgaaaacggcctacttggcgtcattggtgctcatcgtatcgacagcatatgttattaggttgatagcgattctgccttttttccacact
caagcaggtacagaaaaggatacgaaagatggagttaacctactgaaaatacgaaaatcgtcaaagaaaccgctcaagattttt
gtattcttaggatcgggaggtcatactggtgaaatgatccgtcttctagaaaattaccaggatcttttactgggtaagtcgattgtgta
cttgggttattctgatgaggcttccaggcaaagattcgcccactttataaaaaaaatttggtcattgcaaagtaaaatactatgaattca
tgaaagctagggaagttaaagcgactctcctacaaagtgtaaagaccatcattggaacgttggtacaatcttttgtgcacgtggtta
gaatcagatttgctatgtgtggttcccctcatctgtttttattgaatgggcctggaacatgctgtataatatccttttggttgaaaattatg
gaacttcttttgccccctgttgggttcctcccatatagtttatgtagaatcgctggcaaggattaatactcctagtctgaccggaaaaat
attatattgggtagtggatgaattcattgtccagtggcaagaattgagggacaattatttaccaagatccaagtggttcggcatcctt
gtttaa

The alg14 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 4 as follows:

ATGAAAACCGCTTACCTGGCTTCTCTGGTTCTGATCGTTTCTACCGCTTACGTT
ATCCGTCTGATCGCTATCCTGCCGTTCTTCCACACCCAGGCTGGTACCGAAAA
AGACACCAAAGACGGTGTTAACCTGCTGAAAATCCGTAAATCTTCTAAAAAA
CCGCTGAAAATCTTCGTTTTCCTGGGTTCTGGTGGTCACACCGGTGAAATGAT
CCGTCTGCTGGAAAACTACCAGGACCTGCTGCTGGGTAAATCTATCGTTTACC
TGGGTTACTCTGACGAAGCTTCTCGTCAGCGTTTCGCTCACTTCATCAAAAAA
TTCGGTCACTGCAAAGTTAAATACTACGAATTCATGAAAGCTCGTGAAGTTA
AAGCTACCCTGCTGCAGTCTGTTAAAACCATCATCGGTACCCTGGTTCAGTCT
TTCGTTCACGTTGTTCGTATCCGTTTCGCTATGTGCGGTTCTCCGCACCTGTTC

CTGCTGAACGGTCCGGGTACCTGCTGCATCATCTCTTTCTGGCTGAAAATCAT
GGAACTGCTGCTGCCGCTGCTGGGTTCTTCTCACATCGTTTACGTTGAATCTC
TGGCTCGTATCAACACCCCGTCTCTGACCGGTAAAATCCTGTACTGGGTTGTT
GACGAATTCATCGTTCAGTGGCAGGAACTGCGTGACAACTACCTGCCGCGTT
CTAAATGGTTCGGTATCCTGGTTTAA.

**[0054]** The prokaryotic host cell of present invention has eukaryotic mannosyltransferase activity which comprises Alg1 activity and Alg2 activity. The Alg1 activity and Alg2 activity is achieved with a wild-type nucleic acid molecule or a codon optimized nucleic acid sequence as follows. As shown in Figure 10B, these enzymes add mannose units to GlcNAc units. The alg1 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 5 as follows:

atgttttggaaattcctcggtggttacttgccttaataatattatacctttccataccgttagtggtttattatgttataccctacttgttttat
ggcaacaagtcgaccaaaaaaaggatcatcatatttgtgctgggtgatgtaggacactctccaaggatatgctatcacgctataag
tttcagtaagttaggttggcaagtcgagctatgcggttatgtggaggacactctacccaaaattatttccagtgatccaaatatcacc
gtccatcatatgtcaaacttgaaaagaaagggaggcggaacatcagttatatttatggtaaagaaggtgctttttcaagttttaagtat
tttcaaattactttgggaattgagaggaagcgattacatactagttcaaaatccaccgagcataccattcttccgattgctgtgctat
acaagttgaccggttgtaaactaattattgattggcacaatctagcatattcgatattgcaactaaaatttaaaggaaacttttaccatc
ctttagtgttgatatcttacatggtagagatgatattcagcaaatttgctgattataacttgactgttactgaagcaatgaggaaatattt
aattcaaagctttcacttgaatccaaagagatgtgctgttctctacgaccgcccggcttcccaatttcaaccttggcaggtgacattt
ctcgtcaaaaagccctaactaccaaagcctttataaagaattatattcgcgatgattttgatacagaaaaaggcgataaaattattgt
gacttcaacatcattcacccctgatgaagatattggtattttattaggtgccctaaagatttacgaaaactcttatgtcaaatttgattca
agtttgcctaagatcttgtgttttataacgggtaaaggaccactaaaggagaaatatatgaagcaagtagaagaatatgactggaa
gcgctgtcaaatcgaatttgtgtggttgtcagcagaggattacccaaagttattacaattatgcgattacggagtttccctgcatactt
caagttcagggttggacctgccaatgaaaattttagatatgtttggctcaggtcttcctgttattgcaatgaactatccagtgcttgac
gaattagtacaacacaatgtaaatgggttaaaatttgttgatagaagggagcttcatgaatctctgatttttgctatgaaagatgctga
tttataccaaaaattgaagaaaaatgtaacgcaggaagctgagaacagatggcaatcaaattgggaacgaacaatgagagattt
gaagctaattcattga.

The alg1 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 6 as follows:

ATGTTCCTGGAAATCCCGCGTTGGCTGCTGGCTCTGATCATCCTGTACCTGTC
TATCCCGCTGGTTGTTTACTACGTTATCCCGTACCTGTTCTACGGTAACAAAT
CTACCAAAAAACGTATCATCATCTTCGTTCTGGGTGACGTTGGTCACTCTCCG
CGTATCTGCTACCACGCTATCTCTTTCTCTAAACTGGGTTGGCAGGTTGAACT
GTGCGGTTACGTTGAAGACACCCTGCCGAAAATCATCTCTTCTGACCCGAAC
ATCACCGTTCACCACATGTCTAACCTGAAACGTAAAGGTGGTGGTACCTCTGT

TATCTTCATGGTTAAAAAAGTTCTGTTCCAGGTTCTGTCTATCTTCAAACTGCT
GTGGGAACTGCGTGGTTCTGACTACATCCTGGTTCAGAACCCGCCGTCTATCC
CGATCCTGCCGATCGCTGTTCTGTACAAACTGACCGGTTGCAAACTGATCATC
GACTGGCACAACCTGGCTTACTCTATCCTGCAGCTGAAATTCAAAGGTAACTT
CTACCACCCGCTGGTTCTGATCTCTTACATGGTTGAAATGATCTTCTCTAAATT
CGCTGACTACAACCTGACCGTTACCGAAGCTATGCGTAAATACCTGATCCAG
TCTTTCCACCTGAACCCGAAACGTTGCGCTGTTCTGTACGACCGTCCGGCTTC
TCAGTTCCAGCCGCTGGCTGGTGACATCTCTCGTCAGAAAGCTCTGACCACCA
AAGCTTTCATCAAAAACTACATCCGTGACGACTTCGACACCGAAAAAGGTGA
CAAAATCATCGTTACCTCTACCTCTTTCACCCCGGACGAAGACATCGGTATCC
TGCTGGGTGCTCTGAAAATCTACGAAAACTCTTACGTTAAATTCGACTCTTCT
CTGCCGAAAATCCTGTGCTTCATCACCGGTAAAGGTCCGCTGAAAGAAAAAT
ACATGAAACAGGTTGAAGAATACGACTGGAAACGTTGCCAGATCGAATTCGT
TTGGCTGTCTGCTGAAGACTACCCGAAACTGCTGCAGCTGTGCGACTACGGT
GTTTCTCTGCACACCTCTTCTTCTGGTCTGGACCTGCCGATGAAAATCCTGGA
CATGTTCGGTTCTGGTCTGCCGGTTATCGCTATGAACTACCCGGTTCTGGACG
AACTGGTTCAGCACAACGTTAACGGTCTGAAATTCGTTGACCGTCGTGAACT
GCACGAATCTCTGATCTTCGCTATGAAAGACGCTGACCTGTACCAGAAACTG
AAAAAAAACGTTACCCAGGAAGCTGAAAACCGTTGGCAGTCTAACTGGGAA
CGTACCATGCGTGACCTGAAACTGATCCACTAA.

The alg2 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 7 as follows:

atgattgaaaaggataaaagaacgattgcttttattcatccagacctaggtattgggggcgctgaaaggttagtcgtcgatgcagc
attaggtctacagcaacaaggacatagtgtaatcatctatactagtcactgtgataaatcacattgtttcgaagaagttaaaaacggc
caattaaaagtcgaagtttatggtgatttttaccgacaaacttttgggtcgtttttttattgttttcgcaacaattagacagctttatttag
ttattcaattgatcctacagaaaaaagtgaatgcgtaccaattaattatcattgatcaactgtctacatgtattccgcttctgcatatcttt
agttctgccactttgatgttttattgtcatttcccgaccaattattggctcaaagagctgggctattgaagaaaatatacagactacc
atttgacttaatagaacagttttccgtgagtgctgccgatactgttgtggtaaattcaaatttcactaagaatacgttccaccaaacgtt
caagtatttatccaatgatccagacgtcatttatccatgcgtggatttatcaacaatcgaaattgaagatattgacaagaaatttttcaa
aacagtgtttaacgaaggcgatagattttacctaagtataaatcgttttgagaaaaaaaaggatgttgcgctggctataaaggctttt
gcgttatctgaagatcaaatcaatgacaacgttaagttagttatttgcggtggttatgacgagagggttgcagaaaatgtggagtac
ttgaaggaactacagtctctggccgatgaatacgaattatcccatacaaccatatactaccaagaaataaagcgcgtctccgatttta
gagtcattcaaaaccaataatagtaaaattatatttttaacttccatttcatcatctctgaaagaattactgctcgaaagaaccgaaatg
ttattgtatacaccagcatatgagcactttggtattgttcctttagaagccatgaaattaggtaagcctgtactagcagtaaacaatgg
aggtcctttggagactatcaaatcttacgttgctggtgaaaatgaaagttctgccactgggtggctaaaacctgccgtccctattca
atgggctactgcaattgatgaaagcagaaagatcttgcagaacggttctgtgaactttgagaggaatggcccgctaagagtcaag
aaatacttttctagggaagcaatgactcagtcatttgaagaaaacgtcgagaaagtcatatggaaagaaaaaagtattatccttgg
gaaatattcggtatttcattctctaattttatttgcatatggcatttataaaaattctacccaataatccatggccctttctatttatggcca
cttttatggtattatattttaagaactacttatggggaatttactgggcatttgtattcgctctctcctacccttatgaagaaatataa.

The alg2 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 8 as follows:

ATGATCGAAAAAGACAAACGTACCATCGCTTTCATCCACCCGGACCTGGGTA
TCGGTGGTGCTGAACGTCTGGTTGTTGACGCTGCTCTGGGTCTGCAGCAGCAG
GGTCACTCTGTTATCATCTACACCTCTCACTGCGACAAATCTCACTGCTTCGA
AGAAGTTAAAAACGGTCAGCTGAAAGTTGAAGTTTACGGTGACTTCCTGCCG
ACCAACTTCCTGGGTCGTTTCTTCATCGTTTTCGCTACCATCCGTCAGCTGTAC
CTGGTTATCCAGCTGATCCTGCAGAAAAAAGTTAACGCTTACCAGCTGATCAT
CATCGACCAGCTGTCTACCTGCATCCCGCTGCTGCACATCTTCTCTTCTGCTA
CCCTGATGTTCTACTGCCACTTCCCGGACCAGCTGCTGGCTCAGCGTGCTGGT
CTGCTGAAAAAAATCTACCGTCTGCCGTTCGACCTGATCGAACAGTTCTCTGT
TTCTGCTGCTGACACCGTTGTTGTTAACTCTAACTTCACCAAAAACACCTTCC
ACCAGACCTTCAAATACCTGTCTAACGACCCGGACGTTATCTACCCGTGCGTT
GACCTGTCTACCATCGAAATCGAAGACATCGACAAAAAATTCTTCAAAACCG
TTTTCAACGAAGGTGACCGTTTCTACCTGTCTATCAACCGTTTCGAAAAAAAA
AAAGACGTTGCTCTGGCTATCAAAGCTTTCGCTCTGTCTGAAGACCAGATCAA
CGACAACGTTAAACTGGTTATCTGCGGTGGTTACGACGAACGTGTTGCTGAA
AACGTTGAATACCTGAAAGAACTGCAGTCTCTGGCTGACGAATACGAACTGT
CTCACACCACCATCTACTACCAGGAAATCAAACGTGTTTCTGACCTGGAATCT
TTCAAAACCAACAACTCTAAAATCATCTTCCTGACCTCTATCTCTTCTTCTCTG
AAAGAACTGCTGCTGGAACGTACCGAAATGCTGCTGTACACCCCGGCTTACG
AACACTTCGGTATCGTTCCGCTGGAAGCTATGAAACTGGGTAAACCGGTTCT
GGCTGTTAACAACGGTGGTCCGCTGGAAACCATCAAATCTTACGTTGCTGGT
GAAAACGAATCTTCTGCTACCGGTTGGCTGAAACCGGCTGTTCCGATCCAGT
GGGCTACCGCTATCGACGAATCTCGTAAAATCCTGCAGAACGGTTCTGTTAA
CTTCGAACGTAACGGTCCGCTGCGTGTTAAAAAATACTTCTCTCGTGAAGCTA
TGACCCAGTCTTTCGAAGAAAACGTTGAAAAAGTTATCTGGAAAGAAAAAAA
ATACTACCCGTGGGAAATCTTCGGTATCTCTTTCTCTAACTTCATCCTGCACAT
GGCTTTCATCAAAATCCTGCCGAACAACCCGTGGCCGTTCCTGTTCATGGCTA
CCTTCATGGTTCTGTACTTCAAAAACTACCTGTGGGGTATCTACTGGGCTTTC
GTTTTCGCTCTGTCTTACCCGTACGAAGAAATCTAA

**[0055]** The prokaryotic host cell of the present invention has eukaryotic flippase activity in the form of Rft1 activity. As shown in Figure 10B, Rft1 (or PglK) shifts the oligosaccharide assembly of GlcNAc units and mannose units from the cytoplasmic side of the inner membrane of the prokaryote host to the periplasm side. The Rft1 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 9:

atggcgaaaaaaaactcacaattgccctctactagtgagcagatcttggaaaggtccacaacaggagctaccttcctcatgatgg
gccaacttttcaccaaactggtaacgttcatactaaataatttgttgatcaggtttctgtcgcccagaatttcggtatcacggccttct
agaatttatacagggcacagtgttatttttagcagagatgcgattcgtctgtcgacgttgagaatctcagactccggtaatggaata
atcgatgatgacgacgaggaggagtaccaggaaactcattacaagtctaaagttttgcaaaccgcagtcaattttgcttacattcc
gttttggatcgggtttccactgtccattggtcttatcgcctggcagtacagaaacatcaacgcgtatttcatcactcttccattcttcag
gtggtcgatttttcttatctggctgagtatcatcgtggagctgttaagcgagccattcttcatcgtcaaccagtttatgttgaactatgc
cgcaaggtcaagatttgaaagcatcgcggtgactacaggatgtattgtcaatttatagttgtttatgccgttcagcaatcccgctac
ccaatgggggttgtcacatcggacattgacaaagaaggcatcgccatattggcatttgccttgggaaagttagcacattcgatcac
cctgctagcatgttactactgggactatctcaagaatttcaaaccaaagaaattgttcagtaccaggctaacgaagataaaaacgc
gtgaaaataacgaattgaagaaaggctacccaaagagcacatcttatttttccaaaacgacattttacagcacttcaaaaaagttta
ttttcaactatgttttaagcatttgttgacagagggtgataagttgattatcaattctttatgtactgtggaagaacaaggcatttacgct
ctattgtcgaactatggatcgctactaacaagattattatttgcgccgatcgaagaatctctgcggttatttttggcccgtttattatcct
cgcataaccctaaaaatttaaaactatctattgaagtcctggtgaatttaacaaggttttacatatacttatcgttaatgatcattgtattt
gggcctgccaattcatccttttattgcagttcttgattggctcgaaatggtccactacttccgttttggacactataagagtctactgct
ttacatcccatttttatcgcttaatggtatttttgaagctttttccagagtgtagccactggtgaccaaatttgaaacattcatattttat
gatggccttttctggtattttcctgctcaattcctggcttcttattgaaaaactcaaactatcaatcgaaggcttgatattgagtaacatc
attaacatggtgttgagaatattgtattgtggagttttcttgaataaatttcatagggaactgtttacagattcctcttttttcttcaattttaa
ggatttcaaaacagttattattgctggctcaacgatctgtctacttgactggtggtttattgggtacgttaaaaatttacaacaatttgtt
gttaacgtattattcgcaatgggattgttagcgttaattttggtcaaggagcgccaaaccatacaatcttttattaacaagagggcgg
ttccaattctaaagatgtataa.

The rft1 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 10 as follows:

ATGGCTAAAAAAAAACTCTCAGCTGCCGTCTACCTCTGAACAGATCCTGGAAC
GTTCTACCACCGGTGCTACCTTCCTGATGATGGGTCAGCTGTTCACCAAACTG
GTTACCTTCATCCTGAACAACCTGCTGATCCGTTTCCTGTCTCCGCGTATCTTC
GGTATCACCGCTTTCCTGGAATTCATCCAGGGTACCGTTCTGTTCTTCTCTCGT
GACGCTATCCGTCTGTCTACCCTGCGTATCTCTGACTCTGGTAACGGTATCAT
CGACGACGACGACGAAGAAGAATACCAGGAAACCCACTACAAATCTAAAGT
TCTGCAGACCGCTGTTAACTTCGCTTACATCCCGTTCTGGATCGGTTTCCCGCT
GTCTATCGGTCTGATCGCTTGGCAGTACCGTAACATCAACGCTTACTTCATCA
CCCTGCCGTTCTTCCGTTGGTCTATCTTCCTGATCTGGCTGTCTATCATCGTTG
AACTGCTGTCTGAACCGTTCTTCATCGTTAACCAGTTCATGCTGAACTACGCT
GCTCGTTCTCGTTTCGAATCTATCGCTGTTACCACCGGTTGCATCGTTAACTTC
ATCGTTGTTTACGCTGTTCAGCAGTCTCGTTACCCGATGGGTGTTGTTACCTCT
GACATCGACAAAGAAGGTATCGCTATCCTGGCTTTCGCTCTGGGTAAACTGG
CTCACTCTATCACCCTGCTGGCTTGCTACTACTGGGACTACCTGAAAAACTTC
AAACCGAAAAAAACTGTTCTCTACCCGTCTGACCAAAATCAAAACCCGTGAAA
ACAACGAACTGAAAAAAGGTTACCCGAAATCTACCTCTTACTTCTTCCAGAA

CGACATCCTGCAGCACTTCAAAAAAGTTTACTTCCAGCTGTGCTTCAAACACC
TGCTGACCGAAGGTGACAAACTGATCATCAACTCTCTGTGCACCGTTGAAGA
ACAGGGTATCTACGCTCTGCTGTCTAACTACGGTTCTCTGCTGACCCGTCTGC
TGTTCGCTCCGATCGAAGAATCTCTGCGTCTGTTCCTGGCTCGTCTGCTGTCTT
CTCACAACCCGAAAAACCTGAAACTGTCTATCGAAGTTCTGGTTAACCTGAC
CCGTTTCTACATCTACCTGTCTCTGATGATCATCGTTTTCGGTCCGGCTAACTC
TTCTTTCCTGCTGCAGTTCCTGATCGGTTCTAAATGGTCTACCACCTCTGTTCT
GGACACCATCCGTGTTTACTGCTTCTACATCCCGTTCCTGTCTCTGAACGGTA
TCTTCGAAGCTTTCTTCCAGTCTGTTGCTACCGGTGACCAGATCCTGAAACAC
TCTTACTTCATGATGGCTTTCTCTGGTATCTTCCTGCTGAACTCTTGGCTGCTG
ATCGAAAAACTGAAACTGTCTATCGAAGGTCTGATCCTGTCTAACATCATCA
ACATGGTTCTGCGTATCCTGTACTGCGGTGTTTTCCTGAACAAATTCCACCGT
GAACTGTTCACCGACTCTTCTTTCTTCTTCAACTTCAAAGACTTCAAAACCGTT
ATCATCGCTGGTTCTACCATCTGCCTGCTGGACTGGTGGTTCATCGGTTACGT
TAAAAACCTGCAGCAGTTCGTTGTTAACGTTCTGTTCGCTATGGGTCTGCTGG
CTCTGATCCTGGTTAAAGAACGTCAGACCATCCAGTCTTTCATCAACAAACGT
GCTGTTTCTAACTCTAAAGACGTTTAA.

**[0056]** The prokaryotic host cell of the present invention has eukaryotic oligosaccharyl transferase activity in the form of STT3 activity. As shown in Figure 10B, the STT3 enzyme (or the PlgB enzyme) transports the oligosaccharide assembly from the inner membrane to an acceptor protein which is transported to the outer membrane of the host cell. The STT3 wild-type nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 11:

atgggatccgaccggtcgtgtgttttgtctgtgtttcagaccatcctcaagctcgtcatcttcgtggcgatttttggggctgccatatc
atcacgtttgtttgcagtcatcaaatttgagtctattatccatgaattcgacccctggttcaattatagggctaccaaatatctcgtcaa
caattcgttttacaagtttttgaactggtttgacgaccgtacctggtaccccctcggaagggttactggagggactttatatcctggtt
tgatgacgactagtgcgttcatctggcacgccctgcgcaactggttgggcttgcccattgacatcagaaacgtttgtgtgctatttg
cgccactatttctggggtcaccgcctgggcgacttacgaatttacgaaagagattaaagatgccagcgctgggcttttggctgct
ggttttatagccattgtccccggttatatatctagatcagtggcggggtcctacgataatgaggccattgccattacactattaatggt
cactttcatgttttggattaaggcccaaaagactggctctatcatgcacgcaacgtgtgcagctttattctacttctacatggtgtcgg
cttggggtggatacgtgttcatcaccaacttgatcccactccatgtcttttgctgattttgatgggcagatattcgtccaaactgtattc
tgcctacaccacttggtacgctattggaactgttgcatccatgcagatcccatttgtcggtttcctacctatcaggtctaacgaccac
atggccgcattgggtgttttcggtttgattcagattgtcgccttcggtgacttcgtgaagggccaaatcagcacagctaagtttaaag
tcatcatgatggtttctctgttttgatcttggtccttggtgtggtcggacttctgccttgacctatatggggttgattgccccttggact
ggtagattttattcgttatgggataccaactacgcaaagatccacattcctatcattgcctccgtttccgaacatcaacccgtttcgtg
gcccgctttcttctttgataccccacttttgatctggctattcccccgccggtgtattcctactattcctcgacttgaaagacgagcacgtt
tttgtcatcgcttactccgttctgtgttcgtactttgccggtgttatggttagattgatgttgactttgacaccagtcatctgtgtgtccgc
cgccgtcgcattgtccaagatatttgacatctacctggatttcaagacaagtgaccgcaaatacgccatcaaacctgcggcactac

tggccaaattgattgtttccggatcattcatcttttatttgtatcttttcgtcttccattctacttgggtaacaagaactgcatactcttctcc
ttctgttgttttgccatcacaaaccccagatggtaaattggcgttgatcgacgacttcagggaagcgtactattggttaagaatgaac
tctgatgaggacagtaaggttgcagcgtggtgggattacggttaccaaattggtggcatggcagacagaaccactttagtcgata
acaacacgtggaacaatactcacatcgccatcgttggtaaagccatggcttcccctgaagagaaatcttacgaaattctaaaaga
gcatgatgtcgattatgtcttggtcatctttggtggtctaattgggtttggtggtgatgacatcaacaaattcttgtggatgatcagaatt
agcgagggaatctggccagaagagataaaagagcgtgatttctataccgcagagggagaatacagagtagatgcaagggctt
ctgagaccatgaggaactcgctactttacaagatgtcctacaaagatttcccacaattattcaatggtggccaagccactgacaga
gtgcgtcaacaaatgatcacaccattagacgtcccaccattagactacttcgacgaagtttttacttccgaaaactggatggttaga
atatatcaattgaagaaggatgatgcccaaggtagaactttgagggacgttggtgagttaaccaggtcttctacgaaaaccagaa
ggtccataaagagacctgaattaggcttgagagtctaa

The STT3 codon optimized nucleic acid molecule has the nucleotide sequence of SEQ ID NO: 12 as follows:

ATGGGTTCTGACCGTTCTTGCGTTCTGTCTGTTTTCCAGACCATCCTGAAACTG
GTTATCTTCGTTGCTATCTTCGGTGCTGCTATCTCTTCTCGTCTGTTCGCTGTT
ATCAAATTCGAATCTATCATCCACGAATTCGACCCGTGGTTCAACTACCGTGC
TACCAAATACCTGGTTAACAACTCTTTCTACAAATTCCTGAACTGGTTCGACG
ACCGTACCTGGTACCCGCTGGGTCGTGTTACCGGTGGTACCCTGTACCCGGGT
CTGATGACCACCTCTGCTTTCATCTGGCACGCTCTGCGTAACTGGCTGGGTCT
GCCGATCGACATCCGTAACGTTTGCGTTCTGTTCGCTCCGCTGTTCTCTGGTGT
TACCGCTTGGGCTACCTACGAATTCACCAAAGAAATCAAAGACGCTTCTGCT
GGTCTGCTGGCTGCTGGTTTCATCGCTATCGTTCCGGGTTACATCTCTCGTTCT
GTTGCTGGTTCTTACGACAACGAAGCTATCGCTATCACCCTGCTGATGGTTAC
CTTCATGTTCTGGATCAAAGCTCAGAAAACCGGTTCTATCATGCACGCTACCT
GCGCTGCTCTGTTCTACTTCTACATGGTTTCTGCTTGGGGTGGTTACGTTTTCA
TCACCAACCTGATCCCGCTGCACGTTTTCCTGCTGATCCTGATGGGTCGTTAC
TCTTCTAAACTGTACTCTGCTTACACCACCTGGTACGCTATCGGTACCGTTGC
TTCTATGCAGATCCCGTTCGTTGGTTTCCTGCCGATCCGTTCTAACGACCACA
TGGCTGCTCTGGGTGTTTTCGGTCTGATCCAGATCGTTGCTTTCGGTGACTTCG
TTAAAGGTCAGATCTCTACCGCTAAATTCAAAGTTATCATGATGGTTTCTCTG
TTCCTGATCCTGGTTCTGGGTGTTGTTGGTCTGTCTGCTCTGACCTACATGGGT
CTGATCGCTCCGTGGACCGGTCGTTTCTACTCTCTGTGGGACACCAACTACGC
TAAAATCCACATCCCGATCATCGCTTCTGTTTCTGAACACCAGCCGGTTTCTT
GGCCGGCTTTCTTCTTCGACACCCACTTCCTGATCTGGCTGTTCCCGGCTGGT
GTTTTCCTGCTGTTCCTGGACCTGAAAGACGAACACGTTTTCGTTATCGCTTA
CTCTGTTCTGTGCTCTTACTTCGCTGGTGTTATGGTTCGTCTGATGCTGACCCT
GACCCCGGTTATCTGCGTTTCTGCTGCTGTTGCTCTGTCTAAAATCTTCGACAT
CTACCTGGACTTCAAAACCTCTGACCGTAAATACGCTATCAAACCGGCTGCTC
TGCTGGCTAAACTGATCGTTTCTGGTTCTTTCATCTTCTACCTGTACCTGTTCG
TTTTCCACTCTACCTGGGTTACCCGTACCGCTTACTCTTCTCCGTCTGTTGTTC
TGCCGTCTCAGACCCCGGACGGTAAACTGGCTCTGATCGACGACTTCCGTGA

AGCTTACTACTGGCTGCGTATGAACTCTGACGAAGACTCTAAAGTTGCTGCTT
GGTGGGACTACGGTTACCAGATCGGTGGTATGGCTGACCGTACCACCCTGGT
TGACAACAACACCTGGAACAACACCCACATCGCTATCGTTGGTAAAGCTATG
GCTTCTCCGGAAGAAAAATCTTACGAAATCCTGAAAGAACACGACGTTGACT
ACGTTCTGGTTATCTTCGGTGGTCTGATCGGTTTCGGTGGTGACGACATCAAC
AAATTCCTGTGGATGATCCGTATCTCTGAAGGTATCTGGCCGGAAGAAATCA
AAGAACGTGACTTCTACACCGCTGAAGGTGAATACCGTGTTGACGCTCGTGC
TTCTGAAACCATGCGTAACTCTCTGCTGTACAAAATGTCTTACAAAGACTTCC
CGCAGCTGTTCAACGGTGGTCAGGCTACCGACCGTGTTCGTCAGCAGATGAT
CACCCCGCTGGACGTTCCGCCGCTGGACTACTTCGACGAAGTTTTCACCTCTG
AAAACTGGATGGTTCGTATCTACCAGCTGAAAAAAGACGACGCTCAGGGTCG
TACCCTGCGTGACGTTGGTGAACTGACCCGTTCTTCTACCAAAACCCGTCGTT
CTATCAAACGTCCGGAACTGGGTCTGCGTGTTTAA.

**[0057]** The successful expression of eukaryotic proteins, especially membrane proteins, in *E. coli* and other bacteria is a nontrivial task (Baneyx et al., "Recombinant Protein Folding and Misfolding in Escherichia coli," *Nat Biotechnol* 22:1399-1408 ((2004), which is hereby incorporated by reference in its entirety). Thus, consideration has to be given to numerous issues in order to achieve high expression yields of correctly folded and correctly localized proteins (e.g., insertion into the inner membrane). All of these factors collectively dictate whether the eukaryotic proteins will be functional when expressed inside *E. coli* cells.

**[0058]** In one embodiment of the present invention, eukaryotic glycosyltransferases can be codon optimized to overcome limitations associated with the codon usage bias between *E. coli* (and other bacteria) and higher organisms, such as yeast and mammalian cells. Codon usage bias refers to differences among organisms in the frequency of occurrence of codons in protein-coding DNA sequences (genes). A codon is a series of three nucleotides (triplets) that encodes a specific amino acid residue in a polypeptide chain. Codon optimization can be achieved by making specific transversion nucleotide changes, i.e. a purine to pyrimidine or pyrimidine to purine nucleotide change, or transition nucleotide change, i.e. a purine to purine or pyrimidine to pyrimidine nucleotide change. Exemplary codon optimized nucleic acid molecules corresponding to wild-type eukaryotic dolichyl-linked UDP-GlcNAc transferase (SEQ ID NOs: 1 and 3), eukaryotic mannosyltransferase (SEQ ID NOs: 5 and 7), eukaryotic flippase (SEQ ID

NO: 9), and eukaryotic oligosaccharyl transferase (SEQ ID NO: 11) are set forth above as SEQ ID NOs: 2, 4, 6, 8, 10, and 12, respectively.

**[0059]** Figures 13 thru 18 are sequence alignments showing specific nucleotides in the wildtype sequences of Alg1, Alg2, Alg3, Alg4, Rft1, and Sttc3, respectively, subject to transversion and transition changes to achieve codon optimized nucleotide sequences. An exemplary optimized sequence is shown in the sequence alignment and identified as "optimized" and the wildtype sequence is identified as "query". The location of nucleotide changes in the wildtype sequences are shown using the following convention: "|" indicates an unchanged nucleotide (i.e. the nucleotide of the wildtype sequence is not changed in the optimized sequence); "*" indicates the location of a transversion change (e.g. adenine "A" changed to a cytosine "C" or thymine "T"; guanine "G" changed to C or T; C changed to A or G; and T changed to A or G); and "#" indicates the location of a transition change (e.g. A to G or G to A; C to T or T to C). Although an exemplary optimized sequence is shown in each of Figures 13 thru 18, one of skill in the art will readily appreciate that not all of the identified nucleotide changes must be made to achieve a codon optimized sequence and that in the case of a transversion change, two nucleotide changes are possible at each location (i.e. a purine can be changed to either pyrimidine (C or T) and a pyrimidine can be changed to either purine (A or G)).

**[0060]** The nucleic acid molecules and homologs, variants and derivatives of the *alg* genes comprising sequences have at least 75% identity to SEQ ID NO:6, 77% identity to SEQ ID NO:8, 77% identity to SEQ ID NO:2, and 77% identity to SEQ ID NO:4.

**[0061]** In another embodiment, the nucleic acid molecule of the present invention encodes a polypeptide encoded by the polynucleotides of SEQ ID NO:2, 4, 6, 8. Preferably, the nucleic acid molecule encodes a polypeptide sequence of at least 75%, 77%, 80%, 85%, 90%, or 95% identical to SEQ ID NO:2, 4, 6, 8, with the identity values, rising to 80%, 85%, 90%, 95%, 98%, 99%, 99.9%, or even higher.

**[0062]** In further embodiments, the nucleic acid molecules, homologs, variants, and derivatives of the flippase genes have a nucleotide sequence at least 76% identity to

SEQ ID NO:10. Further, the nucleic acid molecule of the present invention encodes a polypeptide encoded by the polynucleotides of SEQ ID NO: 10. Preferably, the nucleic acid molecule encodes a polypeptide sequence of at least 76%, 80%, 85%, 90% or 95% identical to SEQ ID NO: 10, with the identity values increasing to 98%, 99%, 99.9%, or even higher.

**[0063]** In various other embodiments, the nucleic acid molecule and homologs, variants and derivatives of the OST genes have at least 79% identity to SEQ ID NO:12. In another embodiment, the nucleic acid molecule of the present invention encodes a polypeptide encoded by the polynucleotides of SEQ ID NO: 12. Preferably, the nucleic acid molecule encodes a polypeptide sequence of at least 79%, 80%, 85%, 90%, or 95% identical to SEQ ID NO: 12, with the identity values increasing to 98%, 99%, 99.9%, or even higher.

**[0064]** The present invention also encompasses nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. As defined above, and as is well known in the art, stringent hybridizations are performed at about 25°C below the thermal melting point ($T_m$) for the specific DNA hybrid under a particular set of conditions, where the $T_m$ is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. Stringent washing can be performed at temperatures about 5°C lower than the $T_m$ for the specific DNA hybrid under a particular set of conditions.

**[0065]** The polynucleotides or nucleic acid molecules of the present invention refer to the polymeric form of nucleotides of at least 10 bases in length. These include DNA molecules (e.g., linear, circular, cDNA, chromosomal, genomic, or synthetic, double stranded, single stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hair-pinned, circular, or in a padlocked conformation) and RNA molecules (e.g., tRNA, rRNA, mRNA, genomic, or synthetic) and analogs of the DNA or RNA molecules of the described as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native inter-nucleoside bonds, or both. The isolated nucleic acid molecule of the invention includes a nucleic acid molecule free of naturally flanking sequences (i.e., sequences located at the 5' and 3' ends of the nucleic acid molecule) in

the chromosomal DNA of the organism from which the nucleic acid is derived. In various embodiments, an isolated nucleic acid molecule can contain less than about 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, 50 bp, 25 bp or 10 bp of naturally flanking nucleotide chromosomal DNA sequences of the microorganism from which the nucleic acid molecule is derived.

**[0066]** The heterologous nucleic acid molecule is inserted into the expression system or vector in proper sense (5'→3') orientation relative to the promoter and any other 5' regulatory molecules, and correct reading frame. The preparation of the nucleic acid constructs can be carried out using standard cloning methods well known in the art, as described by Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Springs Laboratory Press, Cold Springs Harbor, New York (1989), which is hereby incorporated by reference in its entirety. U.S. Patent No. 4,237,224 to Cohen and Boyer, which is hereby incorporated by reference in its entirety, also describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase.

**[0067]** Suitable expression vectors include those which contain replicon and control sequences that are derived from species compatible with the host cell. For example, if *E. coli* is used as a host cell, plasmids such as pUC19, pUC18, or pBR322 may be used. Other suitable expression vectors are described in *Molecular Cloning: a Laboratory Manual:* 3rd edition, Sambrook and Russell, 2001, Cold Spring Harbor Laboratory Press, which is hereby incorporated by reference in its entirety. Many known techniques and protocols for manipulation of nucleic acids, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Ausubel et al. eds., (1992), which is hereby incorporated by reference in its entirety.

**[0068]** Different genetic signals and processing events control many levels of gene expression (e.g., DNA transcription and messenger RNA ("mRNA") translation) and subsequently the amount of fusion protein that is displayed on the ribosome surface. Transcription of DNA is dependent upon the presence of a promoter, which is a DNA

sequence that directs the binding of RNA polymerase, and thereby promotes mRNA synthesis. Promoters vary in their "strength" (i.e., their ability to promote transcription). For the purposes of expressing a cloned gene, it is desirable to use strong promoters to obtain a high level of transcription and, hence, expression and surface display.

Therefore, depending upon the host system utilized, any one of a number of suitable promoters may also be incorporated into the expression vector carrying the deoxyribonucleic acid molecule encoding the protein of interest coupled to a stall sequence. For instance, when using *E. coli*, its bacteriophages, or plasmids, promoters such as the T7 phage promoter, *lac* promoter, *trp* promoter, *rec*A promoter, ribosomal RNA promoter, the $P_R$ and $P_L$ promoters of coliphage lambda and others, including but not limited, to *lac*UV5, *omp*F, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lac*UV5 (*tac*) promoter or other *E. coli* promoters produced by recombinant DNA or other synthetic DNA techniques may be used to provide for transcription of the inserted gene.

**[0069]** Translation of mRNA in prokaryotes depends upon the presence of the proper prokaryotic signals, which differ from those of eukaryotes. Efficient translation of mRNA in prokaryotes requires a ribosome binding site called the Shine-Dalgarno ("SD") sequence on the mRNA. This sequence is a short nucleotide sequence of mRNA that is located before the start codon, usually AUG, which encodes the amino-terminal methionine of the protein. The SD sequences are complementary to the 3'-end of the 16S rRNA (ribosomal RNA) and probably promote binding of mRNA to ribosomes by duplexing with the rRNA to allow correct positioning of the ribosome. For a review on maximizing gene expression, see Roberts and Lauer, *Methods in Enzymology*, 68:473 (1979), which is hereby incorporated by reference in its entirety.

**[0070]** In accordance with the present invention, the host cell is a prokaryote. Such cells serve as a host for expression of recombinant proteins for production of recombinant therapeutic proteins of interest. Exemplary host cells include *E. coli* and other Enterobacteriaceae, *Escherichia sp., Campylobacter sp., Wolinella sp., Desulfovibrio sp. Vibrio sp., Pseudomonas sp. Bacillus sp., Listeria sp., Staphylococcus sp., Streptococcus sp., Peptostreptococcus sp., Megasphaera sp., Pectinatus sp.,*

*Selenomonas sp., Zymophilus sp., Actinomyces sp., Arthrobacter sp., Frankia sp., Micromonospora sp., Nocardia sp., Propionibacterium sp., Streptomyces sp., Lactobacillus sp., Lactococcus sp., Leuconostoc sp., Pediococcus sp., Acetobacterium sp., Eubacterium sp., Heliobacterium sp., Heliospirillum sp., Sporomusa sp., Spiroplasma sp., Ureaplasma sp., Erysipelothrix, sp., Corynebacterium sp. Enterococcus sp., Clostridium sp., Mycoplasma sp., Mycobacterium sp., Actinobacteria sp., Salmonella sp., Shigella sp., Moraxella sp., Helicobacter sp, Stenotrophomonas sp., Micrococcus sp., Neisseria sp., Bdellovibrio sp., Hemophilus sp., Klebsiella sp., Proteus mirabilis, Enterobacter cloacae, Serratia sp., Citrobacter sp., Proteus sp., Serratia sp., Yersinia sp., Acinetobacter sp., Actinobacillus sp. Bordetella sp., Brucella sp., Capnocytophaga sp., Cardiobacterium sp., Eikenella sp., Francisella sp., Haemophilus sp., Kingella sp., Pasteurella sp., Flavobacterium sp. Xanthomonas sp., Burkholderia sp., Aeromonas sp., Plesiomonas sp., Legionella sp.* and alpha-proteobacteria such as *Wolbachia sp.*, cyanobacteria, spirochaetes, green sulfur and green non-sulfur bacteria, Gram-negative cocci, Gram negative bacilli which are fastidious, Enterobacteriaceae -glucose-fermenting gram-negative bacilli, Gram negative bacilli - non-glucose fermenters, Gram negative bacilli - glucose fermenting, oxidase positive.

**[0071]** In one embodiment of the present invention, the *E. coli* host strain C41(DE3) is used, because this strain has been previously optimized for general membrane protein overexpression (Miroux et al., "Over-production of Proteins in Escherichia coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels," *J Mol Biol* 260:289-298 (1996), which is hereby incorporated by reference in its entirety). Further optimization of the host strain includes deletion of the gene encoding the DnaJ protein (e.g., Δ*dnaJ* cells). The reason for this deletion is that inactivation of *dnaJ* is known to increase the accumulation of overexpressed membrane proteins and to suppress the severe cytotoxicity commonly associated with membrane protein overexpression (Skretas et al., "Genetic Analysis of G Protein-coupled Receptor Expression in Escherichia coli: Inhibitory Role of DnaJ on the Membrane Integration of the Human Central Cannabinoid Receptor," *Biotechnol Bioeng* (2008), which is hereby incorporated by reference in its entirety). Applicants have

observed this following expression of Alg1 and Alg2. Furthermore, deletion of competing sugar biosynthesis reactions is required to ensure optimal levels of *N*-glycan biosynthesis. For instance, the deletion of genes in the *E. coli* O16 antigen biosynthesis pathway (Feldman et al., "The Activity of a Putative Polyisoprenol-linked Sugar Translocase (Wzx) Involved in Escherichia coli O Antigen Assembly is Independent of the Chemical Structure of the O Repeat," *J Biol Chem* 274:35129-35138 (1999), which is hereby incorporated by reference in its entirety) will ensure that the bactoprenol-GlcNAc-PP substrate is available for desired mammalian *N*-glycan reactions. To eliminate unwanted side reactions, the following are representative genes that are deleted from the *E. coli* host strain: *wbbL*, *glcT*, *glf*, *gafT*, *wzx*, *wzy*, *waaL*.

**[0072]** Methods for transforming/transfecting host cells with expression vectors are well-known in the art and depend on the host system selected, as described in Sambrook et al., Molecular Cloning: *A Laboratory Manual*, Cold Springs Laboratory Press, Cold Springs Harbor, New York (1989), which is hereby incorporated by reference in its entirety. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation, and transfection using bacteriophage.

**[0073]** One aspect of the present invention is directed to a glycoprotein conjugate comprising a protein and at least one peptide comprising a D-$X_1$-N-$X_2$-T motif fused to the protein, wherein D is aspartic acid, $X_1$ and $X_2$ are any amino acid other than proline, N is asparagine, and T is threonine.

**[0074]** Another aspect of the present invention is directed to a method of producing a glycosylated protein. This method comprises providing a prokaryotic host cell comprising eukaryotic glycosyltransferase activity, where the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity. The prokaryotic host cell is then cultured under conditions effective to produce a glycosylated protein.

**[0075]** The method of the present invention can be used to produce a glycosylated antibody in accordance with the present invention.

**[0076]** Accordingly, in various aspects, the present invention provides a prokaryotic protein expression system that is engineered to "humanize" *N*-linked proteins as a platform for the stereospecific biosynthesis of a vast array of *N*-linked glycoproteins. In certain embodiments, reconstitution of a eukaryotic *N*-glycosylation pathway in *E. coli* using metabolic pathway and protein engineering techniques results in *N*-glycoproteins with structurally homogeneous human-like glycans. Since native glycosylation pathways are absent in the majority of bacteria, it is contemplated that glyco-engineered bacteria is capable of stereospecific production of *N*-linked glycoproteins with homogenous glycoform synthesized per cell. This ensures that each glyco-engineered cell line will correspond to a unique carbohydrate signature. It is, therefore, an object of the invention to engineer bacteria to produce human-like glycosylation.

**[0077]** The oligosaccharide chain attached by the prokaryotic glycosylation machinery is structurally distinct from that attached by higher eukaryotic and human glycosylation pathways (Weerapana et al., "Asparagine-linked Protein Glycosylation: From Eukaryotic to Prokaryotic Systems," *Glycobiology* 16:91R-101R (2006), which is hereby incorporated by reference in its entirety). In certain embodiments, to begin "humanizing" the bacterial glycosylation machinery (Figure 10A), an object of the present invention is to generate the $Man_3GlcNAc_2$ oligosaccharide structure. In a first aspect, a recombinant pathway comprising the biosynthesis of lipid-linked $Man_3GlcNAc_2$ is constructed in *E. coli* (Figure 10B). The first part of this pathway is the enzymatic synthesis of lipid-linked $Man_3GlcNAc_2$. Specifically, one of several eukaryotic glycosyltransferases is functionally expressed in *E. coli* and the resulting lipid-linked oligosaccharides is analyzed by metabolic labeling of cells with $^3$H-GlcNAc and $^3$H-mannose or with fluorescent lectins (e.g., AlexaFluor-ConA). The $Man_3GlcNAc_2$ oligosaccharide structure represents the core structure of most of the *N*-glycans found in eukaryotic cells. The glycosyltransferases required for the assembly of this structure are known in eukaryotes and most of these enzymes have been functionally expressed in *E. coli*, however to date, no one has been successful in achieving this oligosaccharide

structure. In addition, the substrates of these glycosyltransferases, namely UDP-GlcNAc and GDP-Man, are both present in the cytoplasm of *E. coli*.

**Site-Specific Transfer of $Man_3GlcNAc_2$ Core onto Target Proteins.**

**[0078]** An additional part of the pathway to produce human-like oligosaccharide structures in prokaryotes entails the transfer of the $Man_3GlcNAc_2$ oligosaccharide to N-X-S/T sites on polypeptide chains. This requires functional expression of an integral membrane protein or protein complex known as an oligosaccharyltransferase (OST) that is responsible for the transfer of oligosaccharides to the target protein. Various prokaryotic and eukaryotic OSTs have the ability of to transfer the lipid-linked $Man_3GlcNAc_2$ oligosaccharide onto the target protein. Accordingly, the prokaryotic protein expression system comprises at least one OST activity.

**[0079]** In various aspects, reconstituting a eukaryotic glycosylation pathway in *E. coli* requires the activity of a flippase and an OST (PglK and PglB in *C. jejuni*, respectively, and Rft1 and STT3 in yeast, respectively) (see Figure 10B). The PglK flippase is responsible for translocating the lipid-linked *C. jejuni* heptasaccharide across the inner membrane. Fortuitously, PglK exhibits relaxed specificity towards the glycan structure of the lipid-linked oligosaccharide intermediate (Alaimo et al., "Two Distinct But Interchangeable Mechanisms for Flipping of Lipid-linked Oligosaccharides," *Embo J* 25:967-76 (2006) and Wacker 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* 103:7088-93 (2006), which are hereby incorporated by reference in their entirety). Accordingly, it is contemplated that this enzyme will recognize lipid-linked $Man_3GlcNAc_2$ and thus no further engineering is required. Alternatively, in the unlikely event that PglK does not recognize lipid-linked $Man_3GlcNAc_2$, the present invention provides for expression of a eukaryotic flippase such as, among others Rft1.

**[0080]** "Target proteins", "proteins of interest", or "therapeutic proteins" include without limitation erythropoietin, cytokines such as interferons, G-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, soluble IgE receptor α-chain,

IgG, IgG fragments, IgM, interleukins, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α-1 antitrypsin, DNase II, α-feto proteins, AAT, rhTBP-1 (aka TNF binding protein 1), TACI-Ig (transmembrane activator and calcium modulator and cyclophilin ligand interactor), FSH (follicle stimulating hormone), GM-CSF, GLP-1 w/ and w/o FC (glucagon like protein 1) IL-I receptor agonist, sTNFr (enbrel, aka soluble TNF receptor Fc fusion) ATIII, rhThrombin, glucocerebrosidase, CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4-Ig), receptors, hormones, human vaccines, animal vaccines, peptides, and serum albumin.

**Aglycosylated vs. Glycosylated IgGs**

**[0081]** Another aspect of the present invention relates to a glycosylated antibody comprising an Fv portion which recognizes and binds to a native antigen and an Fc portion which is glycosylated at a conserved asparagine residue.

**[0082]** The glycosylated antibody of the present invention can be in the form of a monoclonal or polyclonal antibody.

**[0083]** A single immunoglobulin molecule is comprised of two identical light (L) chains and two identical heavy (H) chains. Light chains are composed of one constant domain ($C_L$) and one variable domain ($V_L$) while heavy chains are consist of three constant domains ($C_H1$, $C_H2$ and $C_H3$) and one variable domain ($V_H$). Together, the $V_H$ and $V_L$ domains compose the antigen-binding portion of the molecule known as the Fv. The Fc portion is glycosylated at a conserved Asn297 residue (Figure 5A indicated by asterisks). Attachment of *N*-glycan at this position results in an "open" conformation that is essential for effector interaction.

**[0084]** Monoclonal antibodies can be made using recombinant DNA methods, as described in U.S. Patent No. 4,816,567 to Cabilly et al. and Anderson et al., "Production Technologies for Monoclonal Antibodies and their Fragments," *Curr Opin Biotechnol.* 15:456-62 (2004), which are hereby incorporated by reference in their entirety. The polynucleotides encoding a monoclonal antibody are isolated, such as from mature B-

cells or hybridoma cell, such as by RT-PCR using oligonucleotide primers that specifically amplify the genes encoding the heavy and light chains of the antibody, and their sequence is determined using conventional procedures. The isolated polynucleotides encoding the heavy and light chains are then cloned into suitable expression vectors, which are then transfected into the host cells of the present invention, and monoclonal antibodies are generated. In one embodiment, recombinant DNA techniques are used to modify the heavy and light chains with N-terminal export signal peptides (e.g., PelB signal peptide) to direct the heavy and light chain polypeptides to the bacterial periplasm. Also, the heavy and light chains can be expressed from either a bicistronic construct (e.g., a single mRNA that is translated to yield the two polypeptides) or, alternatively, from a two cistron system (e.g., two separate mRNAs are produced for each of the heavy and light chains). To achieve high-level expression and efficient assembly of full-length IgGs in the bacterial periplasm, both the bicistronic and two cistron constructs can be manipulated to achieve a favorable expression ratio. For example, translation levels can be raised or lowered using a series of translation initiation regions (TIRs) inserted just upstream of the bicistronic and two cistron constructs in the expression vector (Simmons et al., "Translational Level is a Critical Factor for the Secretion of Heterologous Proteins in Escherichia coli," *Nat Biotechnol* 14:629-34 (1996), which is hereby incorporated by reference in its entirety). When this antibody producing plasmid is introduced into a bacterial host that also harbors plasmid- or genome-encoded genes for expressing glycosylation enzymes, the resulting antibodies are glycosylated in the periplasm. Recombinant monoclonal antibodies or fragments thereof of the desired species can also be isolated from phage display libraries as described (McCafferty et al., "Phage Antibodies: Filamentous Phage Displaying Antibody Variable Domains," *Nature* 348:552-554 (1990); Clackson et al., "Making Antibody Fragments using Phage Display Libraries," *Nature* 352:624-628 (1991); and Marks et al., "By-Passing Immunization. Human Antibodies from V-Gene Libraries Displayed on Phage," *J. Mol. Biol.* 222:581-597 (1991), which are hereby incorporated by reference in their entirety).

**[0085]** The polynucleotide(s) encoding a monoclonal antibody can further be modified in a number of different ways using recombinant DNA technology to generate

alternative antibodies. In one embodiment, the constant domains of the light and heavy chains of, for example, a mouse monoclonal antibody can be substituted for those regions of a human antibody to generate a chimeric antibody. Alternatively, the constant domains of the light and heavy chains of a mouse monoclonal antibody can be substituted for a non-immunoglobulin polypeptide to generate a fusion antibody. In other embodiments, the constant regions are truncated or removed to generate the desired antibody fragment of a monoclonal antibody. Furthermore, site-directed or high-density mutagenesis of the variable region can be used to optimize specificity and affinity of a monoclonal antibody.

**[0086]** In some embodiments, the antibody of the present invention is a humanized antibody. Humanized antibodies are antibodies that contain minimal sequences from non-human (e.g. murine) antibodies within the variable regions. Such antibodies are used therapeutically to reduce antigenicity and human anti-mouse antibody responses when administered to a human subject. In practice, humanized antibodies are typically human antibodies with minimal to no non-human sequences. A human antibody is an antibody produced by a human or an antibody having an amino acid sequence corresponding to an antibody produced by a human.

**[0087]** Humanized antibodies can be produced using various techniques known in the art. An antibody can be humanized by substituting the complementarity determining region (CDR) of a human antibody with that of a non-human antibody (e.g. mouse, rat, rabbit, hamster, etc.) having the desired specificity, affinity, and capability (Jones et al., "Replacing the Complementarity-Determining Regions in a Human Antibody With Those From a Mouse," *Nature* 321:522-525 (1986); Riechmann et al., "Reshaping Human Antibodies for Therapy," *Nature* 332:323-327 (1988); Verhoeyen et al., "Reshaping Human Antibodies: Grafting an Antilysozyme Activity," *Science* 239:1534-1536 (1988), which are hereby incorporated by reference in their entirety). The humanized antibody can be further modified by the substitution of additional residues either in the Fv framework region and/or within the replaced non-human residues to refine and optimize antibody specificity, affinity, and/or capability.

**[0088]** Bispecific antibodies are also suitable for use in the methods of the present invention. Bispecific antibodies are antibodies that are capable of specifically recognizing and binding at least two different epitopes. Bispecific antibodies can be intact antibodies or antibody fragments. Techniques for making bispecific antibodies are common in the art (Traunecker et al., "Bispecific Single Chain Molecules (Janusins) Target Cytotoxic Lymphocytes on HIV Infected Cells," *EMBO J.* 10:3655-3659 (1991) and Gruber et al., "Efficient Tumor Cell Lysis Mediated by a Bispecific Single Chain Antibody Expressed in Escherichia coli," *J. Immunol.* 152:5368-74 (1994), which are hereby incorporated by reference in their entirety).

**Glycan Screening Technologies**

**[0089]** A further aspect of the present invention pertains to a method for screening bacteria or bacteriophages. This method involves expressing one or more glycans on the surface of a bacteria and attaching a label on the one or more glycans on the surface of the bacteria or on the surface of a bacteriophage derived from the bacteria. The most common bacteriophages used in phage display are M13 and fd filamentous phage, though T4, T7, and λ phage are also used. The label is then analyzed in a high-throughput format.

**[0090]** When a bacteriophage is subjected to labeling and analyzing, the method of the present invention further comprises infecting the bacteria expressing one or more glycans on the cell surface with a helper phage under conditions effective to produce a bacteriophage with one or more glycans on its surface. The bacteriophage is then enriched with one or more glycans on its surface. Alternatively, the use of the helper phage can be eliminated by using a novel 'bacterial packaging cell line' technology (Chasteen et al., "Eliminating Helper Phage From Phage Display," *Nucleic Acids Res* 34:e145 (2006), which is hereby incorporated by reference in its entirety).

**[0091]** The labeling can be carried out with a lectin which recognizes a glycan on the surface of the bacteria or bacteriophage and has a detectable label. Alternatively, the labeling step is carried out with an antibody which recognizes a glycan on the surface of the bacteria or bacteriophage and has a detectable label. Alternatively, by immobilizing a

relevant protein target(s) (e.g., lectin, antibodies) to a solid support such as the surface of a 96-well plate, a cell or phage that displays a protein that binds to one of those targets on its surface will remain while others are removed by washing. Those that remain can be eluted, used to produce more cells (by culturing cells) or phage (by bacterial infection with helper phage) and so produce a cell or phage mixture that is enriched with relevant (i.e. binding) cell or phage. The repeated cycling of these steps is referred to as 'panning'.

**[0092]** This aspect of the present invention permits screening by cell surface display and glycophage display of glycoproteins where engineered bacterial cell lines produce diverse glycans and glycoproteins in a rapid and cost-effective manner. These assays allows for quantitative, high-throughput glycan analysis and rapid isolation of mutants that confer desired phenotypes. The underlying premise for these assays is that both cell surface display and phage display create a unique genotype (i.e., DNA) to phenotype (i.e., protein activity or modification such as glycosylation) linkage. This connection between genotype and phenotype enables large libraries of proteins to be screened and amplified in a process called *in vitro* selection, which is analogous to natural selection. These display technologies can be used to screen at least two different types of libraries. The first strategy is to create libraries of the glycoprotein itself (i.e., using error-prone PCR, DNA shuffling, etc), where variants can be produced with additional glycosylation sites that may be improved with respect to activity or stability following the introduction of additional (but identical) glycan structures. The second strategy is to make a large collection of different glycan structures by making libraries of individual pathway enzymes (i.e., using error-prone PCR, DNA shuffling, etc) or different enzyme combinations such that a combinatorial library of different glycan structures is produced and displayed on the cell or phage surface. The phenotype of the variant glycoprotein or the variant glycan structure is physically coupled to the genotype of the isolated cells (i.e., the sequence of the plasmid) or phages (i.e., the sequence of the packaged DNA known as a phagemid). Thus, the identity of the library clones is easily determined.

**Display of N-Linked Glycoproteins on the Bacterial Cell Surface**

**[0093]** Glycosylation in *E. coli* for high-throughput screening can be carried out with the host cells and methods described above using eukaryotic glycosyltransferase activity, eukaryotic flippase activity, and eukaryotic oligosaccharyl activity. However, in the screening embodiment, activity from other sources can be utilized. For example, such bacterial surface display can be carried out with the *C. jejuni* CjaA protein as an outer membrane anchor (Figure 6A). This protein is suitable primarily, because it is (i) localized to the outer membrane in *C. jejuni* and *E. coli* cells and (ii) glycosylated by the pgl system in *E. coli* (Figure 6A). To determine if the *N*-glycan heptasaccharide on CjaA is surface exposed, pgl+ *E. coli* can be treated with a fluorescently labeled version of the lectin SBA (SBA-Alexa Fluor 488 conjugate, Molecular Probes). The cells further lacked the native *E. coli* WaaL ligase that transfers oligosaccharides from the bactoprenol lipid carrier to the lipid A core molecule (Raetz et al., "Lipopolysaccharide endotoxins," *Annu Rev Biochem* 71:635-700 (2002), which is hereby incorporated by reference in its entirety) (Figure 6B). This ligase is known to have relaxed substrate specificity and is responsible for transfer of the bacterial heptasaccharide from bactoprenolpyrophosphate to the lipid A core, a molecule that is subsequently transferred to the outer side of the outer membrane. When pgl+ cells lacking *waaL* are transformed with the CjaA plasmid and induced to express CjaA, a strong fluorescent signal is detected following SBA-Alexa Fluor labeling (Figure 6C). Importantly, this signal is dependent on the pgl system as a complete loss of fluorescence was observed following SBA-Alexa Fluor labeling of *waaL* mutants carrying the pgl- control vector (Figure 6C). Accordingly, glycan analysis can be performed directly with living *E. coli* cells in a fluorescent format that is compatible with high-throughput screening.

**[0094]** Using a fluorescent version of the lectin Concanavalin A (ConA), which has a high affinity towards the tri-mannose structure of the core glycan, $Man_3GlcNAc_2$ can be assayed on the surface of *E. coli* cells. The basis for this strategy is the observation that bactoprenolpyrophosphate-linked oligosaccharides are the substrates for the *E. coli* WaaL ligase that transfers oligosaccharides from the bactoprenol lipid carrier to the lipid A core molecule (Raetz et al., "Lipopolysaccharide endotoxins," *Annu Rev Biochem*

71:635-700 (2002), which is hereby incorporated by reference in its entirety) (see Figure 6B). Applicants expect the transfer of the $Man_3GlcNAc_2$ oligosaccharide from bactoprenolpyrophosphate to the lipid A core, a molecule that is subsequently transferred to the outer side of the outer membrane. The display of $Man_3GlcNAc_2$ on the surface of *E. coli* cells will be achieved by surface staining using a fluorescent version of ConA (AlexaFluor-ConA). This should make it possible to detect and quantify oligosaccharide biosynthesis using fluorescence activated cell sorting (FACS). Importantly, this measurement does *not* depend on flippase or OST activity. It has been observed that the bacterial oligosaccharide is localized to the outer surface of TG1 pgl+ cells as evidenced by a strong FACS signal following labeling with fluorescent SBA. Identical labeling of TG1 pglmut cells resulted in identical fluorescence profile, indicating that transfer of the oligosaccharide to lipid A did not depend on PglB. Finally, control cells lacking the pgl expression vector resulted in no detectable cell fluorescence. Applicants anticipate that this assay will allow optimization of oligosaccharide expression in *E. coli* with different inducible promoters and, if necessary, different signal peptides to direct correct insertion into the plasma membrane will be used. For instance, despite the promising expression results described in the present invention, it may prove useful to employ SRP- or YidC-dependent targeting (Luirink et al., "Biogenesis of Inner Membrane Proteins in Escherichia coli," *Annu Rev Microbiol* 59:329-55 (2005), which is hereby incorporated by reference in its entirety) of each Alg membrane protein in combination with an *E. coli* host strain such as C41(DE3) that has been specifically engineered for high-level expression of heterlogous membrane proteins (Miroux et al., "Over-production of Proteins in Escherichia coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels," *J Mol Biol* 260:289-98 (1996), which is hereby incorporated by reference in its entirety). Moreover, since deletion of *waaL* eliminates oligosaccharide transfer to lipid A (Figure 6C), the ConA labeling strategy can be used in combination with a surface displayed glycoprotein (e.g., CjaA, see Figure 6) to assay for glycoprotein variants with improved or new properties (e.g., increased activity or stability) or pathway enzymes such as glycosyltransferase, flippase or OST with improved or new activities (e.g., ability to create different or novel glycan structures) (see

Figure 10B). This can be accomplished as follows: the DNA encoding the protein or peptide of interest is itself a surface protein (e.g., *C. jejuni* CjaA) or is ligated in-frame to a cell surface protein (e.g., *E. coli* ClyA, OmpA, OmpX, etc). Multiple cloning sites are sometimes used to ensure that the fragments are inserted in all three possible frames so that the cDNA fragment is translated in the proper frame. The gene encoding the cell surface hybrid protein is cloned in an expression vector and transformed into bacterial cells such as TG1 or XL1-Blue *E. coli*. For creating glycoprotein variants, the incorporation of many different DNA fragments encoding either target glycoprotein as fusion to the cell surface protein gene generates a surface displayed library from which members of interest can be isolated. For creating pathway enzymes with new or improved activities, a DNA library of the enzyme is co-transformed into bacteria along with a plasmid expressing a reporter cell surface displayed glycoprotein (e.g., CjaA) that serves as carrier for the glycan structure or library of glycan structures. Co-transformed bacteria are then screened for the presence of a particular glycan structure attached to the surface displayed carrier.

***E. coli* GlycoPhage Display System**

**[0095]** Another aspect of the present invention relates to a bacterial phage display system for glycans. The GlycoPhage display system is a powerful tool for engineering novel glyco-phenotypes with one embodiment being shown in Figure 7. This is based on a modified version of filamentous phage display (Smith, G. P., "Filamentous Fusion Phage: Novel Expression Vectors that Display Cloned Antigens on the Virion Surface," *Science* 228:1315-7 (1985), which is hereby incorporated by reference in its entirety), where phagemids expressing AcrA of *C. jejuni* lacking the N-terminal lipid anchor sequence (Nita-Lazar et al., "The N-X-S/T Consensus Sequence is Required But Not Sufficient for Bacterial N-linked Protein Glycosylation," *Glycobiology* 15:361-7 (2005), which is hereby incorporated by reference in its entirety) fused to the N-terminus of the minor phage coat protein g3p were constructed. The pectate lyase B signal sequence (*pelB*) was cloned upstream of the *acrA* coding sequence for Sec-dependent translocation to the periplasmic space of *E. coli*. Expression of the fusion protein was directed by the

arabinose inducible and glucose repressible pBAD promoter (Miyada et al., "Regulation of the araC Gene of Escherichia coli: Catabolite Repression, Autoregulation, and Effect on araBAD Expression," *Proc Natl Acad Sci U S A* 81:4120-4 (1984), which is hereby incorporated by reference in its entirety). A 24-amino acid linker was juxtaposed between the expressed AcrA and the g3p domain on phagemid pAcrA-g3p. This linker sequence contained a hexa-histidine tag and an enterokinase cleavage site directly followed by an amber stop codon (UAG), that was transcribed as glutamine in *E. coli supE* strains (e.g., XL1-Blue, ER2738 or TG1) with an efficiency of 80-90% (Miller et al., "Effects of Surrounding Sequence on the Suppression of Nonsense Codons," *J Mol Biol* 164:59-71 (1983), which is hereby incorporated by reference in its entirety). Inclusion of the phage F1 intergenic region (ori M13) on these vectors allowed for packaging of single-stranded phagemid after superinfection with helper phage. This technique can be used to assay for glycoprotein variants with improved or new properties (e.g., increased activity or stability) or pathway enzymes such as glycosyltransferase, flippase or OST with improved or new activities (e.g., ability to create different or novel glycan structures) (see Figure 10B). For creating glycoprotein variants, the DNA encoding the protein or peptide of interest is ligated to the pIII or pVIII gene. Multiple cloning sites are sometimes used to ensure that the fragments are inserted in all three possible frames so that the cDNA fragment is translated in the proper frame. The phage gene and insert DNA hybrid is then transformed into bacterial cells such as TG1 or XL1-Blue *E. coli*. The phage particles will not be released from the *E. coli* cells until they are infected with helper phage, which enables packaging of the phage DNA and assembly of the mature virions with the relevant protein fragment as part of their outer coat on either the minor (pIII) or major (pVIII) coat protein. The incorporation of many different DNA fragments into the pIII or pVIII genes generates a library from which members of interest can be isolated. For the creation of pathway enzymes with improved or new activities such as the ability to synthesize different glycan structures, a DNA library of the enzyme(s) is co-transformed into bacteria along with a plasmid expressing a reporter phage displayed glycoprotein (e.g., AcrA or MBP with N- or C-terminal glyc-tag) that serves as a carrier for the glycan structure or library of glycan structures. Co-transformed bacteria are used to create phage

libraries that are then screened for the presence of a particular glycan structure attached to the phage displayed carrier.

**[0096]** The above disclosure generally describes the present invention. A more specific description is provided below in the following examples. The examples are described solely for the purpose of illustration and are not intended to limit the scope of the present invention. Changes in form and substitution of equivalents are contemplated as circumstances suggest or render expedient. Although specific terms have been employed herein, such terms are intended in a descriptive sense and not for purposes of limitation.

## EXAMPLES

### Example 1 - *N*-linked Protein Glycosylation in Genetically Modified *E. coli*

**[0097]** The experiments of Wacker *et al.* are reproduced here, where the *C. jejuni* pgl genetic locus was functionally transferred to *E. coli*, conferring on these cells the ability to perform *N*-linked protein glycosylation (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety). For these studies, the plasmid pACYC184-*pgl* (pgl+) and a control plasmid derived from pACYC184-*pgl* that carried an insertion mutation in the *pglB* gene encoding the essential OST (pACYC184-*pglB*::kan; pgl-) were employed. BL21(DE3) *E. coli* cells were co-transformed with either a pgl+ or pgl- vector along with a second vector encoding the *C. jejuni* glycoprotein PEB3. His-tagged PEB3 was expressed in the periplasm in pgl+ and pgl- cells and purified from the periplasmic fraction using nickel affinity chromatography; Ni-NTA Spin Kit (QIAGEN). Purified PEB3 was serially-diluted and detected by Western blotting using an anti-polyhistidine antibody (Sigma). Glycosylated PEB3 was detected using the GalNAc-specific lectin soy bean agglutinin (SBA) which binds to the terminal α-linked GalNAc of the heptasaccharide glycan. As expected, it was observed that PEB3 was expressed efficiently in both pgl+ and pgl- cells (Figure 3A), but only the PEB3 from

pgl+ cells cross-reacted with the lectin SBA (Figure 3B), which binds to the terminal α-linked GalNAc of the glycan and indicates a fully glycosylated protein.

**Example 2 - *N*-linked Glycosylation of MBP with a Peptide Tag in Glyco-engineered *E. coli***

**[0100]** *E. coli* maltose binding protein (MBP) was fused to a gene encoding four consecutive glycosylation sequons (GAT CAG AAC GCG ACC GGC GGT GAC CAA AAT GCC ACA GGT GGC GAT CAA AAC GCC ACC GGC GGT GAC CAG AAT GCG ACA) (SEQ ID NO: 13) in the *SacI* and *HindIII* sites of pTRC99A [Amersham Biosciences]. The gene encodes a peptide tag of four consecutive DQNAT SEQ ID NO: 14 peptides separated by two glycine residues. DQNAT (SEQ ID NO: 14) sequons were efficiently glycosylated by PglB during *in vitro* experiments (Chen et al., "From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-linked Glycosylation in C. jejuni," *Biochemistry* 46:5579-85 (2007), which is hereby incorporated by reference in its entirety). Such a tag fused to the C-terminus of MBP, also appended with a C-terminal 6xHis tag for purification, was expressed in BL21(DE3) *E. coli* transformed with pACYC-pgl and pACYC-pgl*mut* (PglB W458A, D459A) (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety). Further, the *C. jejuni* glycoprotein cjAcrA, MBP with an N-terminal tag prior to the mature domain of MBP, MBP lacking its native secretion signal peptide with a C-terminal tag, and MBP & green fluorescent protein (GFPmut2) with a C-terminal tag and a Tat-specific (ssTorA) signal peptide were expressed in an identical manner and purified by nickel affinity chromatography (Ni-NTA Spin Kit, Quiagen). Tags at the N-terminus or C-terminus of mature MBP were determined to be glycosylated by Western blot with anti-HIS serum (Promega) against the protein and hR6P serum that was raised against the bacterial heptasaccharide. Glycosylation was dependent on both a functional PglB and secretion to the periplasm, as neither MBP generated in *E. coli* transformed with pACYC-pgl*mut* nor lacking a secretion signal peptide were glycosylated. Glycosylation occurred via the twin-arginine translocation (Tat) pathway as

evidenced by the glycosylation of MBP and green fluorescent protein (GFP) targeted for secretion in this manner. The anti-heptasaccharide serum revealed at least three discrete bands characteristic of multiple attached N-glycans.

**[0101]** These results show that a peptide containing four consecutive D-$X_1$-N-$X_2$-T sequons were efficiently glycosylated by PglB during *in vitro* experiments (Chen et al., "From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-linked Glycosylation in C. jejuni," *Biochemistry* 46:5579-85 (2007), which is hereby incorporated by reference in its entirety). A GlycTag fused to the C-terminus of MBP was expressed in pgl+ and pgl- *E. coli* and purified to 20 mg/L. The resulting protein was efficiently glycosylated at multiple sites (Figures 4C and 4D). Similar results were seen when the GlycTag was moved to the N-terminus of MBP. MBP-GlycTag fusions generated in pgl- *E. coli* or expressed without a secretion signal peptide were not glycosylated (Figure 4C), confirming that glycosylation was dependent upon PglB and export to the periplasm, respectively. The GlycTag was compatible with other secretion pathways such as the twin-arginine translocation (Tat) pathway as evidenced by the glycosylation of MBP and GFP targeted for Tat-dependent export (Figure 4C). In certain aspects, glycosylation of the GlycTag on MBP is more efficient than the glycosylation of even the natural glycoprotein *C. jejuni* AcrA (Figure 4C). Since MBP has recently been demonstrated as a model protein carrier for glycoconjugate vaccines (Fernandez et al., "Potential Role for Toll-like Receptor 4 in Mediating Escherichia Coli Maltose-binding Protein Activation of Dendritic Cells," *Infect Immun* 75:1359-63 (2007), which is hereby incorporated by reference in its entirety), it is envisioned that the MBP-GlycTag fusions can serve as potent glycoconjugate vaccines against the pathogenic bacterium *C. jejuni* or, as new glycan structures are generated, against other infectious agents. As many as 12 glycans per protein are possible if GlycTags are introduced to both N- and C-termini as well as inserted into permissive sites within MBP (Betton et al., "Creating a Bifunctional Protein by Insertion of Beta-lactamase into the Maltodextrin-binding Protein," *Nat Biotechnol* 15:1276-9 (1997), which is hereby incorporated by reference in its entirety). These MBP glycoconjugates would contain far more glycans than any naturally occurring glycoprotein (Ben-Dor et al., "Biases and Complex Patterns in the Residues Flanking

Protein N-Glycosylation Sites," *Glycobiology* 14:95-101 (2004), which is hereby incorporated by reference in its entirety).

**Example 3 - IgG M18.1 Glycosylation in *E. coli***

**[0102]** This example describes glycosylation of complex human glycoproteins in the periplasm of glyco-engineered *E. coli*. Specifically, a full-length human immunoglobulin (IgG M18.1) against anthrax toxin was expressed in pMAZ360 M18.1 (Mazor et al., "Isolation of Engineered, Full-length Antibodies from Libraries Expressed in Escherichia coli," *Nat Biotechnol* 25:563-5 (2007), which is hereby incorporated by reference in its entirety)) was mutated via site-directed mutagenesis (Quik Change Kit, Qiagen) such that the glutamine residue at residue 295 in the IgG heavy chain ($C_H2$) was mutated to aspartic acid to introduce the *bacterial* glycosylation motif D-$X_1$-N-$X_2$-S/T (Figure 5A) using primers (5' – gacaaagccgcgggaggaggattacaacagcacgtaccgtg – 3' and 5' – cacggtacgtgctgttgtaatcctcctcccgcggctttgtc – 3') (SEQ ID NO:15 AND SEQ ID NO: 18, respectively). Following expression in the periplasm of BL21(DE3) *E. coli* transformed with pACYC-pgl or pACYC-pgl*mut* (PglB W458A, D459A) (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety), IgG M18.1 were purified from cell lysate via Protein A/G affinity chromatography (Nab Protein AG Spin Kit, Pierce) and subject to SDS-PAGE in non-reducing 12% SDS gels and Western blotted with detection via anti-human IgG (Promega)and hR6P antiserum raised against the bacterial heptasaccharide. A characteristic IgG banding pattern was seen for IgG M18.1 isolated from BL21(DE3) *E. coli* transformed with pACYC-pgl or pACYC-pgl*mut*. Only IgG M18.1 from BL21(DE3) *E. coli* transformed with pACYC-pgl cross-reacted with bacterial *N*-glycan specific anti-serum (hR6P). This IgG banding pattern for pgl+ and pgl- cells is seen in Figures 5B and 5C. However, only IgG M18.1 from pgl+ cells cross-reacted with bacterial *N*-glycan specific anti-serum (hR6P) (Figure 5B and 5C). These results indicate that human IgGs can be glycosylated in the periplasm of glyco-engineered *E. coli* cells. Accordingly, in various embodiments, the present invention provides glycosylated human IgGs produced in glyco-engineered *E. coli* cells.

**Example 4 - Display of *N*-linked Glycoproteins on the Bacterial Cell Surface**

**[0103]** *E. coli* BW25113 *waaC*:Kan transformed with pACYC-pgl (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) expressing the *C.jejuni* CjaA protein from plasmid pCjaA as an outer membrane anchor displayed the bacterial heptasaccharide on the cell surface. Plasmid pCjaA was constructed by inserting the coding region for *C. jejuni* CjaA into pBAD18 appended with the coding region for a C-terminal FLAG epitope tag. Display was detected by incubating cells with soybean agglutinin conjugated to fluorescent dye (SBA-Alexa Fluor 488 conjugate, Molecular Probes) and analyzed by flow cytometry. *E. coli* BW25113 *waaC*:Kan transformed with pCjaA and pACYC-pgl*mut* (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) or pCjaA alone, did not result in fluorescent labeling. Glycan attachment was confirmed by subjecting total cellular protein from *E. coli* BW25113 *waaC*:Kan expressing the *C.jejuni* CjaA protein from plasmid pCjaA transformed with either pACYC-pgl or pACYC-pgl*mut* (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) to Western blot analysis followed by probing with hR6P antiserum raised against the bacterial heptasaccharide.

**Example 5 - *E. coli* GlycoPhage Display System**

**[0104]** Expression of AcrA-g3p from phagemid pAcrA-g3p was performed in *E. coli* TG1 cells that carry either the native (pgl+) or a mutated version (pglmut) of the *C. jejuni* glycosylation locus and the appearance of AcrA-g3p was analyzed in whole cell lysates. Immunoblot analysis with AcrA-specific antiserum showed a signal in cell lysates of both TG1 pgl+ and TG1 pglmut after 3, 5 and 16 h of induction with 50 mM arabinose (Figures 8A and 8B, lanes 3 to 5). Anti-AcrA cross-reacting proteins with an

apparent molecular mass of about 80 kDa corresponded well to the calculated mass of the AcrA-g3p fusion protein of 80.8 kDa. The same lysates were probed with glycosylation-specific antiserum (R12) that had been raised against *C. jejuni* whole cell extracts and shows a strong preference towards the glycosylated form of AcrA (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) (Figure 8C). Here, immunoreactive bands with a molecular mass of about 80 kDa can only be detected in whole cell lysates of TG1 pgl+ cells after 3, 5, and 16 h of induction (Figure 8C, lanes 3 to 5). These data prove that the AcrA-g3p fusion protein was glycosylated by the *C. jejuni* pgl system.

**Example 6 - Time-dependent Expression and Glycosylation of AcrA-g3p**

**[0105]** *E. coli* TG1 expressed a fusion of *C. jejuni* AcrA to the g3p phage coat protein from a plasmid comprised of the pAra-AcrA-g3p. In pAra-Acra-g3p, the pectate lyase B signal sequence (*pelB*) was cloned upstream of the *acrA* coding sequence for Sec-dependent translocation to the periplasm of *E. coli*. Expression of the fusion protein was directed by the arabinose inducible and glucose repressible pBAD promoter. A 24-amino acid linker was juxtaposed between the expressed AcrA and the g3p domain. This linker sequence contained a hexa-histidine tag and an enterokinase cleavage site directly followed by an amber stop codon (UAG), that is transcribed as glutamine in *E. coli supE* strains (e.g., TG1). Inclusion of the phage F1 intergenic region (ori M13) on these vectors allowed for packaging of single-stranded phagemid after superinfection with helper phage. TG1 cells harboring pAra-AcrA-g3p were transformed with either pACYC-pgl or pACYC-pgl*mut* (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) and whole cell lysates were prepared from either non-induced cells, or from cells incubated with 50 mM arabinose for 1 h, 3 h, 5 h, and 16 h. Proteins were separated by10% SDS-PAGE with protein standards, transferred to a nitrocellulose membrane, and visualized with AcrA-specific serum or with R12 antiserum raised against the bacterial heptasaccharide.

**Example 7 - Quantification of Glycophage Enrichment by SBA Biopanning**

**[0106]** Phages produced from TG1 cells harboring pAra-AcrA-g3p and either pACYC-pgl or pACYC-pgl*mut* (Wacker et al., "N-linked Glycosylation in Campylobacter jejuni and its Functional Transfer into E. coli," *Science* 298:1790-3 (2002), which is hereby incorporated by reference in its entirety) were applied to immobilized soybean agglutinin (SBA) for column purification. The total colony forming units (CFUs) present in each fraction of the SBA panning procedure were determined by infection of fresh TG1 cells and are the means of at least three independent experiments. The number of phages subjected to SBA panning and the resulting CFUs after fresh infection varied by less than 6%. The fractions were: Fraction 1, CFUs applied to the SBA column; fraction 2, SBA flow-through; fractions 3 and 4, PBS washing steps; fraction 5, 6, and 7, washing steps with 30 mM galactose in PBS; fraction 8, 9, and 10, elution steps with 300 mM galactose in PBS. The presence of AcrA was visualized with anti-AcrA serum and the presence of the bacterial heptasaccharide was visualized with R12 antiserum raised against the bacterial glycan such that: Lane 1, raw phage preparation; lane 2, SBA flow-through; lanes 3 and 4, wash fractions with PBS; lanes 5 to 7, wash fractions with 30 mM galactose in PBS; lanes 8 to 10, elution fractions with 300 mM galactose in PBS. In lanes 1 to 4, $1x10^8$ phages were applied to SDSPAGE. In lanes 5 to 10, $3.5 \times 10^7$, $1.2 \times 10^4$, $4.0x10^3$, $1.3x\ 10^6$, $2.5x10^6$, $1.2x10^6$ phages prepared from TG1 cells harboring pAra-AcrA-g3p and pACYC-pgl or $1.5x10^6$, $3.5x10^3$, $3.0x10^3$, $4.5x10^3$, $0.5x10^4$, $1.5x10^3$ phages prepared from TG1 cells harboring pAra-AcrA-g3p and pACYC-pgl*mut* were analyzed, respectively.

**[0107]** Phage titers of $<9.0x10^{11}$ for TG1 pgl+ and $<8.7x10^{11}$ for TG1 pglmut, each expressing pAcrA-g3p, per ml of culture supernatant were obtained. In order to determine whether glycosylated AcrA-g3p fusion protein was present in the phage preparation, an SBA bio-panning procedure was developed that allows specific enrichment of glycophages (Figure 7). Phage preparations from TG1 pgl+ or TG1 pglmut, each expressing pAcrA-g3p, were mixed with agarose bound SBA, unbound

phages were removed by successive washing steps with PBS and PBS containing 30 mM galactose. Galactose binds to SBA with an equilibrium association constant of $2x10^2$ $M^{-1}$ and could therefore be used to compete with the bound oligosaccharide (Swamy et al., "Thermodynamic and Kinetic Studies on Saccharide Binding to Soya-Bean Agglutinin," *Biochem* J 234:515-22 (1986), which is hereby incorporated by reference in its entirety). Similar titers were found in the respective wash fractions for both phage preparations. In contrast, a $10^3$-fold increase in phage titer ($10^3$ to $10^6$ cfu/ml) was observed in the eluate with 300 mM galactose for phages from TG1 pgl+ expressing pAcrAg3p, while the titer stayed at the background level of $10^3$ cfu/ml for phages from TG1 pglmut expressing pAcrA-g3p (Figure 9A). This PglB-dependent accumulation of SBA-bound phage demonstrates the production of infective glycophage and their specific enrichment by the panning procedure. Next, the presence of glycosylated AcrA-g3p fusion protein in the fractions of both SBA panning experiments was confirmed. Upon immunodetection after SDS-PAGE separation of total phage protein, a signal corresponding to AcrA-g3p was detected with AcrA-specific antibodies (Figures 9A and 9B, lane 1). AcrA-g3p specific bands were also present in the flow-through and in the PBS washing steps (Figure 9A and 9B, lanes 2, 3, and 4). A clear glycosylation-specific signal was present with the R12 antiserum in the elution fraction when phages produced from TG1 pgl+ expressing pAcrA-g3p were used for panning (Figure 9B, panel c, lanes 8, 9, and 10). The R12 antiserum shows a high preference to the glycosylated form of AcrA but also reacts with non-glycosylated AcrA when present in high amounts (Figure 9B, panels c and d, lanes 1 to 4). Importantly, the AcrA fusion protein detected in phage preparations eluted with high galactose concentrations (Figure 9B, panel c, lanes 8, 9, and 10) migrates slower than the fusion protein detected in the flowthrough (Figure 9B, panel c, lane 1), which agrees with the glycosylation of the protein. As expected, phages derived from glycosylation deficient strain TG1 pglmut expressing pAcrA-g3p did not display R12 reacting fusion protein (Figure 9B, panel d, lanes 8, 9, and 10). Therefore, phages produced in the presence of a functional *C. jejuni pgl* operon displayed glycosylated AcrA on the surface and these glycophage were enriched by SBA panning.

**[0108]** To further test the specificity of this method, purified glycophage ($8x10^6$ cfu) produced by TG1 (pgl+, pAcrA-g3p) was mixed with an excess (1 to $10^4$-fold) of aglycosylated phage produced by TG1 (pglmut, pAcrA-g3p) and applied to a single SBA panning step. This experimental setup allowed the differentiation between glycophage and aglycosylated phage by restriction analyses of their phagemids. Glycophage contained the amber stop codon in the AcrA-g3p expression cassette while the aglycosylated phage did not. About 96% ($7.7x10^6 \pm 0.3x10^6$) phages were recovered when only glycophage was applied to SBA panning. When glycophage were mixed with an equal amount or $10^2$ to $10^4$-fold excess of aglycosylated phage, an average of 96% ($7.8x10^6 \pm 0.2x10^6$), 93% ($7.4x10^6 \pm 0.5x10^6$), and 79% ($6.3x10^6 \pm 1.0x10^6$) of phage were bound by SBA, respectively, and subsequently found in the eluate. Applying exclusively aglycosylated phage ($8x10^{10}$) significantly lowered the amount of infectious particles ($3.8x10^4 \pm 0.2x10^4$) that were bound to SBA. To demonstrate that phages in the elution fraction were indeed glycophage, 12 phagemids from each reconstitution were analyzed by *EagI-EheI* digestion that allowed the differentiation between glycophage and non-glycosylated phage. At least 11/12 phagemids showed the restriction pattern of phagemid pAcrA-g3p that was packed into glycophage produced by TG1 (pgl+, pAcrA-g3p) cells. In the elution fraction where only glycophage was used (positive control), 12/12 phagemids showed the glyco-phagemid specific DNA fragments. These data unequivocally establish that: (i) glycosylated proteins carrying the *N*-linked heptasaccharide are amenable to multivalent display on filamentous phage; (ii) the glycophage purification procedure works efficiently, enrichment factors as high as $10^4$ were obtained per round of SBA panning, and (iii) glycophage did not lose infectivity even when subjected to two rounds of SBA panning.

**Example 8 - Expression and Localization of Yeast Glycosyltransferases in *E. coli***

**[0109]** The generation of the $Man_3GlcNAc_2$ oligosaccharide structure requires the functional expression of several eukaryotic glycosyltransferases in *E. coli* and represents a classical example of "pathway engineering" (see Figure 10B).

*WecA-catalyzed transfer of first GlcNAc to lipid carrier.*

**[0110]** Bactoprenylpyrophosphate serves as a carrier for the assembly of an oligosaccharide at the cytoplasmic side of the inner membrane (Feldman et al., "Engineering N-linked Protein Glycosylation With Diverse O Antigen Lipopolysaccharide Structures in Escherichia coli," *Proc. Nat'l Acad. Sci. USA* 102:3016-3021 (2005), which is hereby incorporated by reference in its entirety). In *E. coli*, bactoprenol-*PP*-GlcNAc is generated from bactoprenol-*P* and UDP-GlcNAc by the WecA protein (Valvano, M. A., "Export of O-specific Lipopolysaccharide," *Frontiers in Bioscience* 8:S452-S471 (2003), which is hereby incorporated by reference in its entirety). Therefore, this endogenous substrate can be used as a starting molecule in an artificial pathway that creates the $Man_3GlcNAc_2$ oligosaccharide. The reducing end GlcNAc residue is essential for the substrate recognition by eukaryotic OSTs (Tai et al., "Substrate Specificity of the Glycosyl Donor for Oligosaccharyl Transferase," *J Org Chem* 66:6217-28 (2001), which is hereby incorporated by reference in its entirety) but also fulfills the requirements of a prokaryotic OST substrate (Wacker et al., "Substrate Specificity of Bacterial Oligosaccharyltransferase Suggests a Common Transfer Mechanism for the Bacterial and Eukaryotic Systems," *Proc. Nat'l Acad. Sci. USA* 103:7088-7093 (2006), which is hereby incorporated by reference in its entirety).

**Example 9 - Expression of yeast Alg13/14 in *E. coli***

**[0111]** The β1,4 GlcNAc transferase for the addition of the second GlcNAc residue has recently been identified in yeast (Bickel et al., "Biosynthesis of Lipid-linked Oligosaccharides in Saccharomyces cerevisiae - Alg13p AND Alg14p Form a Complex Required for the Formation of GlcNAc(2)-PP-dolichol," *J. Biol. Chem.* 280:34500-34506 (2005), which is hereby incorporated by reference in its entirety). This enzyme is a complex of two proteins, encoded by the *ALG13* and the *ALG14* locus from *Saccharomyces cerevisiae*. Alg14 is an integral membrane protein, whereas Alg13 is peripherally associated with the cytoplasmic side of the ER membrane by virtue of its association with Alg14. The reason for testing Δ*dnaJ* cells is that inactivation of *dnaJ* is known to increase membrane protein expression and suppress the severe cytotoxicity

associated with their expression (Skretas et al., "Genetic Analysis of G Protein-coupled Receptor Expression in Escherichia coli: Inhibitory Role of DnaJ on the Membrane Integration of the Human Central Cannabinoid Receptor," *Biotechnol Bioeng* (2008), which is hereby incorporated by reference in its entirety).

**[0112]** Total protein from MC4100 *E. coli* cells expressing Alg13 appended with a C-terminal 6xHIS tag from plasmid pBAD18-Cm. Alg14 appended with a C-terminal FLAG epitope tag was subject to centrifugation at 20,000xg for 20 mins. The supernatant represents the soluble fraction and the pellet, the insoluble fraction. The soluble fraction was further spun at 100,000xg for 1 hr and the supernatant and pellet were collected as the soluble and membrane fractions, respectively. The soluble cytoplasmic fraction collected 0, 1, 2, and 3 hrs post induction with 0.2% arabionse was probed with anti-6xHIS antibody (Promega) to detect Alg13-6xHIS. Western blot analysis was used to compare fractions isolated from MC4100 and MC4100 Δ*dnaJ* cells and probed with anti-FLAG antibody to detect Alg14-FLAG collected at 3 hours post induction. Soluble expression of Alg13 in the cytoplasm (Figure 11A) and correct insertion of Alg14 in the inner membrane (Figure 11B) were observed.

**[0113]** Next, GlcNAc transferase activity will be tested in extracts derived from transformed *E. coli* cells (Bickel et al., "Biosynthesis of Lipid-linked Oligosaccharides in Saccharomyces cerevisiae - Alg13p and Alg14p Form a Complex Required for the Formation of GlcNAc(2)-PP-dolichol," *J. Biol. Chem.* 280:34500-34506 (2005), which is hereby incorporated by reference in its entirety) and the *in vivo* formation of bactoprenol-*PP*-$GlcNAc_2$ will be analyzed by labeling the cells with $^3$H-GlcNAc and analysis of glycolipids using standard methods (Bickel et al., "Biosynthesis of Lipid-linked Oligosaccharides in Saccharomyces cerevisiae - Alg13p AND Alg14p Form a Complex Required for the Formation of GlcNAc(2)-PP-dolichol," *J. Biol. Chem.* 280:34500-34506 (2005), which is hereby incorporated by reference in its entirety).

**Example 10 - Expression of Alg1 and Alg2 in *E. coli***

**[0114]** The process of the present invention further requires the expression of active Alg1 protein, the β1,4 mannosyltransferase, and bifunctional Alg2

mannosyltransferase, catalyzing the addition of both the α1,3 and α1,6 mannose residue to the oligosaccharide. Each of these enzymes has been previously expressed in an active form in *E. coli* (O'Reilly et al., "In vitro Evidence for the Dual Function of Alg2 and Alg11: Essential Mannosyltransferases in N-linked Glycoprotein Biosynthesis," *Biochemistry* 45:9593-603 (2006) and Wilson et al., "Dolichol is Not a Necessary Moiety for Lipid-linked Oligosaccharide Substrates of the Mannosyltransferases Involved in In vitro N-linked-oligosaccharide Assembly," *Biochem J* 310 ( Pt 3):909-16 (1995), which are hereby incorporated by reference in their entirety)

**[0115]** Total protein from MC4100 Δ*dnaJ E. coli* cells expressing Alg1 and Alg2 each appended with a C-terminal 6xHIS tag and, the case of Alg2, an N-terminal thioredoxin (TrxA) solubility tag was subject to centrifugation at 20,000xg for 20 mins, the supernatant was collected and further spun at 100,000xg for 1 hr and the pellet from this spin was collected as the membrane fraction. Membrane fractions were harvested from cells at 3, 4 and 5 hours post induction. Blots were probed with anti-6xHIS antibody (Promega). As shown in Figure 12, each localizes correctly in the inner membrane of *E. coli*.

**[0116]** Next, mannosyltransferase activity will need to be tested in extracts derived from transformed *E. coli* cells according to an established protocol (O'Reilly et al., "In vitro Evidence for the Dual Function of Alg2 and Alg11: Essential Mannosyltransferases in N-linked Glycoprotein Biosynthesis," *Biochemistry* 45:9593-603 (2006) and Schwarz et al., "Deficiency of GDP-Man:GlcNAc2-PP-dolichol Mannosyltransferase Causes Congenital Disorder of Glycosylation Type Ik," *Am J Hum Genet* 74:472-81 (2004), which are hereby incorporated by reference in their entirety).

**Prophetic Example 11 - Construction of an Artificial *Alg* Operon**

**[0117]** After verification that each enzyme can be functionally expressed in *E. coli*, a gene cluster encoding all four of the above yeast enzymes will be constructed on a single plasmid backbone. Co-expression of the four Alg enzymes will be performed in wt, Δ*dnaJ* mutants, and in the strain C41(DE3) that has been previously optimized for

membrane protein expression (Miroux et al., "Over-production of Proteins in Escherichia coli: Mutant Hosts That Allow Synthesis of Some Membrane Proteins and Globular Proteins at High Levels," *J Mol Biol* 260:289-98 (1996), which is hereby incorporated by reference in its entirety). Applicants expect that co-expression of the four Alg enzymes will result in the *in vivo* formation of bactoprenol-*PP*-$GlcNAc_2Man_3$. This will be confirmed by metabolic labeling cells with $^3H$-mannose for 30 min at 37°C. Bactoprenol-linked oligosaccharides will be extracted, released, and analyzed by high-performance liquid chromatography (HPLC), as described in Korner et al., "Abnormal Synthesis of Mannose 1-phosphate Derived Carbohydrates in Carbohydrate-deficient Glycoprotein Syndrome Type I Fibroblasts with Phosphomannomutase Deficiency," *Glycobiology* 8:165-71 (1998), which is hereby incorporated by reference in its entirety.

**[0118]** Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions, and the like can be made without departing from the spirit of the invention and these are therefore, considered to be within the scope of the present invention as defined the claims which follow.

**WHAT IS CLAIMED:**

1. A prokaryotic host cell comprising eukaryotic glycosyltransferase activity, wherein the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity.

2. The prokaryotic host cell of claim 1, wherein the dolichyl-linked UDP-GlcNAc transferase activity comprises Alg13 activity and Alg14 activity.

3. The prokaryotic host cell of claim 2, wherein the Alg13 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 1 or 2.

4. The prokaryotic host cell of claim 2, wherein the Alg14 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 3 or 4.

5. The prokaryotic host cell of claim 1, wherein the mannosyltransferase activity comprises Alg1 activity and Alg2 activity.

6. The prokaryotic host cell of claim 5, wherein the Alg1 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 5 or 6.

7. The prokaryotic host cell of claim 2, wherein the Alg2 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 7 or 8.

8. The prokaryotic host cell of claim 1 further comprising eukaryotic flippase activity.

9. The prokaryotic host cell of claim 8, wherein the eukaryotic flippase activity comprises Rft1 activity.

10. The prokaryotic host cell of claim 2, wherein the Rft1 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 9 or 10.

11. The prokaryotic host cell of claim 1 further comprising eukaryotic oligosaccharyl transferase activity.

12. The prokaryotic host cell of claim 11, wherein the oligosaccharyl transferase activity comprises STT3 activity.

13. The prokaryotic host cell of claim 12, wherein the STT3 activity is encoded by a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NOs: 11 or 12.

14. The prokaryotic host cell of claim 1 further comprising a protein of interest.

15. The prokaryotic host cell of claim 1, wherein the eukaryotic glycosyltransferase activity produces oligosaccharide compositions selected from the group consisting of $GlcNAc_2$, $Man_1GlcNAc_2$, $Man_2GlcNAc_2$, and $Man_3GlcNAc_2$.

16. A glycoprotein produced by the prokaryotic host cell of claim 1.

17. A glycoprotein conjugate comprising a protein and at least one peptide comprising a D-$X_1$-N-$X_2$-T motif fused to the protein, wherein D is aspartic acid, $X_1$ and $X_2$ are any amino acid other than proline, N is asparagine, and T is threonine.

18. A method of producing a glycosylated protein, said method comprising:

providing a prokaryotic host cell comprising eukaryotic glycosyltransferase activity, wherein the eukaryotic glycosyltransferase activity is eukaryotic dolichyl-linked UDP-GlcNAc transferase activity and eukaryotic mannosyltransferase activity and

culturing the prokaryotic host cell under conditions effective to produce a glycosylated protein.

19. The method of claim 18, wherein the dolichyl-linked UDP-GlcNAc transferase activity comprises Alg13 activity and Alg14 activity.

20. The method of claim 18, wherein the mannosyltransferase activity comprises Alg1 activity and Alg2 activity.

21. The method of claim 18, wherein the prokaryotic host cell further comprises a eukaryotic flippase activity.

22. The method of claim 18, wherein the eukaryotic flippase activity comprises Rft1 activity.

23. The method of claim 18 further comprising a eukaryotic oligosaccharyl transferase activity.

24. The method of claim 23, wherein the oligosaccharyl transferase activity comprises STT3 activity.

25. The method of claim 18, wherein the glycosylated protein is a glycosylated antibody comprising:

an Fv portion which recognizes and binds to a native antigen and

an Fc portion which is glycosylated at a conserved asparagine residue.

26. A method for screening bacteria or bacteriophages, said method comprising:

expressing one or more glycans on the surface of a bacteria;

attaching a label on the one or more glycans on the surface of the bacteria or on the surface of a bacteriophage derived from the bacteria; and

analyzing the label in a high-throughput format.

27. The method of claim 26, wherein bacteria comprises glycosyltransferase activity, wherein the glycosyltransferase activity is dolichyl-linked UDP-GlcNAc transferase activity and mannosyltransferase activity.

28. The method of claim 27, wherein the dolichyl-linked UDP-GlcNAc transferase activity comprises Alg13 activity and Alg14 activity.

29. The method of claim 27, wherein the mannosyltransferase activity comprises Alg1 activity and Alg2 activity.

30. The method of claim 27, wherein the bacteria further comprises flippase activity.

31. The method of claim 30, wherein the flippase activity comprises Rft1 activity or PglK activity.

32. The method of claim 27, wherein the bacteria further comprises oligosaccharyl transferase activity.

33. The method of claim 32, wherein the oligosaccharyl transferase activity comprises PglB activity or STT3 activity.

34. The method of claim 26, wherein a bacteria is subjected to said labeling and said analyzing.

35. The method of claim 26, wherein a bacteriophage is subjected to said labeling and said analyzing, said method further comprising:

infecting the bacteria expressing one or more glycans on the cell surface with a helper phage under conditions effective to produce a bacteriophage with one or more glycans on its surface.

36. The method of claim 35 further comprising:

enriching bacteriophages with one or more glycans on their surfaces.

37. The method of claim 26, wherein said labeling is carried out with a lectin which recognizes a glycan on the surface of the bacteria or bacteriophage and has a detectable label.

38. The method of claim 26, wherein said labeling is carried out with an antibody which recognizes a glycan on the surface of the bacteria or bacteriophage and has a detectable label.

40. A glycosylated antibody comprising:
an Fv portion which recognizes and binds to a native antigen and
an Fc portion which is glycosylated at a conserved asparagine residue.

41. The glycosylated antibody of claim 39, wherein the antibody is monoclonal.

42. The glycosylated antibody of claim 39, wherein the antibody is polyclonal.