

POLYSIALIC ACID, BLOOD GROUP ANTIGENS AND GLYCOPROTEIN EXPRESSION

STATEMENT OF FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0001] This invention was made with government support under grant numbers 1R43GM093483-01, 5R43AI091336-01 and 5R43AI091336-02 by the National Institutes of Health. The government has certain rights in this invention.

CROSS-REFERENCE TO RELATED APPLICATIONS

[0002] This application is related to U.S. Provisional Application No. 61/801,948, filed March 15, 2013, which is herein incorporated by reference, in its entirety, for all purposes.

SEQUENCE LISTING

[0003] This application contains a Sequence Listing which has been submitted via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on [DATE], is named [.txt] and is [#####] bytes in size.

FIELD OF INVENTION

[0004] The disclosure herein generally relates to the field of glycobiology and protein engineering. More specifically, the embodiments described herein relate to oligosaccharide compositions and production of therapeutic glycoproteins in recombinant hosts.

BACKGROUND

[0005] Protein and peptide drugs have had a huge clinical impact and constitute a \$70 billion market. Unfortunately, the efficacy of protein drugs is often compromised by

limitations arising from proteolytic degradation, uptake by cells of the reticuloendothelial system, renal removal, and immunocomplex formation. This can lead to elimination from the blood before effective concentrations are reached, and can result in unacceptably short therapeutic windows. The predominant factors that contribute to these pharmacokinetic limitations are stability and immunogenicity. Efforts have been made to address these problems, including changing the primary structure, conjugating glycans or polymers to the protein, or entrapping the protein in nanoparticles to improve residence time and reduce immunogenicity. The most popular approach to date has been conjugation to monomethoxy poly(ethyleneglycol) (mPEG) commonly referred to as PEGylation. PEGylation can endow protein and peptide drugs with longer circulatory half-lives and reduce immunogenicity. A number of PEGylated drugs are now used clinically (e.g., asparaginase, interferon α , tumor necrosis factor and granulocyte-colony stimulating factor). However, PEG is not biodegradable via normal detoxification mechanisms and the administration of PEGylated proteins has been found to elicit anti-PEG antibodies.

[0006] PEGylation is a well-accepted approach to enhance stability and reduce immunogenicity, whereby protein is conjugated to poly(ethyleneglycol) (PEG) [1]. Such PEGylation involves the covalent attachment of either linear or branched chains of PEG via a chemically reactive side-chain, such as a hydroxysuccinimidylester or an aldehyde group, for linking to either the α or ϵ amino groups on the protein [2]. PEGylation can endow protein and peptide drugs with longer circulatory half-lives and reduced immunogenicity, as PEG is water-soluble and increases the size of the protein and reduces proteolytic cleavage by occluding cleavage sites [1]. The value of PEGylation was demonstrated for several proteins, including: (i) asparaginase [3], an enzyme used in the treatment of leukemia, and (ii) adenosine deaminase [4], which participates in purine metabolism. PEGylation was also used to enhance the activity of immunological factors such as granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF) [5], tumor necrosis factor (TNF), interferon α -2a (IFN α -2a) and IFN α -2b [1]. While PEGylation is a chemical modification that can enhance pharmacokinetic properties, it is not without drawbacks. First, the heterogeneity of PEGylation yields many different isoforms of varying biological activity. This is primarily a result of the polydisperse nature of the polymer. Second, concerns have been raised about introducing a synthetic polymer into the human body that does not appear to be completely biodegradable [6]. Third, the extended half-life of PEGylated proteins that is often observed can be accompanied by reduced biological activity

related to the structural change in the molecules as a result of conjugation [2]. Fourth, the process of PEGylation is expensive and requires several *in vitro* chemical reactions and multiple purifications [7]. Thus, while PEGylation has been clinically proven as a method to increase circulatory half-lives and reduce immunogenicity, clearly it is not the optimal solution.

[0007] An emerging clinical alternative to PEGylation is polysialylation which involves attachment of a polymer of natural N-acetylneuraminic acid (polysialic acid or PSA) to the protein. PSA is highly hydrophilic with similar hydration properties to PEG, is inconspicuous to the innate and adaptive immune systems, and is naturally synthesized and displayed on human cells. PSA has recently been developed for clinical use with polysialylated versions of insulin and erythropoietin each displaying improved tolerance and pharmacokinetics. Unfortunately, as with PEGylation, the PSA conjugation process is technically challenging and expensive making the final product cost prohibitive to the healthcare consumer. PSA conjugation requires the separate production and purification of the target protein and PSA, as well as the *in vitro* reductive amination of the nonreducing end of PSA to allow chemical linkage to primary amine groups on the protein.

[0008] PSA conjugation has proven to be a very effective method to increase the active life of therapeutic proteins and prevent them from being recognized by the immune system. PSA conjugation has several performance advantages over PEGylation and is currently being tested in the clinic.

[0009] Molecules that are inconspicuous to the innate and adaptive immune systems are likely to survive for prolonged periods in circulation. Polysialic acid (PSA; polymers of N-acetylneuraminic (sialic) acid) is one such molecule and offers a natural alternative to PEG as a conjugate that can modify serum persistence of proteins. PSA is a human polymer found almost exclusively on neural cell adhesion molecule (NCAM) where it has an antiadhesive function in brain development [8]. When used for protein and therapeutic peptide drug delivery, conjugated PSA provides a protective microenvironment. This increases the active life of the therapeutic protein in circulation and prevents it from being recognized by the immune system. Unlike PEG, PSA is metabolized as a natural sugar molecule by tissue sialidases [9]. The highly hydrophilic nature of PSA results in similar hydration properties to PEG, giving it a high apparent molecular weight in the blood. This increases circulation time since no receptors with PSA specificity have been identified to date [10].

[0010] While PSA is naturally found in the human body, it is also synthesized as a capsule by bacteria such as *Neisseria meningitidis* and certain strains of *E. coli* [11]. These

polysialylated bacteria use molecular mimicry to evade the defense systems of the human body. Bacterial PSA is completely non-immunogenic, even when coupled to proteins, and is chemically identical to PSA in the human body to the extent that PSA has been developed for clinical use. Reductive amination of the nonreducing end of oxidized PSA allows *in vitro* chemical conjugation via primary amine groups on proteins, and the therapeutic benefits of PSA conjugation have been demonstrated with asparaginase [12] and insulin [13] for the treatment of leukemia and diabetes, respectively. Recent clinical data from trials with polysialylated insulin and erythropoietin showed that these biopharmaceuticals were well tolerated with enhanced pharmacokinetics [14]. Recently, two exciting discoveries have increased enthusiasm for PSA conjugation. First, it was observed that chemically polysialylated antitumor Fab fragments resulted in a 5-fold increase in bioavailability with a corresponding 3-fold increase in tumor uptake compared to unmodified Fab [15]. Second, site-specific (rather than random) coupling of PSA to engineered C-terminal thiols lead to antibody fragments with full immunoreactivity, increased blood half-life, higher tumor uptake, and improved specificity ratios [14]. PSA conjugation may add significant therapeutic value and polysialylated antibody fragments may be a viable alternative to whole IgGs by improving serum half-life and ameliorating concerns associated with Fc-domains.

[0011] Unfortunately, even PSA conjugation is not without its drawbacks. While effective in a therapeutic context, the production process of PSA conjugation is intensive and comes with a significant capital and processing cost. Currently, production involves a laborious eight-step process including: (i) fermentation of *E. coli* K1 and (ii) purification of its capsular coating, (iii) fermentation of *E. coli* expressing therapeutic protein and (iv) purification of therapeutic protein, (v) chemical cleavage of PSA from membrane anchor, (vi) purification of PSA, (vii) chemical crosslinking PSA to primary amine groups on the therapeutic protein by reductive amination of the nonreducing end of oxidized PSA, and (viii) purification of PSA-conjugated protein. This eight-step process requires two fermentations, two *in vitro* chemical reactions, and four purifications. The process is further complicated by the fact that standard amine-directed chemical conjugation of PSA results in random attachment patterns of undesirable heterogeneity [14]. To address this problem, site-specific, thiol-directed chemical conjugation can be used. However, this requires the addition of multiple C-terminal thiols, which are problematic to express in *E. coli* fermentation and require a mammalian expression system [14].

[0012] Accordingly, what is needed, therefore, is a method and composition for recombinant production of therapeutic proteins linked to an oligosaccharide composition that

is structurally homogeneous and human-like produced in a controlled, rapid and cost-effective manner.

SUMMARY

[0013] Described herein are methods and compositions for the recombinant production of human or human-like glycans including polysialic acid and blood group antigens on proteins. The present invention provides methods and compositions for the recombinant production of human or human-like glycans including A antigen, H antigen, B antigen, T antigen, sialyl T antigen, Lewis X antigen and polysialylated antigen. The methods further provide for the production of non-native carbohydrates containing human glycans in prokaryotic host cells and attaching them as *N*-linked glycans to proteins. Various host cells are engineered to express proteins required to produce the necessary sugar nucleotides and glycosyltransferase activities required to synthesize specified oligosaccharide structures.

[0014] In certain aspects, methods and compositions are provided for producing an oligosaccharide composition comprising: culturing a recombinant host cell to express GalNAc transferase activity (EC 2.4.1.-) (EC 2.4.1.290) and galactosyltransferase activity (EC 2.4.1.-) (EC 2.4.1.309); wherein the host cell produces an oligosaccharide composition comprising one or more GalNAc, galactose or galactose-GalNAc residues linked to a lipid carrier.

[0015] Additional embodiments provide expression of one or more enzyme activities selected from fucosyltransferase (EC 2.4.1.69); sialyltransferase (EC 2.4.99.4, EC 2.4.99.-, EC 2.4.99.8); and N-acetylglucosaminyl transferase (EC 2.4.1.-) for the production of an oligosaccharide composition comprising at least one fucose, sialic acid or GlcNAc residues linked to a lipid carrier.

[0016] Certain embodiments provide expression of one or more activities selected from UndP N-acetylglucosaminyl transferase (EC 7.8.33), UndPP GalNAc epimerase (EC 5.1.3.c) and UndP bacillosamine transferase (EC 2.7.8.36).

[0017] Certain embodiments provide expression of α 1,3-N-acetylgalactosamine transferase activity (EC 2.4.1.-, EC 2.4.1.306).

[0018] Certain embodiments provide expression of one or more activities selected from β 1,3 galactosyltransferase (EC 2.4.1.-) β 1,4 galactosyltransferase (EC 2.4.1.22) and α 1,3 galactosyl transferase activity (EC 2.4.1.309).

[0019] Certain embodiments provide expression of one or more activities selected from α 1,2 fucosyltransferase (EC 2.4.1.69), α 1,3 fucosyltransferase (EC 2.4.1.152), and α 1,3/1,4 fucosyltransferase (EC 2.4.1.65).

[0020] Certain embodiments provide expression of β 1,3N-acetylglucosaminyl transferase activity (EC 2.4.1.101).

[0021] Certain embodiments provide expression of one or more activities selected from α 2,3 NeuNAc transferase (EC 2.4.99.4), α 2,6 NeuNAc transferase (EC 2.4.99.1), bifunctional α 2,3 α 2,8 NeuNAc transferase (EC 2.4.99.-, EC 2.4.99.4, EC 2.4.99.8) and α 2,8 polysialyltransferase (EC 2.4.99.8).

[0022] Certain embodiments provide expression of undecaprenyl-phosphate α -N-acetylglucosaminyltransferase activity (EC 2.7.8.33).

[0023] Certain embodiments provide expression of N-acetyl- α -D-glucosaminyl-diphosphoditrans, octakis-undecaprenol 4- epimerase activity (EC 5.1.3.c).

[0024] Certain embodiments provide expression of undecaprenyl phosphate N,N'-diacetylbaucillosamine 1-phosphate transferase activity (EC 2.7.8.36).

[0025] Additional embodiments provide an attenuation in at least one of the enzyme activities selected from N-acetylneuraminate lyase (EC 4.1.3.3), undecaprenyl-phosphate glucose phosphotransferase (EC 2.7.8.-)(EC 2.7.8.31) and O-antigen ligase activity.

[0026] Certain embodiments provide expression of one or more activities selected from N-acetylneuraminate synthase (EC2.5.1.56), N-acetylneuraminate cytidyltransferase (EC 2.7.7.43), UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14) and N-acetylneuraminate acetyltransferase (EC 2.3.1.45).

[0027] Certain embodiments provide expression of GalNAc epimerase activity (EC 5.1.3.2).

[0028] Certain embodiments provide expression of Gal epimerase activity (EC 5.1.3.2).

[0029] Certain embodiments provide expression of one or more enzyme activities selected from GDP-mannose 4,6 dehydratase (EC 4.2.1.47), GDP-fucose synthetase (EC 1.1.1.271), GDP-mannose mannosyl hydrolase (EC 3.2.1.42), mannose-1-phosphate guanylttransferase (EC 2.7.7.13) and phosphomannomutase (EC 5.4.2.8).

[0030] Certain embodiments provide expression of one or more enzyme activities selected from UDP-N-acetylbaucillosamine N-acetyltransferase (EC 2.3.1.203), UDP-N-acetylglucosamine 4,6 dehydratase (EC 4.2.1.135) and UDP-N-acetylbaucillosamine transaminase (EC2.6.1.34).

[0031] In one embodiment, the invention provides a glycoprotein composition comprising an N-linked sialic acid residue on the glycoprotein. Preferably, the glycoprotein composition comprising the N-linked sialic acid residue comprises one of following glycoforms: (Sia $\alpha 2,8$)_n – Sia $\alpha 2,8$ – Sia $\alpha 2,3$ – Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GalNAc $\alpha 1,3$ – GlcNAc; (Sia $\alpha 2,8$)_n – Sia $\alpha 2,8$ – Sia $\alpha 2,3$ – Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GlcNAc; (Sia $\alpha 2,8$)_n – Sia $\alpha 2,8$ – Sia $\alpha 2,3$ – Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GalNAc, and (Sia $\alpha 2,8$)_n – Sia $\alpha 2,8$ – Sia $\alpha 2,3$ – Gal $\beta 1,3$ – (GalNAc $\alpha 1,3$). Alternatively, enzyme activities that convert UDP-GlcNAc to CMP-NeuNAc are introduced and expressed in a select host system. For instance, Neu enzyme activities that convert UDP-GlcNAc to CMP-NeuNAc comprise NeuB (synthase), NeuC (epimerase), and NeuA (synthase). In addition, enzyme activities required to synthesize polysialic acid and/or an acetylated form including NeuE, NeuS (polysialyltransferase), NeuD (acyltransferase family), NeuO (PSA O-acetyltransferase), and KpsCS are expressed. In certain embodiments, PSA is produced using minimal genes *neuES* and *kpsCS* to produce [$\alpha(2\rightarrow 3)$ Neu5Ac]_n; [$\alpha(2\rightarrow 6)$ Neu5Ac]_n; [$\alpha(2\rightarrow 8)$ Neu5Ac]_n [$\alpha(2\rightarrow 9)$ Neu5Ac]_n, or [$\alpha(2\rightarrow 8)$ Neu5Ac – $\alpha(2\rightarrow 9)$ Neu5Ac]_n, or a combination thereof. In yet further embodiments, the glycoprotein composition has a defined degree of polymerization from about 1 to about 500, preferably between 2 and 125 sialic acid residues.

[0032] In various other aspects of the invention, a combination of glycosyltransferase enzymes are expressed to produce glycans containing, for example, H-antigen (Fuc $\alpha 1,2$ – Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GlcNAc); T-antigen (Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GlcNAc; Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GalNAc β) and Sialyl T-antigen (Sia $\alpha 2,3$ – Gal $\beta 1,3$ – GalNAc $\alpha 1,3$ – GlcNAc).

[0033] While various host cells can be engineered to produce oligosaccharides and glycoprotein compositions, a preferred expression system involves prokaryotic host cells. Prokaryotic host cells further comprise an oligosaccharyl transferase activity (EC 2.4.1.119) capable of transferring the oligosaccharide composition onto an N-glycosylation acceptor site of the protein of interest.

[0034] In preferred aspects, the invention provides methods and host cells comprising a heterologous protein of interest. In certain embodiments, the protein of interest comprises desired oligosaccharide composition. Accordingly, the invention provides various oligosaccharide compositions produced as described herein.

[0035] In other preferred aspects, the glycoprotein compositions produced by the host cell are described herein. In more preferred aspects, the glycoproteins enhance pharmacokinetic

properties such as improved serum half-life, enhanced stability, reduced immunogenicity or non-immunogenic or illicit a desired immune response.

[0036] Still in other aspects, cell cultures comprising the host cell are provided.

Additionally, various methods for producing an oligosaccharide composition comprising culturing recombinant prokaryotic host cells are provided. Various methods for producing a glycoprotein composition comprising culturing recombinant prokaryotic host cells are also provided.

BRIEF DESCRIPTION OF THE FIGURES

[0037] **Figure 1** depicts representative biosynthetic pathways for the recombinant production of oligosaccharides containing various human and human-like glycan structures and polysialic acid (PSA).

[0038] **Figure 2** represents FACS analysis of the engineered human T antigen on the cell surface of bacteria detected by RCA (**left**), SBA (**center**) and glycosylated hGH detected by a SDS-PAGE (**right**).

[0039] **Figure 3** represents a MS of a recombinantly expressed human T antigen treated with buffer or β 1,3 galactosidase.

[0040] **Figure 4** represents Western blot analysis of recombinantly produced aglycosylated MBP8xDQNAT (pMW07) or MBP8xDQNAT carrying the T antigen glycan (pJD-07).

[0041] **Figure 5** represents IgM- or IgG- class specific ELISAs using serum derived from mice immunized with MBP8xDQNAT conjugated to a T antigen glycan or aglycosylated MBP8xDQNAT. The rectangle represents the mean.

[0042] **Figure 6** represents ELISA results using serum from mice immunized with T antigen-MBP8xDQNAT (glycosylated) or MBP8x DQNAT (aglycosylated). Wells are coated with T antigen-GFP4xDQNAT or GFP4x DQNAT.

[0043] **Figure 7** represents MS of recombinantly expressed human 2,3 sialyl T antigen on glucagon.

[0044] **Figure 8** represents MS of recombinantly expressed human 2,3 sialyl T antigen on glucagon improved by expression of *neuDBAC* on glucagon plasmid.

[0045] **Figure 9** represents MS of recombinantly expressed human sialyl T antigen on glucagon after treatment with α 2,3 neuraminidase confirming sialylation and linkage.

[0046] **Figure 10** represents MS over time of glucagon alone (**left**), or with the human 2,3 sialyl T antigen (**right**).

[0047] **Figure 11** represents MS of a recombinantly expressed 2,6 sialylated T antigen on glucagon.

[0048] **Figure 12** represents MS of a recombinantly expressed 2,6 sialylated T antigen on glucagon treated with α 2,3 neuraminidase or non-linkage specific neuraminidase.

[0049] **Figure 13** represents a dot blot of recombinant PSA expression on the cell surface of *E. coli* Δ *nanA* supplemented with NeuNAc (**a**); and the expected linkages of an exemplary glycan (**b**).

[0050] **Figure 14** represents a Western blot using the α PSA antibody in the presence of pJLic3BS-07 and NeuNAc supplementation (**top**) and total protein detected by the presence of the hexahistidine tag with α His antiserum (**bottom**).

[0051] **Figure 15** represents a dot blot highlighting the effect of *neuD* expression on cell surface PSA produced by expression of pJLic3BS-07.

[0052] **Figure 16** represents SDS PAGE and Western blot of anti-PSA (**top**) and anti-His (**bottom**) *ex vivo* polysialylation of MBP4xGT with CstII-SiaD fusion plasmid.

[0053] **Figure 17** represents a MS of a recombinantly expressed fucosylated human H antigen glycan with buffer control (**a**) or treated with α 1,2 fucosidase and MS of a recombinantly expressed fucosylated H antigen glycan with expression of GDP-fucose biosynthetic genes (**b**).

[0054] **Figure 18** represents a Western blot of TNF α Fab expressed with pJK-07 glycosylation plasmid.

[0055] **Figure 19** represents MS of recombinant fucosylated glucagon peptide with the human H antigen (**left**) and the glucagon peptide with the human H antigen and additional expression of the GDP-fucose biosynthetic genes (**right**).

[0056] **Figure 20** represents MS of recombinantly expressed fucosylated glucagon peptide treated with buffer only, or α 1,2 fucosidase confirming fucosylation and linkage

[0057] **Figure 21** represents SDS PAGE and Western blot of GH2 expressed with pJK-07 detected with an α hGH antibody (**left**), and MS of GH2 glycosylated with the H antigen (**right**).

[0058] **Figure 22** represents the increase in turbidity over time following vortexing GH2 or GH2-H antigen (**a**) and receptor binding of GH2 and GH2-H antigen.

[0059] **Figure 23** represents ELISA of GH2 or GH2-H antigen detected in rat serum over time following injection of the respective proteins.

ABBREVIATIONS AND TERMS

[0060] The following explanations 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. 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.

[0061] All publications, patents and other references mentioned herein are hereby incorporated by reference in their entireties.

[0062] EC numbers are established by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (available at <http://www.chem.qmul.ac.uk/iubmb/enzyme/>). The EC numbers referenced herein are derived from the KEGG Ligand database, maintained by the Kyoto Encyclopedia of Genes and Genomics, sponsored in part by the University of Tokyo. Unless otherwise indicated, the EC numbers are as provided in the database as of March 2013.

[0063] The accession numbers referenced herein are derived from the NCBI database (National Center for Biotechnology Information) maintained by the National Institute of Health, U.S.A. Unless otherwise indicated, the accession numbers are as provided in the database as of March 2013.

[0064] The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Taylor and Drickamer, *Introduction to Glycobiology*, Oxford Univ. Press (2003); Worthington Enzyme Manual, Worthington Biochemical Corp., Freehold, N.J.; *Handbook of Biochemistry: Section A*

Proteins, Vol I, CRC Press (1976); Handbook of Biochemistry: Section A Proteins, Vol II, CRC Press (1976); Essentials of Glycobiology, Cold Spring Harbor Laboratory Press (1999).

[0065] The term “claim” in the provisional application is synonymous with embodiments or preferred embodiments.

[0066] The term “human-like” with respect to a glycoprotein refers to proteins having attached either *N*-acetylglucosamine (GlcNAc) residue or *N*-acetylgalactosamine (GalNAc) 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.

[0067] “*N*-glycans” or “*N*-linked glycans” refer to *N*-linked saccharide 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 (Glu), galactose (Gal), mannose (Man), fucose (Fuc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), and sialic acid (e.g., *N*-acetyl-neuraminic acid (Neu5Ac, NeuAc, NeuNA, NeuNAc, Sia or NANA). Hexose (Hex) refers to mannose or galactose.

[0068] The term “blood group antigens”, “BGA” or “human antigen” are used interchangeably and comprise an oligosaccharide moiety(ies).

[0069] The term “polysialic acid”, or “PSA” refers to an oligosaccharide structure that comprises at least two NeuNAc residues.

[0070] 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.

[0071] 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.

[0072] 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.

[0073] 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.

[0074] The term “binding affinity” refers to a protein binding to a target receptor. The binding affinity of a glycosylated protein or peptide can range from about 0.01%-30%, or about 0.1% to about 20%, or about 1% to about 15%, or about 2% to about 10% of the binding affinity of the corresponding aglycosylated protein or peptide. Binding affinity of a glycosylated protein or peptide can be increased or reduced at least about 3-fold, or at least about 5-fold, or at least about 6-fold, or at least about 7-fold, or at least about 8-fold, or at least about 9-fold, or at least about 10-fold, or at least about 12-fold, or at least about 15-fold, or at least about 17-fold, or at least about 20-fold, or at least about 30-fold, or at least about 50-fold, or at least about 100-fold less binding affinity compared to the aglycosylated protein or peptide.

[0075] The term “serum persistence” as applied to proteins or peptides refers to the ability of the proteins or peptides to withstand degradation in blood or components thereof, which typically involves proteases in the serum or plasma. The serum degradation resistance can be measured by as shown in Example 20.

[0076] 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.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0077] In various aspects, the present invention provides glycoengineered host cells to recombinantly produce oligosaccharides such as BGA-conjugated or PSA-conjugated proteins in a single fermentation without the added step for *in vitro* chemical modification. Advantageously, glycoengineered host expression technology enables control of the location and stoichiometry of attached polysaccharides and eliminates the need for excess thiols and *in vitro* chemical reactions. Accordingly, in certain embodiments, the present invention provides methods and compositions for producing an oligosaccharide composition comprising: culturing a recombinant host cell to express GalNAc transferase activity (EC 2.4.1.-) (EC 2.4.1.290) and galactosyltransferase activity (EC 2.4.1.-) (EC 2.4.1.309); wherein the host cell produces an oligosaccharide composition comprising one or more GalNAc, galactose or galactose-GalNAc residues linked to a lipid carrier.

[0078] **Figure 1** provides an overview of exemplary biosynthetic mechanisms to produce either BGA-conjugated, sialic acid, or PSA-conjugated proteins in prokaryotes. In preferred embodiments, recombinant oligosaccharide synthesis is initiated by the expression of an α 1,3-N-acetylgalactosamine transferase activity (EC 2.4.1.-, EC 2.4.1.306). Additional embodiments include expression of other galactosyltransferase activity such as WbiP and CgtA to initiate recombinant oligosaccharide synthesis. Alternatively, recombinant oligosaccharide synthesis can be initiated directly on the N-linked site of the protein by expressing UDP-N-acetylglucosamine 4-epimerase activity (Rush et al (2010) JBC 285(3)

1671-1680). Yet another alternative provides bacillosamine to initiate oligosaccharide synthesis. Accordingly, the present invention provides methods for recombinant oligosaccharide synthesis on a GlcNAc residue, a GalNAc residue or bacillosamine, which can be *N*-linked onto a protein of interest.

[0079] Methods and compositions are also provided to express one or more activities selected from UndP N-acetylglucosaminyl transferase (EC 7.8.33), UndPP GalNAc epimerase activity (EC 5.1.3.c) and UndP bacillosamine transferase activity (EC 2.7.8.36).

[0080] Human T Antigen

[0081] In exemplary embodiments, the invention provides methods to recombinantly express the genetic machinery needed for the production of various BGAs. A preferred method to produce the human T antigen comprises the recombinant expression of a GalNAc transferase activity (EC 2.4.1.-) (EC 2.4.1.290) that catalyzes the transfer of a UDP-GalNAc residue onto an acceptor substrate β 1,4GlcNAc. The host cell further expresses a galactosyltransferase enzyme activity (EC 2.4.1.-) (EC 2.4.1.309), which caps the GalNAc acceptor oligosaccharide resulting in a human T antigen. **Figure 3** provide experimental support of a recombinantly produced glycoform that correlates w the structure: Gal β 1,3 – GalNAc α 1,3 –GlcNAc, the human T antigen.

[0082] Human Sialyl T Antigen

[0083] In another aspect of the invention, a method is provided to produce the human sialyl T antigen, which comprises the recombinant expression of a GalNAc transferase activity (EC 2.4.1.-) (EC 2.4.1.290), a galactosyltransferase enzyme activity (EC 2.4.1.-) (EC 2.4.1.309) and a 2,3 NeuNAc transferase activity (EC 2.4.99.4, EC 2.4.99.-, EC 2.4.99.8). **Figure 7** represents a MS of a recombinantly produced glycoform on glucagon peptide that correlates w the structure: Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 –GlcNAc;

[0084] In more preferred embodiments, an improved level of a glycoform is produced by expressing one or more of the enzyme activities selected from sialic acid biosynthesis protein, N-acetylneuraminate synthase (EC 2.5.1.56), N-acetylneuraminate cytidylyltransferase (EC 2.7.7.43), UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14) and N-acetylneuraminate acetyltransferase (EC 2.3.1.45) e.g., neuDBAC. **Figure 8** describes a recombinantly produced glycoform on glucagon peptide with improved level of the sialyl T glycoform on the glucagon peptide via ectopic or increased expression of sugar nucleotide enzyme activities. Addition of sialic acid was confirmed with the treatment of the glycosylated glucagon peptide with α 2,3 neuraminidase **Figure 9**.

[0085] In additional embodiments, α 2,6 sialyl T glycoform is produced by expression of one or more α 2,6 NeuNAc transferase (EC 2.4.99.1). A glucagon peptide comprising a linkage other than the α 2,3 linkage, e.g., α 2,6 sialyl T glycoform is shown in **Figure 11**.

[0086] Polysialic Acid

[0087] In other exemplary embodiments, the present invention provides a method for producing an oligosaccharide composition comprising: culturing a recombinant host cell to express one or more of the enzymes comprising: GalNAc transferase activity (EC 2.4.1.-) that transfers a GalNAc residue onto an acceptor substrate; galactosyltransferase enzyme activity (EC 2.4.1.-); fucosyltransferase enzyme activity (EC 2.4.1.69); and sialyltransferase enzyme activity (EC 2.4.99.4, EC 2.4.99.-, EC 2.4.99.8), wherein the host cell produces a polysialic acid.

[0088] Evidence of PSA on the cell wall is shown in **Figures 13 and 15**. The expected structural linkages of the PSA glycoforms include:

$(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc}\alpha 1,3 - \text{GalNAc}\alpha 1,3 - \text{GlcNAc}$;

$(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc}\alpha 1,3 - \text{GlcNAc}$; and

$(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - (\text{GalNAc } \alpha 1,3)_n$.

[0089] In select embodiments, the invention provides methods to recombinantly express the genetic machinery needed for the PSA production. As described in **Example 12**, the genes representing the capsular biosynthetic loci harboring the *kps* and *neu* genes of *E. coli* K1 and K92 are cloned into plasmid pACYC184 for transformation of a preferred strain of *E. coli*.

[0090] In other select embodiments, the N-linked oligosaccharide compositions comprise or consists of $[\alpha(2 \rightarrow 3)\text{Neu5Ac}]_n$; $[\alpha(2 \rightarrow 6)\text{Neu5Ac}]_n$; $[\alpha(2 \rightarrow 8)\text{Neu5Ac}]_n$; $[\alpha(2 \rightarrow 9)\text{Neu5Ac}]_n$ or a combination thereof.

[0091] Also disclosed are genes for producing the desired PSA oligosaccharide compositions. In certain embodiments, one or more Neu activity such as NeuDBACES and Kps activity such as KpsSCUDEF are expressed. In yet other embodiments, one or more genes encoding KpsMT is attenuated. The invention provides a method for producing an N-linked sialic acid on a glycoprotein comprising: culturing a host cell to produce CMP-Neu5Ac from UDP-GlcNAc; PSA from CMP-Neu5Ac; and expressing an OST activity; wherein the OST activity transfers the sialic acid onto an acceptor asparagine of the resulting glycoprotein.

[0092] Preferably the oligosaccharide structure is N-linked to a protein, comprises a terminal sialic acid residue and is more preferably a polysialic acid that is a polysaccharide

comprising at least 2 sialic acid residues joined to one another through α 2-8 or α 2-9 linkages. A suitable polysialic acid has a weight average molecular weight in the range 2 to 100 kDa, preferably in the range 1 to 35 kDa. The most preferred polysialic acid has a molecular weight in the range of 10-20kDa, typically about 14kDa.

[0093] More preferably, the N-linked PSA glycoprotein comprises about 2-125 sialic acid residues. Polymerized PSA can be transferred onto the glycoprotein, N-linked, some comprising 10-80 sialic acid residues, others 20-60 sialic acid residues, or 40-50 sialic acid residues. The preferred N-linked PSA glycoprotein composition has a defined degree of polymerization.

[0094] In additional embodiments, the glycoprotein composition further comprises a second N-linked oligosaccharide structure for example eukaryotic, human or human-like glycans such as Neu5Ac₁₋₄Gal₁₋₄GlcNAc₁₋₅Man₃GlcNAc₂, Man₃₋₅GlcNAc₁₋₂, GlcNAc₁₋₂, bacterial glycans such as GalNAc- α 1,4-GalNAc- α 1,4-[Glc β 1,3]GalNAc- α 1,4-GalNAc- α 1,4-GalNAc- α 1,3-Bac- β 1,*N*-Asn (GalNAc₅GlcBac, where Bac is bacillosamine or 2,4-diacetamido-2,4,6-trideoxyglucose). A mixture of N-linked PSA and N-linked oligosaccharide composition is also contemplated.

[0095] Glycoengineered *E. coli* have been used to attach diverse lipid-linked O-antigen glycans to corresponding asparagines in acceptor proteins *in vivo* (Feldman MF et al, (2005) Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A. 2005 Feb 22;102(8):3016-21.). Enabling control of the location and stoichiometry of attached polysaccharides such as PSA may be critically important as amine-directed chemical conjugation of PSA is random and results in an unacceptably heterogeneous product. Favorable conjugation has only recently been achieved by site-specific, chemical coupling of PSA to engineered C-terminal thiols.

[0096] The PSA-conjugated protein is expected to improved circulating half-life and provide stability. Because PSA is a natural part of the human body, the recombinant PSA composition, which is chemically and immunologically similar to human PSA and (unlike PEG) is expected to be degraded or metabolized by tissue neuraminidases or sialidases to sialic acid residues. The recombinant PSA compositions are also immunologically invisible as a biodegradable polymer.

[0097] Additional advantages of the recombinant biosynthesis are as follows. While PSA conjugation requires several intricate *in vitro* chemical reactions and multiple purifications, direct recombinant production of PSA via host cell expression obviates the need for *in vitro* chemical reactions. There is no need to isolate PSA from *E. coli* K1 capsules prior to *in vitro*

chemical crosslinking. Random attachment patterns and undesirable heterogeneity resulting from the standard amine-directed chemical conjugation of PSA is also obviated. While site-specific, thiol-directed chemical conjugation can be used, this requires the appendage of multiple C-terminal thiols and expression from a mammalian host. Capital cost and production are kept low for efficient production and processing using the glycoengineered hosts. Therefore, in one aspect of the invention, the methods and host cells serve as a glycoprotein expression system for producing *N*-linked glycoproteins with structurally homogeneous human-like glycans and overcomes many of the above limitations and challenges. The host cells address the clear clinical demand for PSA-conjugated protein therapeutics.

[0098] Human H Antigen

[0099] In further exemplary embodiments, the present invention provides a method for producing an oligosaccharide composition comprising: culturing a recombinant host cell to express one or more of the enzymes comprising: GalNAc transferase activity that catalyzes a GalNAc residue onto an acceptor substrate (EC 2.4.1.-); galactosyltransferase enzyme activity (EC 2.4.1.-); and one or more activities selected from α 1,2 fucosyltransferase (EC 2.4.1.69), α 1,3 fucosyltransferase (EC 2.4.1.152) and α 1,3/1,4 fucosyltransferase (EC 2.4.1.65). GDP-fucose transfer was confirmed with the treatment of the glycans with α 1,2-fucosidase **Figure 17A**. The recombinantly produced glycoform that correlates with the structure: α 1,2 Fuc – Gal β 1,3 – GalNAc α 1,3 – GlcNAc, the human H antigen is shown in **Figure 17B**. The human H antigen was also transferred onto a glucagon peptide by culturing the recombinant host to express a GDP-fucose biosynthetic machinery (**Example 18**). Increased production of fucosylated glycopeptide comprising the H antigen is provided in **Figure 19**. GDP-fucose transfer on glucagon was confirmed with the treatment of the glycans with α 1,2-fucosidase **Figure 20**.

[00100] **Figure 18** indicates a glycosylated TNF α Fab heavy chain with a human H antigen. Accordingly, in an exemplary embodiment, the invention provides a method for recombinant expression of TNF α Fab heavy chain comprising a human H antigen.

[00101] Prokaryotic Expression System

[00102] In preferred aspects, the invention provides a glycoprotein production system that serves as an attractive solution for circumventing the significant hurdles associated with eukaryotic cell culture systems or *in vitro* chemical conjugation. The use of bacteria as a production vehicle is expected to yield structurally homogeneous glycoproteins while at the same time dramatically lowering the cost and time associated with protein drug development

and manufacturing. 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.

[00103] Previously, the ability to attach a foreign glycan to an acceptor protein in *E. coli* has been shown (Wacker et al 2002 N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science 2002 Nov 29;298(5599):1790-3.). Also, the ability to attach foreign glycans to a recombinant protein in a site-directed, stoichiometric manner using our proprietary C-terminal GlycTag has been demonstrated (PCT/US2009/030110). Moreover, the ability to attach lipid-linked polysaccharides (e.g., poly-FucNAc) to acceptor proteins in *E. coli* have been described (Feldman 2005). Recently, Valderrama-Rincon, et. al. (Valderrama-Rincon, et. al. "An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*," Nat. Chem. Biol. AOP (2012)) disclosed a biosynthetic pathway for the biosynthesis and assembly of Man₃GlcNAc₂ on Und-PP in the cytoplasmic membrane of *E. coli*, however, to date, no studies have demonstrated the ability to recombinantly produce BGA or PSA-conjugated proteins directly from an expression platform in a simple fermentation and purification process.

[00104] Nucleic Acid Sequences

[00105] In select embodiments, the invention provides isolated nucleic acid molecules, variants thereof, expression optimized forms of the disclosed genes, and methods of improvement thereon.

[00106] In one embodiment is provided an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of glycosyltransferase gene homologs, variants and derivatives of the wild-type coding sequences. The invention provides nucleic acid molecules comprising or consisting of sequences which are structurally and functionally optimized versions of the wild-type genes. In a preferred embodiment, nucleic acid molecules and homologs, variants and derivatives comprising or consisting of sequences optimized for substrate affinity, specificity and/or substrate catalytic conversion rate, improved thermostability, activity at a different pH and/or optimized codon usage for improved expression in a host cell are provided.

[00107] In a further embodiment is provided nucleic acid molecules and homologs, variants and derivatives comprising or consisting of sequences which are variants of the glycosyltransferase genes having at least 60% identity. In a further embodiment provided

nucleic acid molecules and homologs, variants and derivatives comprising or consisting of sequences which are variants having at least 62%, 65%, 68%, 70%, 75%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 90%, 92%, 95%, 98%, 99%, 99.9% or even higher identity to the wild-type gene.

[00108] In another embodiment, the encoded polypeptides having at least 50%, preferably, at least 55%, 60%, 70%, 80%, 90% or 95%, more preferably, 98%, 99%, 99.9% or even higher identity to the wild-type gene.

[00109] Provided also are 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.

[00110] The nucleic acid molecule includes DNA molecules (e.g., linear, circular, cDNA, chromosomal DNA, double stranded or single stranded) and RNA molecules (e.g., tRNA, rRNA, mRNA) and analogs of the DNA or RNA molecules of the described herein using nucleotide analogs. 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.

[00111] The genes, as described herein, include nucleic acid molecules, for example, a polypeptide or RNA-encoding nucleic acid molecule, separated from another gene or other genes by intergenic DNA (for example, an intervening or spacer DNA which naturally flanks the gene and/or separates genes in the chromosomal DNA of the organism).

[00112] Nucleic acid molecules comprising a fragment of any one of the above-described nucleic acid sequences are also provided. These fragments preferably contain at least 20 contiguous nucleotides. More preferably the fragments of the nucleic acid sequences contain at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or even more contiguous nucleotides.

[00113] In another embodiment, an isolated glycosyltransferase gene encoding nucleic acid molecule hybridizes to all or a portion of a nucleic acid molecule having the nucleotide

sequence set forth in the sequence listings or hybridizes to all or a portion of a nucleic acid molecule having a nucleotide sequence that encodes a polypeptide having the amino acid sequence of any of amino acid sequences as set forth in the sequence listings. Such hybridization conditions are known to those skilled in the art (see, for example, *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995); *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989)). In another embodiment, an isolated nucleic acid molecule comprises a nucleotide sequence that is complementary to a *neu* or *kps* gene encoding nucleotide sequence as set forth herein.

[00114] The nucleic acid sequence fragments display utility in a variety of systems and methods. For example, the fragments may be used as probes in various hybridization techniques. Depending on the method, the target nucleic acid sequences may be either DNA or RNA. The target nucleic acid sequences may be fractionated (e.g., by gel electrophoresis) prior to the hybridization, or the hybridization may be performed on samples in situ. One of skill in the art will appreciate that nucleic acid probes of known sequence find utility in determining chromosomal structure (e.g., by Southern blotting) and in measuring gene expression (e.g., by Northern blotting). In such experiments, the sequence fragments are preferably detectably labeled, so that their specific hybridization to target sequences can be detected and optionally quantified. One of skill in the art will appreciate that the nucleic acid fragments may be used in a wide variety of blotting techniques not specifically described herein.

[00115] It should also be appreciated that the nucleic acid sequence fragments disclosed herein also find utility as probes when immobilized on microarrays. Methods for creating microarrays by deposition and fixation of nucleic acids onto support substrates are well known in the art. Reviewed in *DNA Microarrays: A Practical Approach* (Practical Approach Series), Schena (ed.), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1-60 (1999); *Microarray Biochip: Tools and Technology*, Schena (ed.), Eaton Publishing Company/ BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of which are incorporated herein by reference in their entireties. Analysis of, for example, gene expression using microarrays comprising nucleic acid sequence fragments, such as the nucleic acid sequence fragments disclosed herein, is a well-established utility for sequence fragments in the field of cell and molecular biology. Other uses for sequence fragments immobilized on microarrays are described in Gerhold *et al.*, *Trends Biochem. Sci.* 24:168-173 (1999) and Zweiger, *Trends Biotechnol.* 17:429-436 (1999); DNA Microarrays:

A Practical Approach (Practical Approach Series), Schena (ed.), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1-60 (1999); Microarray Biochip: Tools and Technology, Schena (ed.), Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376), the disclosures of each of which is incorporated herein by reference in its entirety.

[00116] As is well known in the art, enzyme activities are measured in various ways. For example, the pyrophosphorolysis of OMP may be followed spectroscopically. Grubmeyer *et al.*, *J. Biol. Chem.* 268:20299-20304 (1993). Alternatively, the activity of the enzyme is followed using chromatographic techniques, such as by high performance liquid chromatography. Chung and Sloan, *J. Chromatogr.* 371:71-81 (1986). As another alternative the activity is indirectly measured by determining the levels of product made from the enzyme activity. More modern techniques include using gas chromatography linked to mass spectrometry (Niessen, W. M. A. (2001). *Current practice of gas chromatography--mass spectrometry*. New York, N.Y: Marcel Dekker. (ISBN: 0824704738)). Additional modern techniques for identification of recombinant protein activity and products including liquid chromatography-mass spectrometry (LCMS), high performance liquid chromatography (HPLC), capillary electrophoresis, Matrix-Assisted Laser Desorption Ionization time of flight-mass spectrometry (MALDI-TOF MS), nuclear magnetic resonance (NMR), near-infrared (NIR) spectroscopy, viscometry (Knothe, G., R.O. Dunn, and M.O. Bagby. 1997. Biodiesel: The use of vegetable oils and their derivatives as alternative diesel fuels. *Am. Chem. Soc. Symp. Series* 666: 172–208), physical property-based methods, wet chemical methods, etc. are used to analyze the levels and the identity of the product produced by the organisms. Other methods and techniques may also be suitable for the measurement of enzyme activity, as would be known by one of skill in the art.

[00117] Another embodiment comprises mutant or chimeric nucleic acid molecules or genes. Typically, a mutant nucleic acid molecule or mutant gene is comprised of a nucleotide sequence that has at least one alteration including, but not limited to, a simple substitution, insertion or deletion. The polypeptide of said mutant can exhibit an activity that differs from the polypeptide encoded by the wild-type nucleic acid molecule or gene. Typically, a chimeric mutant polypeptide includes an entire domain derived from another polypeptide that is genetically engineered to be collinear with a corresponding domain. Preferably, a mutant nucleic acid molecule or mutant gene encodes a polypeptide having improved activity such as substrate affinity, substrate specificity, improved thermostability, activity at a different pH,

improved solubility, improved expression, or optimized codon usage for improved expression in a host cell.

[00118] Isolated Polypeptides

[00119] In one embodiment, polypeptides encoded by nucleic acid sequences are produced by recombinant DNA techniques and can be isolated from expression host cells by an appropriate purification scheme using standard polypeptide purification techniques. In another embodiment, polypeptides encoded by nucleic acid sequences are synthesized chemically using standard peptide synthesis techniques.

[00120] Included within the scope of the invention are glycosyltransferase polypeptides or gene products that are derived polypeptides or gene products encoded by naturally-occurring bacterial genes. Further, included within the inventive scope, are bacteria-derived polypeptides or gene products which differ from wild-type genes, including genes that have altered, inserted or deleted nucleic acids but which encode polypeptides substantially similar in structure and/or function.

[00121] For example, it is well understood that one of skill in the art can mutate (e.g., substitute) nucleic acids which, due to the degeneracy of the genetic code, encode for an identical amino acid as that encoded by the naturally-occurring gene. This may be desirable in order to improve the codon usage of a nucleic acid to be expressed in a particular organism. Moreover, it is well understood that one of skill in the art can mutate (e.g., substitute) nucleic acids which encode for conservative amino acid substitutions. It is further well understood that one of skill in the art can substitute, add or delete amino acids to a certain degree to improve upon or at least insubstantially affect the function and/or structure of a gene product (e.g., glycosyltransferase activity) as compared with a naturally-occurring gene product, each instance of which is intended to be included within the scope of the invention. For example, the glycosyltransferase activity, enzyme/substrate affinity, enzyme thermostability, and/or enzyme activity at various pHs can be unaffected or rationally altered and readily evaluated using the assays described herein.

[00122] In various aspects, isolated polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules are provided. Preferably the isolated polypeptide has preferably 50%, 60%-70%, 70%-80%, 80%-90%, 90%-95%, 95%-98%, 98.1%, 98.2%, 98.3%, 98.4%, 98.5%, 98.6%, 98.7%, 98.8%, 98.9%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or even higher identity to the sequences optimized for substrate affinity and/or substrate catalytic conversion rate.

[00123] According to other embodiments, isolated polypeptides comprising a fragment of the above-described polypeptide sequences are provided. These fragments preferably include at least 20 contiguous amino acids, more preferably at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or even more contiguous amino acids.

[00124] The polypeptides also include fusions between the above-described polypeptide sequences and heterologous polypeptides. The heterologous sequences can, for example, include sequences designed to facilitate purification, e.g. histidine tags, and/or visualization of recombinantly-expressed proteins. Other non-limiting examples of protein fusions include those that permit display of the encoded protein on the surface of a phage or a cell, alter the subcellular localization of the protein, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region.

[00125] Secretion Signal Sequences

[00126] In selected embodiments, the oligosaccharide-conjugated polypeptide is expressed with a secretion signal sequence. The secretion signal can be an amino terminal sequence that facilitates transit across a membrane. In those embodiments where the host organism is prokaryotic, secretion signal is a leader peptide domain of a protein that facilitates insertion into the membrane or transport through a membrane. The signal sequence is removed after crossing the inner membrane, and proteins may be retained in the periplasmic space.

[00127] Various secretion signals are used, for instance pelB. The predicted amino acid residue sequences of the secretion signal domain from two PelB gene product variants from *Erwinia carotova* are described in Lei et al., Nature, 331 :543-546 (1988). The leader sequence of the PelB protein has previously been used as a secretion signal for fusion proteins (Better et al., Science, 240:1041-1043 (1988); Sastry et al., Proc. Natl. Acad. Sci., USA, 86:5728-5732 (1989); and Mullinax et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990)). Amino acid residue sequences for other secretion signal polypeptide domains from *E. coli* useful in this invention include those described in Oliver, *Escherichia coli* and *Salmonella Typhimurium*, Neidhard, F. C. (ed.), American Society for Microbiology, Washington, D.C., 1 : 56-69 (1987).

[00128] Another typical secretion signal sequence is the gene III (gill) secretion signal. Gene HI encodes Pill, one of the minor capsid proteins from the filamentous phage fd (similar to Ml 3 and rl). Pill is synthesized with an 18 amino acid, amino terminal signal sequence and requires the bacterial Sec system for insertion into the membrane.

[00129] Another typical secretion signal sequence is the SRP secretion signal. SRP secretion signals have been used, for example, to improve production of fusion protein for phage

display (Steiner et al. Nat. Biotechnology, 24:823-831 (2006)). Most commonly used type II secretion signals, such as the PelB secretion signal, use the SecB pathway. Thus, secretion constructs presented herein for expression of human mAb heavy and light chains use an SRP secretion signal, namely the secretion signal of the *E. coli dsbA* gene. Other SRP secretion signals that can be used in the methods, polynucleotides and polypeptides provided herein include SfmC (chaperone), TolB (translocation protein), and TorT (respiration regulator). The sequences of these signals are known in the art.

[00130] Secretion by the *E. coli* SecB mechanism involves attachment of a nascent polypeptide first to trigger factor, TF, and then to SecB. The ScB protein then directs attachment of the completed polypeptide to the Type II secretion complex which secretes the protein into the periplasm. Without being bound by theory, it is thought that some recombinant proteins may fold into forms which secrete poorly by this mechanism. In contrast, the SRP mechanism recognizes a different set of secretion signals and directs co-translation and secretion of nascent polypeptides through the Type II secretion complex into the periplasm. This mechanism can be used to avoid problems that could occur in secretion by the SecB pathway.

[00131] It will be apparent to one of ordinary skill in the art that any suitable secretion signal sequence may be used to facilitate secretion of expressed polypeptides.

[00132] Secretion of proteins into periplasm and medium

[00133] To determine secretion of an active antibody into culture the medium, media samples collected during the expression analysis of the variousP constructs are assayed by ELISA for its antigen binding activity.

[00134] 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.

[00135] 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. Preparation of nucleic acid constructs can alternatively be prepared using homologous recombination in yeast as described by Shanks et al., AEM, 72,2, (2006).

[00136] 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.

[00137] 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 often 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, *recA* promoter, ribosomal RNA promoter, the P_R and P_L promoters of coliphage lambda and others, including but not limited, to *lacUV5*, *ompF*, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lacUV5* (*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.

[00138] 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.

[00139] Host Cells

[00140] In accordance with the present invention, the host cell may be 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*, *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.

[00141] 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 may be required to ensure optimal levels of *N*-glycan biosynthesis. For instance, the deletion of genes in the *E. coli* O 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 other reactions. To eliminate unwanted side reactions, the following are representative genes that may be deleted from the *E. coli* host strain: *wbbL*, *glcT*, *glf*, *gafT*, *wzx*, *wzy*, *waaL*, *nanA*, *wcaJ*.

[00142] 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). 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.

[00143] One aspect of the present invention is directed to a glycoprotein conjugate comprising a protein and at least one peptide comprising a D-X₁-N-X₂-T motif fused to the protein, wherein D is aspartic acid, X₁ and X₂ are any amino acid other than proline, N is asparagine, and T is threonine.

[00144] Various host cells can be used to recombinantly produce PSA. In select embodiments, host cells are genetically modified to remove the existing native glycosyltransferases and are engineered to express the glycosyltransferases of the invention for PSA production. To remove the existing glycosylation, e.g., eukaryotic host cells are engineered to express endoglycosidase or amidase that cleave between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides from N-linked glycoproteins. Since glycosylation is essential, one may not be able to entirely eliminate the native glycan. In other embodiments, sialic acid bearing glycans may be engineered in the host cell and used as substrates for polysialiation such as ST8Sia II, ST8Sia IV, or NeuS to transfer multiple α 2-8 sialic acids to acceptor N-glycans.

[00145] In preferred aspects, the invention provides methods for recombinant production of various glycoproteins *in vivo*. In one embodiment, PSA-conjugated glucagon peptide is produced in glycoengineered *E. coli*. Using a glycosylation tag (GlycTag)

[PCT/US2009/030110], glucagon peptide from glycoengineered *E. coli* harboring the PSA genetic machinery is expressed and purified. Conjugation of PSA is confirmed by Western blot analysis using commercially available anti-PSA antibodies.

[00146] Alternative Expression Systems

[00147] Use of eukaryotic expression systems such as mammalian, yeast, fungi, plant or insect cells can be employed to produce PSA-conjugated proteins. In these embodiments, native glycosylation pathways may be disrupted in order to reduce interference with the engineered glycan pathway.

[00148] Production of PSA using yeast or fungal systems

[00149] Expression of a sialyltransferase has been demonstrated in *P. pastoris* (Hamilton, et al, "Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins", Science, vol. 313, pp. 1441-1443 (2006)). By amplifying the *E. coli neuA*, *neuB* and *neuC* genes, a pool of CMP-sialic acid was shown to accumulate in yeast. Yeast or other fungal

systems are suitable expression hosts to express the various glycosyltransferases for the production of human antigens or PSA.

[00150] Expressing PSA operon in plant cell, e.g., tobacco, lemna or algae

[00151] As described in the U.S. Pat. No. 6,040,498, lemna (duckweed) can be transformed using both agrobacterium and ballistic methods. Using protocols described, lemna is transformed and the resulting oligosaccharide composition is transferred onto a target protein. Transgenic plants can be assayed for those that produce proteins with desired human antigens or PSA residues according to known screening techniques.

[00152] Production of PSA using insect cell systems

[00153] The present invention can also be applied to the metabolically transformed cell lines derived from Sf9 cells. Sf9 has been used as a production host for recombinant proteins such as interferons, IL-2, plasminogen activators among others, based on its relative ease at which proteins are cloned, expressed and purified in comparison to mammalian cells. Sf9 more readily accepts foreign genes coding for recombinant proteins than many vertebrate animal cells because it is very receptive to viral infection and replication [Bishop, D. H. L. and Possee, R. D., Adv. Gene Technol., 1, 55, (1990)]. Expression levels of recombinant proteins are extremely high in Sf9 and can approach 500 mg/liter [Webb, N. R. and Summers, M. D., Technique, 2, 173 (1990)]. The cell line performs a number of key post-translational modifications; however, they are not identical to those in vertebrates and, therefore, may alter protein function [Fraser, M. J., In Vitro Cell. Dev. Biol., 25, 225 (1989)]. Despite this, the majority of recombinant proteins that undergo post-translational modification in insect cells are immunologically and functionally similar to their native counterparts [Fraser, M. J., In Vitro Cell. Dev. Biol., 25, 225 (1989)]. In contrast to animal cell culture, Sf9 facilitates protein purification by expressing relatively low levels of proteases and having a high ratio of recombinant to native protein expression [Goswami, B. B. and Glazer, R. O. BioTechniques, 10, 626 (1991)].

[00154] Baculoviruses serve as expression systems for the production of recombinant proteins in insect cells. These viruses are pathogenic towards specific species of insects, causing cell lysis [Webb, N. R. and Summers, M. D., Technique, 2, 173 (1990)].

[00155] Recombinant protein expression in insect cells is achieved by viral infection or stable transformation. For the former, the desired gene is cloned into baculovirus at the site of the wild-type polyhedron gene [Webb, N. R. and Summers, M. D., Technique, 2, 173 (1990); Bishop, D. H. L. and Possee, R. D., Adv. Gene Technol., 1, 55, (1990)]. The polyhedron gene is nonessential for infection or replication of baculovirus. It is the principle component of a

protein coat in occlusions which encapsulate virus particles. When a deletion or insertion is made in the polyhedron gene, occlusions fail to form. Occlusion negative viruses produce distinct morphological differences from the wild-type virus. These differences enable a researcher to identify and purify a recombinant virus. In baculovirus, the cloned gene is under the control of the polyhedron promoter, a strong promoter which is responsible for the high expression levels of recombinant protein that characterize this system. Expression of recombinant protein typically begins within 24 hours after viral infection and terminates after 72 hours when the Sf9 culture has lysed.

[00156] Stably-transformed insect cells provide an alternate expression system for recombinant protein production [Jarvis, D. L., Fleming, J.-A. G. W., Kovacs, G. R., Summers, M. D., and Guarino, L. A., *Biotechnology*, 8, 950 (1990); Cavegn, C., Young, J., Bertrand, M., and Bernard, A. R., in *Animal Cell Technology: Products of Today, Prospects for Tomorrow*, Spier, R. E., Griffiths, J. B., and Berthold, W., Eds. (Butterworth-Heinemann, Oxford, 1994, pp. 43-49)]. In these cells, the desired gene is expressed continuously in the absence of viral infection. Stable transformation is favored over viral infection when recombinant protein production requires cellular processes that are compromised by the baculovirus. This occurs, for example, in the secretion of recombinant human tissue plasminogen activator from Sf9 cells [Jarvis, D. L., Fleming, J.-A. G. W., Kovacs, G. R., Summers, M. D., and Guarino, L. A., *Biotechnology*, 8, 950 (1990)]. Viral infection is favored when the recombinant protein is cytotoxic since protein expression is transient in this system.

[00157] Insect cells for in vitro cultivation have been produced and several cell lines are commercially available. This process includes using insect cells capable of culture as described herein regardless of the source. The preferred cell line is Lepidoptera Sf9 cells. Other cell lines include Drosophila cells from the European Collection of Animal Cell Cultures (Salisbury, UK) or cabbage looper *Trichoplusia ni* cells including High Five available from Invitrogen Corp. (San Diego, Calif.) Sf9 insect cells from either Invitrogen Corporation or American Type Culture Collection (Rockville, Md.) are the preferred cell line and were cultivated in the bioreactor freely suspended in serum-free EX-CELL 401 Medium purchased from JRH Biosciences (Lenexa, Kans.) and maintained at 27° C.

[00158] Oligosaccharide Compositions

[00159] The prokaryotic system can yield homogenous glycans at a relatively high yield. In preferred embodiments, the oligosaccharide composition comprises or consists essentially of a single glycoform in at least 50, 60, 70, 80, 90, 95, 99 mole %. In further embodiments, the

oligosaccharide composition consists essentially of two desired glycoforms of at least 50, 60, 70, 80, 90, 95, 99 mole %. In yet further embodiments, the oligosaccharide composition consists essentially of three desired glycoforms of at least 50, 60, 70, 80, 90, 95, 99 mole %. The present invention, therefore, provides stereospecific biosynthesis of a vast array of novel oligosaccharide compositions and *N*-linked glycoproteins including glycans for BGA and PSA. Methods for estimating glycan or glycoprotein homogeneity and yield may include Mass Spectrometry, NMR, Lectin blotting, fluorophore-assisted carbohydrate electrophoresis (FACE), or chromatography methods [16-18].

[00160] Select PSA oligosaccharide compositions include:

[00161] $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$; $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$; $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - (\text{GalNAc } \alpha 1,3)_n - \text{GlcNAc}$.

[00162] Select Sialyl T Antigen oligosaccharide compositions include:

[00163] $\text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$; $\text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc } \alpha 1,3$; $\text{Sia } \alpha 2,6 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$.

[00164] Select H Antigen oligosaccharide compositions include:

[00165] $\text{Fuc } \alpha 1,2 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$; $\text{Fuc } \alpha 1,2 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc}$.

[00166] Select T Antigen oligosaccharide compositions include:

[00167] $\text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}$; and $\text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc } \alpha 1,3$.

[00168] Other select PSA oligosaccharide compositions include:

[00169] $[\beta \text{GlcNAc}] [\alpha \text{GalNAc}] [\beta \text{GalNAc}] \text{Gal}[\beta 1,3] [\alpha (2 \rightarrow 3) \text{Neu5Ac}]_n$; $[\alpha (2 \rightarrow 6) \text{Neu5Ac}]_n$; $[\alpha (2 \rightarrow 8) \text{Neu5Ac}]_n$; $[\alpha (2 \rightarrow 8) \text{Neu5Ac} - \alpha (2 \rightarrow 9) \text{Neu5Ac}]$ or $[\alpha (2 \rightarrow 9) \text{Neu5Ac}]_n$.

[00170] Various oligosaccharide compositions produced using the methods and compositions of the invention include but are not limited to the following:

[00171] $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}\beta 1-$;

[00172] $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc}\beta 1-$;

[00173] $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc}\alpha 1-$;

[00174] $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{Bac}\alpha 1-$;

[00175] $\text{Sia}\alpha 2,8 - \text{Sia}\alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc}\alpha 1,3 - \text{GlcNAc}\beta 1-$;

[00176] $\text{Sia}\alpha 2,8 - \text{Sia}\alpha 2,3 - \text{Gal}\beta 1,3 - \text{GalNAc}\alpha 1,3 - \text{GalNAc}\alpha 1-$;

- [00177] Sia α 2,8 – Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00178] Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00179] Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – GalNAc1-;
- [00180] Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00181] Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00182] Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- [00183] Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00184] Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00185] Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- [00186] Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00187] Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00188] Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- [00189] Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00190] GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00191] GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- [00192] GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- [00193] Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – GlcNAc β 1-;
- [00194] Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – GalNAc α 1-;
- [00195] Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – Bac α 1-;
- [00196] Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- [00197] Gal β 1,3 – GalNAc α 1,3 – Bac α 1-; and
- [00198] Gal β 1,3 – GalNAc α 1,3 – GalNAc 1-.

[00199] Target Glycoproteins

[00200] Various examples of suitable target glycoproteins may be produced according to the invention, which include without limitation: 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, glucagon, glucagon peptides, GLP-1 w/ and w/o FC (glucagon like protein 1), GLP-1

receptor agonist e.g., exenatide, direct thrombin inhibitor e.g., bivalirudin, IGF-1 e.g., mecasermin, parathyroid hormone e.g., teriparatide, plasma kallikrein inhibitor e.g., ecallantide, IL-1 receptor agonist, sTNFr (aka soluble TNF receptor Fc fusion), CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4-Ig), receptors, hormones such as human growth hormone, erythropoietin, peptides, stapled peptides, human vaccines, animal vaccines, serum albumin and enzymes such as ATIII, rhThrombin, glucocerebrosidase and asparaginase.

[00201] Antibodies, fragments thereof and more specifically, the Fab regions such as adalimumab, atorolimumab, fresolimumab, golimumab, lerdelimumab, metelimumab, morolimumab, sifalimumab, ipilimumab, tremelimumab, bertilimumab, briakinumab, canakinumab, fezakinumab, ustekinumab, adecatumumab, belimumab, cixutumumab, conatumumab, figitumumab, intetumumab, iratumumab, lexatumumab, lucatumumab, mapatumumab, necitumumab, ofatumamb, panitumumab, pritumumab, rilotumumab, robatumumab, votumumab, zalutumumab, zanolimumab, denosumab, stamulumab, efungumab, exbivirumab, foravirumab, libivirumab, rafivirumab, regavirumab, sevirumab, tuvirumab, nebacumab, panobacumab, raxibacumab, ramucirumab, gantenerumab.

[00202] Full-length monoclonal antibodies have traditionally been produced in mammalian cell culture due to their parental hybridoma source, the complexity of the molecule, and the desirability of glycosylation of the monoclonal antibodies. Generally, *Escherichia coli* is the host system of choice for the expression of antibody fragments such as Fv, scFv, Fab or F(ab')₂. These fragments can be made relatively quickly in large quantities with the retention of antigen binding activity. However, because antibody fragments lack the Fc domain, they do not bind the FcRn receptor and are cleared quickly. Full-length antibody chains can also be expressed in *E. coli* as insoluble aggregates and then refolded in vitro, but the complexity of this method limits its usefulness. Accordingly, the antibodies are produced in the periplasm.

[00203] In contrast to the widespread uses of bacterial systems for expressing antibody fragments, there have been few attempts to express and recover at high yield functional intact antibodies in *E. coli*. Because of the complex features and large size of an intact antibody, it is often difficult to achieve proper folding and assembly of the expressed light and heavy chain polypeptides, which results in poor yield of reconstituted tetrameric antibody. Furthermore, antibodies made in prokaryotes are not glycosylated. Since glycosylation is required for Fc receptor mediated activity, it is conventionally considered that *E. coli* would not be a useful system for making intact antibodies. (Pluckthun and Pack (1997) Immunotech

3:83-105; Kipriyanov and Little (1999) *Mol. Biotech.* 12:173-201.). Recombinant oligosaccharide synthesis changes this paradigm.

[00204] Recent developments in research and clinical studies suggest that in many instances, intact antibodies are preferred over antibody fragments. An intact antibody containing the Fc region tends to be more resistant to degradation and clearance *in vivo*, thereby having longer biological half life in circulation. This feature is particularly desirable where the antibody is used as a therapeutic agent for diseases requiring sustained therapies.

[00205] Currently, anti-TNF antibodies are produced in mammalian cells and are glycosylated. The cost of producing antibodies in mammalian cells (frequently in CHO cells) is high and the procedure is complex. Glycosylation of antibodies has two effects: first, it can increase the lifetime of the antibody in the blood serum, so that it circulates for many days or even weeks. This may be because of decreased kidney clearance or because of greater resistance to proteolysis. Second, as provided herein, glycosylation in the constant region of the antibody is important for activating the “effector functions” of the antibody, which are triggered when an antibody binds to a target that is attached to a cell surface. These functions are linked to activation of the immune system and can lead to natural killer (NK) mediated cell killing.

[00206] The present invention relates in part to glycoprotein compositions comprising peptides characterized as having enhanced pharmacokinetic properties such as improved serum half-life, enhanced stability, reduced immunogenicity or non-immunogenic or illicit a desired immune response. Example 19 provides recombinantly expressed human growth hormone placental variant (GH2) comprising a H antigen. **Figure 21** represents a mass that correlates to the GH2 glycosylated with the H antigen. Stability and binding were measured as shown in **Figure 22**. In certain embodiments, the glycoprotein composition is configured to have reduced or increased binding affinity for a target receptor of the corresponding peptide as compared to the aglycosylated peptide.

[00207] The invention further provides novel peptides characterized as having increased serum persistence as more fully described in Example 20. The *in-vivo* half-life assay in rat model provides evidence of increased serum persistence of GH2 comprising a H antigen as compared to the aglycosylated GH2 as evidenced in **Figure 23**. Accordingly, the present invention in part demonstrates that the glycoproteins comprise enhanced pharmacokinetic properties such as improved serum half-life, enhanced stability, reduced immunogenicity or non-immunogenic or illicit a desired immune response.

[00208] Pharmaceutical Compositions and Pharmaceutical Administration

[00209] Another aspect of the invention is a composition as defined above which is a pharmaceutical composition and further comprises one or more pharmaceutically acceptable excipients. The pharmaceutical composition may be in the form of an aqueous suspension. Aqueous suspensions contain the novel compounds in admixture with excipients suitable for the manufacture of aqueous suspensions. The pharmaceutical compositions may be in the form of a sterile injectable aqueous or homogeneous suspension. This suspension may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents.

[00210] Pharmaceutical compositions may be administered orally, intravenously, intraperitoneally, intramuscularly, subcutaneously, intranasally, intradermal, topically or intratracheal for human or veterinary use.

[00211] The protein, peptide, antibody and antibody-portions of the invention can be incorporated into pharmaceutical compositions suitable for administration to a subject. Typically, the pharmaceutical composition comprises an antibody or antibody portion of the invention and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible. Examples of pharmaceutically acceptable carriers include one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Pharmaceutically acceptable substances or minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the protein, peptide, antibody or antibody portion.

[00212] The compositions of this invention may be in a variety of forms. These include, for example, liquid, semi-solid and solid dosage forms, such as liquid solutions (e.g., injectable and infusible solutions), dispersions or suspensions, tablets, pills, powders, liposomes and suppositories. The preferred form depends on the intended mode of administration and therapeutic application. Typical preferred compositions are in the form of injectable or infusible solutions, such as compositions similar to those used for passive immunization of humans with other antibodies. The preferred mode of administration is parenteral (e.g., intravenous, subcutaneous, intraperitoneal, intramuscular). In a preferred embodiment, the antibody is administered by intravenous infusion or injection. In another preferred embodiment, the antibody is administered by intramuscular or subcutaneous injection.

[00213] Therapeutic compositions typically must be sterile and stable under the conditions of manufacture and storage. The composition can be formulated as a solution, microemulsion, dispersion, liposome, or other ordered structure suitable to high drug concentration. Sterile injectable solutions can be prepared by incorporating the active compound (i.e., protein, peptide, antibody or antibody portion) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof. The proper fluidity of a solution can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prolonged absorption of injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[00214] The protein, peptide, antibody and antibody-portions of the present invention can be administered by a variety of methods known in the art, although for many therapeutic applications, the preferred route/mode of administration is intravenous injection or infusion. As will be appreciated by the skilled artisan, the route and/or mode of administration will vary depending upon the desired results. In certain embodiments, the active compound may be prepared with a carrier that will protect the compound against rapid release, such as a controlled release formulation, including implants, transdermal patches, and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Many methods for the preparation of such formulations are patented or generally known to those skilled in the art. See, e.g., Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

[00215] In certain embodiments, an antibody or antibody portion of the invention may be orally administered, for example, with an inert diluent or an assimilable edible carrier. The compound (and other ingredients, if desired) may also be enclosed in a hard or soft shell gelatin capsule, compressed into tablets, or incorporated directly into the subject's diet. For oral therapeutic administration, the compounds may be incorporated with excipients and used

in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. To administer a compound of the invention by other than parenteral administration, it may be necessary to coat the compound with, or co-administer the compound with, a material to prevent its inactivation.

[00216] 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.

EXAMPLE 1

[00217] Plasmid Construction

[00218] Plasmids in this study were constructed using standard homologous recombination in yeast (Shanks RM, Caiazza NC, Hinsa SM, Toutain CM, O'Toole GA: **Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria**. *Appl Environ Microbiol* 2006, **72**(7):5027-5036.)). Plasmids were recovered from yeast and transferred to *E. coli* strain DH5 α for confirmation via PCR and/or sequencing. The following list describes plasmids constructed during the course of this study. The plasmid name is followed by the inserted genes/sequences in order from 5'-3' followed by the vector in parentheses. Glycan expression plasmids were constructed in vector pMW07 (Vaderrama-Rincon et al.). Protein expression plasmids were typically constructed in vector pTRCY. Sugar nucleotide synthesis plasmids were cloned in pTrcY, pMQ70.

[00219] In order of figures:

[00220] pMW07: (vector) pBAD, ChlorR, *ura3*, CEN ORI [19]

[00221] pDis-07: *galE*, *pglB*, *pglA* (pMW07)

[00222] pDisJ-07: *galE*, *pglB*, *pglA*, *wbnJ* (pMW07)

[00223] pTrcY: (vector) pTRC, AmpR, pBR322 ORI, 2 μ

[00224] pMBP-hGH-Y: *ssdsbA-malE* (no signal sequence)-*hexahistidine*-TEV-*hGH* (pTrcY)

[00225] pTrc-spMBP-GT-MBP-GT: *ssmalE*-4*xdqnat-malE*-4*xdqnat-hexahistidine* (pTrc99a) [20]

[00226] pDisJD-07: *galE*, *pglB*, *pglA*, *neuD*, *neuB*, *neuA*, *neuC*, *wbnJ* (pMW07)

[00227] pTrc-spTorA-GFP-GT: *sstorA-gfp*-4*xdqnat-hexahistidine* (pTrc99a). [20]

- [00228] pJDLST-07: *galE, pglB, pglA, neuD, neuB, neuA, neuC, lst, wbnJ* (pMW07)
- [00229] pMG4X-Y: *ssdsbA-malE-3XTEV-glucagon-4X dqnat hexahistidine* (pTrcY)
- [00230] pMG1X-Y: *ssdsbA-malE-3XTEV-glucagon-1X dqnat-hexahistidine* (pTrcY)
- [00231] pMG1X D-Y: *ssdsbA-malE-3xTEV-glucagon-1X dqnat-hexahistidine, neuDBAC* (pTrcY)
- [00232] pJDPdST6-07: *galE, pglB, pglA, neuD, neuB, neuA, neuC, Pdst6* (pMW07)
- [00233] pJCstIIS-07: *galE, pglB, pglA, neuS, neuB, neuA, neuC, cstII260, wbnJ* (pMW07)
- [00234] pJLic3BS-07: *galE, pglB, pglA, neuS, neuB, neuA, neuC, lic3B, wbnJ* (pMW07)
- [00235] pNeuD-Y: *neuD* (pTrcY)
- [00236] pMBP4X-Y: *ssdsbA-malE-4X GlycTag-hexahistidine* (pTrcY)
- [00237] pCstII*SiaD-Y: *cstIII53S260-siaD* (pTrcY)
- [00238] pCstIISiaD-Y: *cstII260-siaD* (pTrcY)
- [00239] pJK-07: *galE, pglB, pglA, wbnJK* (pMW07)
- [00240] pGNF-70: *galE(Cj), galE(K12), gmd, fcl, gmm, cpsBG* (pMQ70)
- [00241] pTnfαFab4X-Y: *tnfα light chain, tnfα heavy chain-4X dqnat-hexahistidine* (pTrcY)
- [00242] pMG1X GNF-Y: *ssdsbA-malE-3XTEV-glucagon-1X dqnat-hexahistidine, galE (Cj), galE(K12), wbnK, gmd, fcl, gmm, cpsBG* (pTrcY)
- [00243] pMG1x KGF-Y: *ssdsbA-malE-3XTEV-glucagon-1X dqnat-hexahistidine, galE(Ec), wbnK, gmd, fcl, gmm, cpsBG* (pTrcY)
- [00244] pG4-His-GNF-Y (*ssdsbA-malE-1X TEV-hGHv-hexahistidine, galE Cj, galE Ec, gmd, fcl, gmm, cpsBG*) (pTrcY)
- [00245] **Strains** (in order of figures)
- [00246] MC4100
- [00247] MC4100 $\Delta waaL$
- [00248] MC4100 $\Delta waaL \Delta nanA$
- [00249] MC4100 $\Delta nanA$
- [00250] LPS1 $\Delta waaL$
- [00251] LPS1
- [00252] *E. coli* MC4100 was selected as a host for functional testing because it does not natively express glycan structures containing sialic acid and it has served as a functional host for glycosylation previously (Vaderrama-Rincon et al. "An engineered eukaryotic protein glycosylation pathway in *E. coli*," *Nat Chem Bio* 8, 434–436 (2012)). The mutations in the *waaL*, and *nanA* genes were transduced from the corresponding mutant in the Keio collection. The *kan* cassette was later removed from the MC4100 $\Delta nanA$ strain. For surface

expression of glycans, plasmids of interest were used to transform MC4100, MC4100 Δ *nanA*, or MC4100 Δ *nanA* Δ *waaL*. Protein glycosylation experiments were performed in strains as indicated.

[00253] Media and Reagents

[00254] Antibiotic selection was maintained at: 100 μ g/mL ampicillin (Amp), 25 μ g/mL chloramphenicol (Chlor), 10 μ g/mL tetracycline (Tet) and 50 μ g/mL kanamycin (Kan). Routine growth of *E. coli* cultures was performed in LB medium supplemented with glucose at 0.2% (w/v) and antibiotics as necessary. For expression of PSA plasmids, LB medium was supplemented with sialic acid (Sigma or Millipore) at a final concentration of 0.25% (w/v) and the medium was adjusted to pH ~7.5 and sterilized. Plasmids for glycan and protein expression were induced with the addition of L-arabinose at 0.2% or isopropyl β -D-thiogalactoside (IPTG) at 100 mM respectively. Yeast FY834 was maintained on YPD medium and synthetic defined -Uracil medium was used to select or maintain yeast plasmids.

[00255] Cell-surface glycan detection

[00256] Dot blots were performed using 2.5 μ l or 4 μ l of overnight LB culture from strain indicated. Cells were spotted on a nitrocellulose membrane and PSA glycans were detected by immunoblot as below. For flow cytometry cultures were inoculated in LB supplemented with antibiotics as appropriate.. Analysis was performed using lectins as indicated and a BD FACScalibur flow cytometer.

[00257] Protein expression and purification

[00258] Strains to be harvested for analysis of *N*-glycosylation were inoculated into LB with the appropriate antibiotics and incubated with shaking at 30°C until the cultures reached an OD₆₀₀ of 1.5-2. Plasmids for glycan expression were induced with the addition of arabinose and production of the acceptor protein was induced with IPTG. Cultures were harvested 16-18h post induction. Cell lysis and purification of glycoproteins was performed using the Ni-NTA kit (Qiagen) for small scale cultures (50-100mL). Larger preparations were in binding buffer (50mM Tris, 30mM Na₂HPO₄, 30mM Imidazole, 500mM NaCl pH=7.4) purified using HisTrap FF column (GE Healthcare) followed by elution with binding buffer containing a final concentration of 500mM imidazole. Purification over a DEAE HiTrap FF column (GE Healthcare) typically followed using 20mM Tris pH 6.8 as the binding buffer and elution with a gradient of 0-500mM NaCl in the same buffer. For purification of glycoprotein containing the T antigen glycan, protein was exchanged to 10 mM HEPES pH 7.5, 0.15 M NaCl, 0.1 mM CaCl₂, 0.01 mM MnCl₂ and further separated using Peanut agglutinin (PNA)-agarose (Vector labs). Galactose was used to isolate glycoprotein.

[00259] Protein analysis

[00260] Proteins were separated by SDS-polyacrylamide gels (Lonza), and Western blotting was performed as described previously (DeLisa MP, et al., Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc Natl Acad Sci U S A* 2003, **100**(10):6115-6120.). Briefly, proteins were transferred onto polyvinylidene fluoride (PVDF) membranes and membranes were probed with one of the following: anti-6x-His antibodies conjugated to HRP (Sigma), anti-PSA-NCAM (Millipore), or PNA-Biotin (Vector Labs). In the case of the anti-PSA antiserum, anti-mouse IgG-HRP (Promega) was used as the secondary antibody. For PNA-Biotin, Streptavidin-HRP (Vector Labs) was used for secondary detection.

EXAMPLE 2**[00261] Engineering *E. coli* for expression of the human Thomsen-Friedenreich antigen (T-antigen)**

[00262] The T antigen glycan (T-antigen, Gal β 1,3 GalNAc α -) is a structure found at the core of many human related human glycans. In order to assemble a glycan containing the human T antigen in *E. coli*, a plasmid was constructed for expression of the glycosyltransferase and sugar nucleotide epimerase activities necessary to produce this structure using the native UndPP-GlcNAc as a substrate. Plasmid pMW07 (Valderrama-Rincon et al.) was used as the vector because it contains a low copy number origin of replication (ORI), an inducible pBAD promoter, and a yeast ORI allowing for cloning via homologous recombination in *Saccharomyces cerevisiae*. The sequence of pMW07 is provided as SEQ ID NO: 1.

[00263] To generate a disaccharide glycan with the structure GalNAc α 1,3 GlcNAc, a plasmid was constructed to express the *C. jejuni* GalNAc transferase PglA, and the epimerase GalE to promote synthesis of the UDP-GalNAc substrate. The gene encoding the OST PglB from *C. jejuni* was also included for use in glycosylation in the future. A PCR fragment including *galE*, *pglB*, and *pglA* along with linearized pMW07 was used to co-transform *S. cerevisiae* and cloning was performed by homologous recombination in yeast as previously described (Shanks et al.). The sequences of these genes are provided as SEQ ID NOs: 2, 3 and 4 respectively. Plasmid was isolated from colonies selected on synthetic defined –uracil medium and used to transform *E. coli* DH5a for confirmation of construct. The resulting plasmid was designated pDis-07.

[00264] The human Thomsen-Friedenreich or T-antigen glycan consists of Gal β 1-3GalNAc α structure. Galactose transferase WbnJ from *E. coli* 086 was selected as the glycosyltransferase to incorporate the terminal galactose residue because it is reported to attach galactose in a β 1,3 linkage to a GalNAc residue and is a native bacterial enzyme (Yi W, Shao J, Zhu L, Li M, Singh M, Lu Y, Lin S, Li H, Ryu K, Shen J *et al*: **Escherichia coli O86 O-Antigen Biosynthetic Gene Cluster and Stepwise Enzymatic Synthesis of Human Blood Group B Antigen Tetrasaccharide**. *Journal of the American Chemical Society* 2005, 127(7):2040-2041.). The *wbnJ* gene was amplified from a synthetic plasmid from Mr. Gene and homologous recombination in yeast was used to combine the resulting PCR product and linearized pDis-07 plasmid. The resulting plasmid is named pDisJ-07 and contains the following genes as a synthetic operon under control of a pBAD promoter: (5'-3') *galE*, *pglB*, *pglA*, *wbnJ*. The sequence of *wbnJ* is included as SEQ ID NO: 5.

[00265] In their native context, the substrates for both glycosyltransferases PglA and WbnJ are saccharides assembled on the lipid undecaprenylpyrophosphate (UndPP). As part of the *E. coli* K12 LPS synthesis pathway, a GlcNAc residue is first added to UndPP via the activity of native WecA and the resulting GlcNAc is then transferred to the lipid A core oligosaccharide in the periplasm by the WaaL ligase. Finally, the lipid A moiety is transported to the outer membrane resulting in cell-surface display of the glycans. Cells carrying deletions in the *waaL* gene are unable to transport UndPP-linked glycans to the cell surface and thus, this mutation is useful for confirming that a glycan is linked to UndPP.

[00266] The *waaL* (*rfaL*) gene has been previously mutated as part of the Keio collection and the resulting strain *rfaL*734(del)::kan (JW3597-1) (Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection**. *Mol Syst Biol* 2006, 2.) was obtained from the Yale Coli Genetic Stock Center (CGSC). P1 *vir* phage was used to transduce the *waaL* mutation into an MC4100 recipient to make strain MC4100 Δ *waaL*::kan. Plasmid pCP20 was used to then remove the *kan* cassette (Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products**. *Proceedings of the National Academy of Sciences* 2000, 97(12):6640-6645.) resulting in strain MC4100 Δ *waaL*.

[00267] Flow cytometry was used to analyze the cell surface glycans produced by *E. coli* MC4100 expressing pDisJ-07 to confirm the presence of a galactose-terminal structure compared to control plasmid pDis-07. Cultures were inoculated in 1.5mL tubes containing 1000 μ L LB supplemented with 25 μ g/ml chloramphenicol and 0.2% arabinose (v/v). After a

24 hour incubation shaking at 30 °C, the cultures were pelleted and resuspended in 200 µl PBS. 100 µl aliquots of each were heated at 95 °C for 10 minutes and cooled to room temperature. 400 µl PBS was added to each sample and 3 µl of fluorescein labeled Soy Bean Agglutinin (SBA, Vector laboratories) or Ricinus Communis Agglutinin I (RCA I, vector laboratories) which preferentially binds to galactose terminal glycans. Samples were incubated on a rocking platform at room temperature for 10 minutes in the dark prior to flow cytometry.

[00268] Flow cytometry with the RCA I lectin was suggests the presence of a galactose terminal glycan on the cell surface of MC4100 cells expressing pDisJ-07 but not pDis-07 (**Figure 2, left**). This result is consistent with the previously reported function of the WbnJ enzyme as a galactosyl transferase. Cell-surface labeling with SBA-fluorescein (**Figure 2, center**) was reduced in cells expressing the pDisJ-07 plasmid compared to the pDis-07 plasmid suggesting a reduction in the amount of available terminal GalNAc residues. In a MC4100 $\Delta waaL$ mutant, fluorescence was greatly reduced for cells expressing either plasmid suggesting that these are both synthesized as UndPP-linked glycans.

EXAMPLE 3

[00269] *In vivo* synthesis of proteins carrying an *N*-glycan terminating in the human T antigen

[00270] The OST PglB is utilized to transfer UndPP-linked oligosaccharides to specific asparagine residues. This requires a target protein bearing the PglB recognition site consisting of the D/E X₁ N X₂ S/T sequon to be localized to the periplasm and the presence of an appropriate glycan substrate. For this study, we also constructed vector pTRCY for use in expression of glycoproteins.

[00271] pTRCY was cloned via homologous recombination in *S. cerevisiae* by adding the URA3 gene and the yeast 2 micron ORI to pTRC99a thus generating a novel vector capable of replicating in yeast. The URA3 gene and 2micron ORI were amplified with primers containing homology to vector pTRC99a for insertion between the pBR322 ORI and *lacI* gene. The sequence of vector pTRCY is provided as SEQ ID NO: 6.

[00272] hGH was cloned as a c-terminal translational fusion following a signal peptide from *E. coli* DsbA, MBP, hexahistidine tag, and a tev cleavage site. The hGH gene was further modified to contain a single glycosylation acceptor site DQNAT and the final construct is named pMBP-hGH-Y. The sequence of the gene fusion is supplied as SEQ ID NO: 7.

[00273] Strains MC4100 $\Delta nanA\Delta waaL$ bearing plasmids pDisJ-07 and pMBP-hGH-Y or pMBP-hGH-Y alone were grown under ampicillin (100 $\mu\text{g}/\text{ml}$) and chloramphenicol (25 $\mu\text{g}/\text{ml}$) or ampicillin (100 $\mu\text{g}/\text{ml}$) selection respectively. pDisJ-07 is induced with the addition of 0.2% (v/v) arabinose and IPTG (0.1mM) after approximately 16h to induce protein production. The protein was partially purified by nickel affinity chromatography and treated with TEV protease (Sigma) to release hGH prior to analysis by SDS-PAGE and Coomassie staining. The visible mobility shift in the presence of the pDisJ-07 plasmid is consistent with glycosylation (**Figure 2, right**).

EXAMPLE 4

[00274] **Confirm identity and linkage of the galactose residue in the human T antigen**

[00275] To further probe the identity of the glycan produced upon expression of pDisJ-07, we extracted the lipid-linked oligosaccharides and analyzed the released glycans by mass spectrometry. A 1:100 inoculum was used to seed 4 250mL cultures containing LB supplemented with 25 $\mu\text{g}/\text{ml}$ chloramphenicol. Cultures were grown at 30 °C and induced when the ABS_{600} reached ~ 2.0 . Cells were harvested after ~ 20 hours for isolation of lipid-linked oligosaccharides by the method of Gao and Lehrman (Gao N, Lehrman M: **Non-radioactive analysis of lipid-linked oligosaccharide compositions by fluorophore-assisted carbohydrate electrophoresis**. *Methods Enzymol* 2006, **415**:3-20.). Briefly, pellet was resuspended in 10 mL methanol and lysed by sonication. Material was dried at 60 °C and subsequently resuspended in 1mL 2:1 chloroform: methanol (v/v, CM) via sonication and material was washed two times in CM. The pellet was then washed in water then lipids were extracted with 10:10:3 chloroform: methanol: water (v/v/v, CMW) followed by methanol. The CMW and methanol extracts were combined and loaded onto a DEAE cellulose column. CMW was used to wash the column and lipid-linked oligosaccharides were eluted with 300mM NH_4OAc in CMW. The lipid-linked oligosaccharides were extracted with chloroform and dried.

[00276] To release the glycans from the lipids, the material was resuspended in 1.5mL 0.1N HCl in 1:1isopropanol: water (v/v). The solution was heated at 50 °C for 2 hours and then dried at 75 °C. Residue was suspended in water saturated butanol and the aqueous phase containing the glycans was dried, resuspended in water, and purified with AG50W-H8(hydrogen atom) cation exchange resin followed by Ag1-X8 (formate form) anion exchange resin.

[00277] To confirm the identity of the terminal glycan, the sample was divided and half was treated with β 1,3 galactosidase (NEB) and half with a water control. Samples were incubated at 37 °C for 48 hours. Mass spectrometry revealed a major peak (m/z 609) in the buffer control sample (**Figure 3, top**) consistent with the expected size of the T antigen glycan. In the sample treated with galactosidase, a major peak was detected (m/z 447) consistent with the expected size of the disaccharide GalNAc GlcNAc suggesting loss of the terminal β 1,3 galactose.

EXAMPLE 5

[00278] Immunization with the human T antigen

[00279] The human T antigen is frequently found to be aberrantly expressed in cancers and is thus known as a pancarcinoma antigen. It has been estimated that up to 90% of carcinomas carry the T antigen on the cell surface including carcinomas of the breast, colon, bladder, lung, prostate, liver, and stomach [21, 22]. Because of its specific expression in multiple cancers, the T antigen is of interest as a target of anti-cancer immunotherapy treatments.

[00280] To enable preparation of an immunogen bearing multiple T antigen glycans on a carrier protein *in vivo*, a plasmid was obtained (pTrc-spMBP-GT-MBP-GT) that encodes the MBP protein fused to a 4x GlycTag (bearing 4 DQNAT motifs) at both the N- and C-termini and a 6x-his tag for purification purposes [20]. A second plasmid (pJD-07) was constructed to express a uniform glycan terminating in the T antigen. pJD-07 was cloned using homologous recombination in yeast by insertion of the *neuDBAC* genes into pDisJ-07. pJD-07 contains the following genes as a synthetic operon under control of a pBAD promoter: (5'-3') *galE*, *pglB*, *pglA*, *neuD*, *neuB*, *neuA*, *neuC*, and *wbnJ*.

[00281] *E. coli* strain MC4100 $\Delta waaL$ was transformed with plasmids pTrc-spMBP-GT-MBP-GT and either pJD-07 for expression of target protein glycosylated with the T antigen glycan or pMW07 for expression of aglycosylated target protein. The target MBP protein expressed from plasmid pTrc-spMBP-GT-MBP-GT contains a total of 8 glycosylation sites (MBP8xDQNAT). Strains were cultured under selection with ampicillin (100ug/mL) and chloramphenicol (25 μ g/mL) at 30°C and induced at an ABS_{600} of approximately 1.5 with 0.2% (v/v) arabinose and 0.1mM IPTG for ~16hours. The MBP target protein was purified on a HisTrap FF column (GE Healthcare) followed by DEAE HiTrap FF column (GE Healthcare) and eluted with a NaCl gradient (0-500 mM) in 20mM Tris pH 6.8. The glycosylated protein was affinity purified with Peanut agglutinin (PNA)-agarose (Vector labs) to isolate protein conjugated to the T antigen glycan. Resulting proteins were separated

by PAGE and analyzed by Western blot with α 6x-His (left), or biotin conjugated PNA (5 μ g/mL, Vector labs) and peroxidase-conjugated streptavidin (1:3333, Vector labs, right) to confirm glycosylation. As expected, the MBP expressed with glycosylation plasmid pJD-07 migrated more slowly than the negative control (pMW07) and reacted with the PNA lectin consistent with glycosylation (**Figure 4**).

[00282] Female C3H mice at approximately 8-10 weeks of age were utilized for this study in groups of 5 with feed and water provided *ad libitum*. Aglycosylated MBP8xDQNAT and T-antigen-MBP8xDQNAT prepared as described above were adjusted to a concentration of 0.4mg/mL in PBS. Immediately prior to use, proteins were mixed with an equal volume of Sigma Adjuvant System (Sigma) and mice were immunized through the intraperitoneal (IP) route with 20 μ g of protein in a volume of 0.1 mL on days 0, 7, and 13. Serum samples were collected on day -1 (prior to immunization), and on days 14 and 21.

[00283] **Analysis of antibody response**

[00284] ELISA was used to determine the presence of specific antibodies in the resulting serum. To assess the immune response to the carrier protein, aglycosylated MBP8xDQNAT prepared above was adjusted to a concentration of 2 μ g/mL in Coating Buffer (4.2 g/L NaHCO_3 , 1.78 g/L Na_2CO_3 , pH 9.6) and 50 μ L was applied in triplicate to the wells of a PolySorp microtiter plate (Nunc) and incubated overnight at 4°C. Wells were washed in triplicate with 200 μ L PBS containing Tween-20 (PBST: 4 g/L NaCl, 0.1 g/L KCl, 0.72 g/L Na_2HPO_4 , 0.12 g/L KH_2PO_4 + 0.05% v/v Tween-20) prior to blocking the wells with 200 μ L 10% bovine serum albumin (BSA) in PBST for 60 minutes at room temperature. Serum samples were diluted 1:500 in 1% BSA in PBST, and 50 μ L of each sample was applied in triplicate on coated wells and incubated at room temperature for 60 min. The plates were washed 4 times with 200 μ L PBST then incubated for 60 minutes at room temperature with 50 μ L of a 1:5000 dilution of either HRP conjugated anti-mouse IgM or HRP conjugated anti-mouse IgG specific secondary antibody (Jackson ImmunoResearch Laboratories). The microtiter plates were washed 7 times with 200 μ L PBST and incubated for 10-30 min with 100 μ L of 1-Step Ultra TMB-ELISA (Thermo) at room temperature in the dark. Reactions were stopped with the addition of 100 μ L 2N HCl and absorbances were read at 450 nm (**Figure 5**). For glycosylated and aglycosylated groups, detection of IgM antibodies to MBP8xDQNAT peaks around day 14 (**Figure 5**, top panel), while detection of IgG antibodies increased over the course of the study (**Figure 5**, lower panel).

[00285] A second ELISA was performed to determine the antibody response to the c-terminal portion of the immunogen using GFP modified with a similar tag. A plasmid was

obtained (pTrc-spTorA-GFP-GT) [20] that expresses the GFP protein modified with a 4xGlycTag containing 4 iterations of the DQNAT motif, followed by a 6x-His tag (GFP4xGT). pTrc-spTorA-GFP-GT was used to cotransform MC4100 Δ *waaL* cells with pMW07 or pJD-07. Resulting strains were cultured under selection with ampicillin (100ug/mL) and chloramphenicol (25 μ g/mL) at 30°C and induced at an ABS_{600} of approximately 1.5 with 0.2%v/v arabinose and 0.1mM IPTG for ~16hours. The GFP target protein was purified on a HisTrap FF column (GE Healthcare) followed by DEAE HiTrap FF column (GE Healthcare) and eluted with a NaCl gradient (0-500 mM) in 20mM Tris pH 6.8. The glycosylated protein was additionally affinity purified with Peanut agglutinin (PNA)-agarose (Vector labs) to isolate protein conjugated to the T antigen glycan.

[00286] Resulting T antigen-GFP4xGT, or aglycosylated GFP 4xGT was adjusted to a concentration of 2 μ g/mL in Coating Buffer and 50 μ L was applied in triplicate to the wells of a PolySorp microtiter plate (Nunc) and incubated overnight at 4°C. The wells were washed 3 times with 200 μ L PBST prior to blocking the wells with 10% BSA in PBST for 60 min at room temperature. Serum samples as indicated by were diluted 1:500 in PBST with 1 % BSA, and 50 μ L was applied in triplicate on coated wells and incubated at room temperature for 60 min. Wells were washed 4 times with 200 μ L PBST and incubated for 60 minutes at room temperature with 50 μ L of a 1:5000 dilution of HRP conjugated anti-mouse secondary antibody (Promega). After 7 washes with PBST, reactions were developed by addition of 100 μ L of 1-Step Ultra TMB-ELISA (Thermo) and incubated for 10-30 minutes in the dark. Reactions were stopped with addition of 100 μ L 2N HCl and absorbances were read at 450nm (**Figure 6**). Protein used to coat the wells is indicated beneath the chart and the immunogen used to treat the mice is indicated in the key. Antibody binding to GFP-coated wells was elevated on average (grey rectangles) in serum from mice immunized with aglycosylated MBP8xDQNAT compared with glycosylated T antigen-MBP8xDQNAT suggesting the glycans may have interfered with the antibody response.

EXAMPLE 6

[00287] **Engineering *E. coli* for expression of the human (2,3) sialyl-T antigen**

[00288] The human (2,3) sialyl-T antigen consists of the T antigen glycan modified with a terminal α 2,3 Neuraminic acid (NeuNAc) residue resulting in the following structure: NeuNAc α 2,3 Gal β 1,3 GalNAc α -. To generate a glycan terminating with the sialyl T antigen structure in an *E. coli* host, the plasmid described above expressing genes required to

synthesize the T-antigen glycan (pDisJ-07) was modified to include a gene encoding a sialyltransferase, and genes whose products comprise the cytidine 5' monophospho-*N*-acetylneuraminic acid (CMP-NeuNAc) synthesis pathway in *E. coli* K1.

[00289] A region of DNA was amplified from the *E. coli* K1 genome including the genes *neuB*, *neuA*, and *neuC* using PCR. These encode a Neu5Ac synthase, CMP-Neu5Ac synthetase, and UDP-GlcNAc2-epimerase respectively. The *neuD* gene was also included as it may help to stabilize the *neuB* gene product (Daines DA, Wright LF, Chaffin DO, Rubens CE, Silver RP: **NeuD plays a role in the synthesis of sialic acid in Escherichia coli K1.** *FEMS microbiology letters* 2000, **189**(2):281-284.). The *lst* gene encoding the *N.meningitidis* α 2,3 sialyltransferase was also amplified and both PCR products along with linearized pDisJ-07 were used to co-transform *S. cerevisiae* to make resulting plasmid pJDLST-07 by homologous recombination. The sequences of *neuB*, *neuA*, *neuC*, and *neuD* are provided as Seq ID NOs: 8-11 and the sequence of the *lst* gene is provided as SEQ ID NO: 12. Plasmid pJDLST-07 contains a synthetic operon under control of the pBAD promoter with genes in the following order: *galE*, *pglB*, *pglA*, *neuD*, *neuB*, *neuA*, *neuC*, *lst*, *wbnJ*.

[00290] For use in expressing sialylated glycans, a strain was constructed in which the *nanA* gene encoding the sialic acid aldolase NanA was targeted for disruption. Deletion of the *nanA* gene prevents degradation of sialic acid from external sources (Vimr ER, Troy FA:

Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* 1985, **164**(2):845-853.). The $\Delta nanA::kan$ mutation was introduced into MC4100 *E. coli* via P1 *vir* phage transduction from the corresponding mutant generated as part of the Keio collection (CGSC #10423, Yale genetic stock center)(Baba et al.). The kanamycin cassette was removed by the method of Datsenko and Wanner (Datsenko et al.). To promote glycosylation, the $\Delta waaL::kan$ mutation was subsequently introduced and cured of kanamycin resistance by the same method as described above.

EXAMPLE 7

[00291] ***In vivo* synthesis of proteins carrying an *N*-glycan terminating in the human (2,3) sialyl-T antigen**

[00292] To permit analysis of sialylated glycopeptide by Mass spectrometry, a Glucagon peptide modified with a 1X GlycTag containing a DQNAT motif was cloned. To construct this plasmid, the DsbA signal peptide sequence and the *malE* gene (which encodes MBP) were amplified with primers containing homology to vector pTRCY and the sequence for the TEV protease sites. Similarly, glucagon was amplified from a synthetic oligonucleotide with

primers containing sequence encoding the TEV protease site or the sequence for the 4X GlycTag and 6x-His tag followed by homology to pTRCY. These PCR products were used with linearized pTRCY to co-transform *S. cerevisiae* for cloning by homologous recombination to generate plasmid pMG4X-Y. The related plasmid pMG1X-Y is a derivative of pMG4X-Y made by replacing the 4XGlycTag with a 1X GlycTag. Briefly, pMG4X-Y was linearized and an oligonucleotide encoding the 1XGlycTag was used to replace the 4XGlycTag by homologous recombination in *S. cerevisiae*. The sequence encoding proteins MBP-3TEV-GLUC-4XGlycTag-6H and MBP-3TEV-GLUC-1XGlycTag-6H are provided as SEQ ID NOS: 13 and 14.

[00293] In order to generate glycoprotein *in vivo* containing the human sialyl-T antigen, strain MC4100 Δ *nanA* Δ *waaL* described above was used to promote periplasmic accumulation of sialylated glycans. This strain was co-transformed with plasmid pMG1X-Y encoding a glycosylation acceptor protein and pJDLST-07 which expresses the machinery necessary to synthesize the sialyl-T antigen glycan.

[00294] An overnight culture consisting of MC4100 Δ *nanA* Δ *waaL* pMG1X-Y and pJDLST-07 was used to inoculate a 50mL culture in LB with 100 μ g/ml ampicillin and 25 μ g/ml chloramphenicol. When the ABS_{600} reached approximately 1.5 the culture was induced with arabinose to 0.2% and IPTG to 0.1mM and the cells were harvested by centrifugation approximately 19 hours post-induction. Following cell lysis, protein was purified on a NiNTA column and TEV protease was used to cleave 30 μ l of the resulting eluate. The sample was incubated at 30 °C for 3h and an aliquot was analyzed by mass spectrometry on an AB SCIEX TOF/TOF mass spectrometer using dihydroxybenzoic acid (DHB) as the matrix.

[00295] Mass spectrometry revealed major peaks consistent with the expected size of glucagon modified with the sialyl-T antigen (m/z 6251) and the expected size of glycosylated Glucagon bearing the T antigen terminal glycan (m/z 5960) (**Figure 7**).

EXAMPLE 8

[00296] **Relative improvement of sialylation through expression of *neuDBAC* from TRCY**

[00297] One potential strategy for improving sialylation in this system is to increase the intracellular availability of CMP-NeuNAc. Although the necessary biosynthetic genes are present on plasmid pJDLST-07, it was hypothesized that additional copies could improve sialylation. The genes *neuDBAC* were amplified as a single PCR product and inserted into

pMG1X –Y downstream of the glucagon fusion protein using homologous recombination in *Saccharomyces cerevisiae*. This resulted in creation of plasmid pMG1X D-Y.

[00298] Plasmid pMG1X D-Y was combined with pJDLST-07 in strain MC4100 Δ nanA Δ waaL to test glycosylation in 50mL cultures as described above. Mass spectrometry of the TEV-cleaved peptide product reveals a major peak consistent with the expected size of glucagon modified with the (2,3) sialyl-T antigen containing glycan (m/z 6250). A second smaller peak consistent with the expected size of glucagon modified with the T antigen glycan (m/z 5959) is also detected (**Figure 8**).

EXAMPLE 9

[00299] α 2,3 neuraminidase treatment of sialylated glucagon peptide

[00300] To validate the sialylation of the glucagon peptide a neuraminidase treatment was performed. Strain MC4100 Δ nanA Δ waaL carrying plasmids pMG1X D-Y and pJDLST-07 is grown in a 50mL culture in LB with 100 μ g/ ml ampicillin and 25 μ g/ ml chloramphenicol and induced with 0.2% arabinose and 0.1mM IPTG for approximately 16h. The recombinant protein is purified from the lysate with nickel affinity chromatography and the eluate is buffer exchanged in 50mM Tris pH 8.0 100mM NaCl and concentrated prior to incubation for 3h at 30 °C with TEV protease. The protein is divided and incubated with α 2,3 neuraminidase (NEB) or a buffer control for 2 hours at 37 °C prior to analysis by Mass spectrometry (**Figure 9**). The major peak in the buffer control sample (m/z 6253, black) is consistent with the expected size of glucagon modified with the sialyl-T antigen glycan. In the sialidase treated sample, the major peak (m/z 5961, gray) is consistent with the expected size of the T antigen glycopeptide. No evidence of the sialylated glycopeptide was present following neuraminidase treatment.

EXAMPLE 10

[00301] Determining the effect of glycosylation on stability *in vitro*

[00302] Glycosylation is a well-known strategy for improving the stability of a protein and is a rational approach for improving both *in vivo* or *in vitro* persistence. In order to determine if N-glycosylation in bacteria could be utilized for this purpose, the (2,3) sialyl- T antigen was conjugated to glucagon for analysis.

[00303] Plasmid pMG1X D-Y was combined with pJDLST-07 in strain MC4100 Δ nanA Δ waaL to generate sialylated glucagon and resulting cells were used to inoculate a 100 mL culture containing LB medium supplemented with 100 μ g/mL ampicillin

and 25 µg/mL chloramphenicol. To generate aglycosylated glucagon, Origami2 $\Delta nanA \Delta waaL \Delta gmd::kan$ harboring plasmid pMG1X MCB-07 was used to inoculate a 100mL culture containing LB medium and 100 µg/mL ampicillin. This strain was selected based on our ability to detect the aglycosylated peptide. Both cultures were grown with shaking at 30 °C until an ABS_{600} of ~2.3 was reached then were induced with 0.1mM IPTG (both) and arabinose 0.2% v/v (glycosylated only). Cultures were maintained at 30 °C overnight. The glucagon fusion protein was isolated by Ni affinity (NiNTA, Qiagen) and the eluate was concentrated. 1 µl TEV protease was added to 50 µl of glycosylated or aglycosylated glucagon and the reaction was incubated at 30 °C for 3 hours then transferred to 37 °C. Presence of glucagon was monitored over time by MALDI TOF mass spectrometry. The aglycosylated glucagon was no longer detected after 21 hours of incubation whereas a peak at the expected m/z of glucagon bearing the (2,3) sialyl T antigen was still the most prominent. (**Figure 10**) suggesting that sialylation enhances the persistence *in vitro* of the glucagon peptide.

EXAMPLE 11

[00304] Engineering *E. coli* for expression the T antigen modified with a terminal $\alpha 2,6$ NeuNAc

[00305] The human sialyl-T antigen consists of the T antigen glycan modified with a terminal $\alpha 2,3$ Neuraminic acid (NeuNAc) residue resulting in the following structure: NeuNAc $\alpha 2,3$ Gal $\beta 1,3$ GalNAc α -. A related glycan was also explored differing only in the linkage of the terminal NeuNAc residue: NeuNAc $\alpha 2,6$ Gal $\beta 1,3$ GalNAc α . To generate a glycan terminating with the 2,6 sialylated T antigen structure in an *E. coli* host, the plasmid described above expressing genes required to synthesize the 2,3 sialyl T-antigen glycan (pJDLST-07) was modified by replacing the *lst* gene with the a gene encoding a 2,6 sialyltransferase.

[00306] To express the structure NeuNAc $\alpha 2,6$ Gal $\beta 1,3$ GalNAc $\alpha 1,3$ GlcNAc, a codon-optimized version of *Pdst6* encoding a 2,6 sialyltransferase from *Photobacterium damsela* JT0160 was synthesized by Mr. Gene and amplified by PCR. The *Pdst6* gene was cloned in place of the *lst* gene in pJDLST-07 by homologous recombination in yeast to create pJDPdST6fl-07. The sequence of the *PdST6* gene is provided as SEQ ID NO 15. Plasmid pJDPdST6fl-07 contains a synthetic operon under control of the pBAD promoter with genes in the following order: *galE*, *pglB*, *pglA*, *neuD*, *neuB*, *neuA*, *neuC*, *Pdst6*, *wbnJ*.

[00307] *In vivo* synthesis of proteins carrying an N-glycan terminating in 2,6 Sialic acid

[00308] In order to generate glycoprotein *in vivo* containing the 2,6 sialylated-T antigen, strain MC4100 Δ *nanA* Δ *waaL* described above was used to promote periplasmic accumulation of sialylated glycans. This strain was co-transformed with plasmid pMG1X D-Y encoding a glycosylation acceptor protein and pJDPdST6fl-07 which expresses the machinery necessary to synthesize the 2,6 sialic acid-terminal glycan.

[00309] An overnight culture consisting of MC4100 Δ *nanA* Δ *waaL* pMG1X D-Y and pJDPdST6fl-07 was used to inoculate a 50mL culture in LB with 100 μ g/ml ampicillin and 25 μ g/ml chloramphenicol. When the ABS₆₀₀ reached approximately 1.5 the culture was induced with arabinose to 0.2% and IPTG to 0.1mM, and the cells were harvested by centrifugation approximately 19 hours post-induction. Following cell lysis, protein was purified on a NiNTA column and TEV protease was used to cleave 30 μ l of the resulting eluate. The sample was incubated at 30°C for 3h and an aliquot was analyzed by mass spectrometry on an AB SCIEX TOF/TOF mass spectrometer using dihydroxybenzoic acid (DHB) as the matrix.

[00310] Mass spectrometry revealed major peaks consistent with the expected size of glucagon modified with the 2,6 sialylated T antigen (m/z 6257) and the expected size of glycosylated Glucagon bearing the T antigen terminal glycan (m/z 5964) (**Figure 11**).

[00311] To confirm that the glycan produced from plasmid pJDPdST6fl-07 does not in fact terminate in the 2,3 sialyl T antigen, the glycopeptide generated above was treated with neuraminidases with different specificities. The sialylated glucagon peptide was divided and incubated for 30 minutes at 37 °C with α 2,3 Neuraminidase (NEB) or Neuraminidase (NEB) which is reported to hydrolyze α 2,3-, α 2,6-, and α 2,8-linked sialic acid from glycans or glycoprotein. The resulting peptides were analyzed by Maldi TOF Mass spectrometry (**Figure 12**). The fraction treated with α 2,3 Neuraminidase (**Figure 12, left**) was found to contain a substantial peak at the expected size for the sialylated product (m/z 6256) but this peak was absent from the sample treated with Neuraminidase (**Figure 12, right**). The major peak in the Neuraminidase treated sample (m/z 5963) is consistent with the expected size of the T antigen glycopeptide, consistent with loss of sialic acid. Loss of the peak at m/z 6257 upon treatment with Neuraminidase, but not α 2,3 Neuraminidase, is consistent with the production of sialylated glycopeptide with the expected terminal α 2,6 NeuNAc residue.

EXAMPLE 12

[00312] **Production of a recombinantly produced polysialylated glycan in *E. coli*.**

[00313] There are several bacteria known to produce polysialic acid (PSA) glycans including *E. coli* K1 and strains of *Neisseria meningitidis*. In these strains PSA forms a protective capsular polysaccharide. The PSA capsule is well-studied in *E. coli* K1 but the lipid substrate for PSA synthesis has not been identified. In order to adapt PSA for N-glycosylation, it is likely necessary to direct its synthesis on a substrate appropriate for the OST and provide the necessary disialic acid 'primer' required for the PSA polymerase to extend sialylation. The glycan described herein terminating in the human T antigen is a good candidate for polysialylation because it is efficiently used in glycosylation in this system. To elaborate the T antigen with a disialic acid motif, the genes *cstII* from *C. jejuni* and *lic3B* from *H. influenza* were selected based on their reported bifunctional 2,3 and 2,8 sialyltransferase activities. For polymerization the *neuS* gene was chosen for successive 2,8 sialylation because it is an *E. coli* gene.

[00314] To clone plasmids use in exploring polysialylation, synthetic versions of the *cstII* and *lic3b* genes were obtained (Mr. Gene). The sequences of *cstII* and *lic3b* are supplied as SEQ ID NOs: 16 and 17.

[00315] A truncated version of the gene *cstII* encoding the first 260 amino acids of the bifunctional α 2,3 α 2,8 sialyltransferase was cloned with *neuBAC*, the *E. coli* K1 polysialyltransferase *neuS* (SEQ ID NO: 18), and the genes to synthesize the T antigen glycan using homologous recombination in *Saccharomyces cerevisiae*. The full length bifunctional α 2,3 α 2,8 sialyltransferase *lic3b* was also cloned in the same manner. The resulting plasmids are called pJCstIIS-07 and pJLic3bS-07.

[00316] Plasmid pJCstIIS-07 was used to transform MC4100 $\Delta nanA$ and MC4100 $\Delta nanA \Delta waaL$ for functional testing. A single colony is used to inoculate 1mL of LB medium containing 0.25% NeuNAc (w/v), 25 μ g/ml chloramphenicol and 0.2% (v/v) arabinose. Cultures were grown approximately 18 hours at 30 °C in a 1.5 mL tube and pelleted. After washing with PBS, cultures are normalized by optical density, heated for 10 min at 95°C, and the whole cells are spotted on nitrocellulose when cooled. The membrane is blotted with an anti-PSA antibody followed by anti-mouse-horseradish peroxidase (**Figure 13a**). Reactivity with the PSA antibody suggests that a PSA- glycan is displayed on the cell surface in the presence of *waaL*. The structure of the expected glycan is diagrammed (**Figure 13b**).

[00317] To test the putative PSA-terminal glycan in a glycosylation reaction, the MC4100 $\Delta waaL \Delta nanA$ strain was transformed with pMG4X-Y encoding a glycosylation acceptor protein. The resulting strain was transformed with plasmid pDisJ-07 or pJLlc3B-07. Resulting strains were grown in 50mLs LB +/- 0.25% NeuNAc and appropriate antibiotics.

Cultures are induced at an approximate optical density between 2-4 with 0.2% arabinose and 0.1mM IPTG. Proteins were purified by nickel affinity chromatography, concentrated, and treated with TEV protease prior to analysis by Western blot (**Figure 14**).

[00318] Detection with the α PSA antibody (**Figure 14, top**) showed some reactive material only in the presence of pJLic3BS-07 and NeuNAc supplementation consistent with presence of a PSA glycan. Total target protein is detected by the presence of the hexanitidine tag with α His antiserum (**Figure 14, bottom**).

EXAMPLE 13

[00319] NeuD is important for synthesis of sialylated glycans in *E. coli* MC4100

[00320] The *neuD* gene is part of the genetic locus for PSA synthesis in *E. coli* K1 and other strains that produce sialylated glycans although there are conflicting assignments of NeuD function. In order to confirm the importance of NeuD in the sialylation platform it was cloned as an individual gene into vector pTRCY using homologous recombination in *Saccharomyces cerevisiae*. The resulting plasmid containing NeuD under the control of the Trc promoter is called pNeuD-Y.

[00321] To test pNeuD-Y, this plasmid was used with pJLic3BS-07 to cotransform strain MC4100 Δ *nanA*. A single colony is used to inoculate 1mL of LB medium containing 25 μ g/ml chloramphenicol and 0.2% arabinose. LB medium was made with or without sialic acid at a final concentration of 0.25% (w/v) and was adjusted for pH and filter sterilized. Cultures are grown approximately 18 hours at 30 °C in a 1.5 mL tube and the cultures are pelleted. After washing with PBS, cultures are normalized by optical density and heated for 10min at 95 °C and the whole cells are spotted on nitrocellulose when cooled. The membrane is blotted with an anti-PSA antibody followed by anti-mouse-horseradish peroxidase (**Figure 15**).

[00322] Reactivity with the PSA antibody suggests the presence of a cell surface PSA glycan in the presence of pNeuD-Y or NeuNAc. This result suggests the importance of NeuD in production of sialylated compounds in laboratory *E. coli* (**Figure 15**).

EXAMPLE 14

[00323] *Ex vivo* polysialylation

[00324] As an alternative method to confirm the functionality of polysialyltransferases in laboratory *E. coli*, an *ex vivo* method for polysialylation was utilized. For this method a lysate is generated from a strain expressing a polysialyltransferase and it is combined with CMP-

NeuNAc and an acceptor protein produced in a separate strain. MBP was selected for use as the acceptor protein because it is expressed and glycosylated efficiently in this system.

[00325] To prepare the acceptor protein plasmid, the coding sequence for MBP modified with the DsbA signal peptide and a 4X GlycTag and hexahistidine motif was subcloned from pTRC99-MBP 4XDQNAT (Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP: **Production of Secretory and Extracellular N-Linked Glycoproteins in Escherichia coli**. *Applied and Environmental Microbiology* 2011, 77(3):871-881.). The resulting plasmid is termed pMBP4XGT-Y. CstII was also cloned as a translation fusion to the Neisserial polysialyltransferase SiaD (obtained from Genwiz) to make a self-priming polysialyltransferase as described by Willis et al (Willis LM, Gilbert M, Karwaski M-F, Blanchard M-C, Wakarchuk WW: **Characterization of the α -2,8-polysialyltransferase from Neisseria meningitidis with synthetic acceptors, and the development of a self-priming polysialyltransferase fusion enzyme**. *Glycobiology* 2008, 18(2):177-186.). Two versions were cloned using homologous recombination in *Saccharomyces cerevisiae* resulting in plasmids pCstII-SiaD-Y and pCstII153S-SiaD-Y, the latter of which includes a mutation of isoleucine 53 to cysteine which is reported to improve the α 2,8 sialyltransferase activity. The sequence of *siaD* is provided as SEQ ID NO: 19.

[00326] An acceptor glycoprotein was first prepared by addition of the T antigen-containing glycan to the MBP4XGT protein. Plasmids pMBP4XGT-Y and pDisJ-07 were used to transform strain MC4100 Δ *waaL*. The resulting strain was used to inoculate a 1L culture containing LB, ampicillin (100ug/ml), and chloramphenicol (25ug/ml). The culture was incubated at 30 °C until the optical density reached OD 1.5 and then both glycan and glycoprotein production are induced with 0.2% arabinose and 0.1mM IPTG respectively. The pellet was harvested after 16hours and the his-tagged protein purified by nickel affinity chromatography. Eluted protein is buffer exchanged into *ex vivo* sialylation buffer containing 50mM Tris 7.5, 10mM MgCl₂ and concentrated.

[00327] To prepare the polysialyltransferase lysates, strains MC4100 Δ *waaL* containing plasmid pTRCY, pCstII-SiaD-Y, or pCstII153S-SiaD-Y were grown in 50mL cultures containing LB and ampicillin. When the optical density reached 1.5-1.9, protein expression is induced with the addition of IPTG to a final concentration of 0.1mM and induction is carried out at 20 °C for approximately 16 hours. Pellets were harvested and resuspended in *ex vivo* sialylation buffer. Following cell lysis, the material is centrifuged at 1000 x g for 11 minutes and the supernatant was retained.

[00328] For the *ex vivo* reaction, 20 μ l of the MBP glycoprotein was combined with 30 μ l of the polysialylation or control lysate and CMP-NeuNAc. Reactions are incubated at 37 °C for 45 minutes prior to analysis by SDS-PAGE and Western blot (**Figure 16**). Incubation with anti PSA antiserum (**Figure 16, top panel**) resulted in appearance of high molecular weight material in the presene of both CMP-NeuNAc and lysate containing pCstIII53S-SiaD-Y consistent with the formation of a PSA glycan. It appeared that there was a reduced amount of reactive material generated with the lysate containing the pCstII-SiaD-Y plasmid and none detected with the vector control. The presence of the MBP4XGT protein was confirmed with an anti-Histidine Western blot (**Figure 16, lower panel**).

EXAMPLE 15

[00329] ***In vivo* synthesis of an N-glycan terminating in the human blood group O glycan (H-antigen) in *E. coli***

[00330] The human blood group O determinant or H-antigen consists of a fucosylated glycan that is similar to the human T antigen. The type III H-antigen structure consists of Fucose α 1,2 Galactose β 1,3 GalNAc α -. To synthesize a glycan in *E. coli* terminating in the human H-antigen structure, the genes from the plasmid described above expressing genes required to synthesize the T-antigen glycan were combined with a gene encoding a fucosyltransferase.

[00331] Fucosyltransferase WbnK from *E. coli* O86 was selected because it is a bacterial enzyme that fucosylates a glycan with similar structure in its native context. The sequence of *wbnK* is provides as SEQ ID NO: 20. A PCR product containing the *wbnJ* and *wbnK* genes was generated using a synthetic template from Genewiz. The PCR product was combined with linear pDis-07 plasmid using homologous recombination in yeast to generate plasmid pDisJK-07. The resulting plasmid, pDisJK-07, contains a synthetic operon under control of the pBAD promoter with genes in the following order: *galE*, *pglB*, *pglA*, *wbnJ*, *wbnK*.

[00332] For use in expressing fucosylated blood group H-antigen, the *E. coli* strain LPS1 (Yavuz E, Maffioli C, Ilg K, Aebl M, Priem B: **Glycomimicry: display of fucosylation on the lipo-oligosaccharide of recombinant Escherichia coli K12**. *Glycoconjugate journal* 2011, **28**(1):39-47.) was used to promote accumulation of GDP-fucose (GDP-Fuc). *E. coli* encodes a native pathway for synthesis of GDP-Fuc however this sugar nucleotide is then normally incorporated into the fucose-containing exopolysaccharide colanic acid. To prevent usage of GDP-Fuc in this competing pathway a mutation is present in the gene *wcaJ* (ECK2041) encoding a putative UDP-glucose lipid carrier transferase. To further promote

glycosylation in this strain, a mutation in the *waaL* gene was introduced. The *waaL* (*rfaL*) gene has been previously mutated as part of the Keio collection and the resulting strain *rfaL734(del)::kan* (JW3597-1) (Baba et al.) was obtained from the Yale Coli Genetic Stock Center (CGSC). P1 *vir* phage was used to transduce the *waaL* mutation into the LPS1 recipient to make strain LPS1 $\Delta waaL::kan$.

[00333] To confirm the glycan structure produced by the glycosyltransferases encoded by pDisJK-07, the plasmid was used to transform strain LPS1 $\Delta waaL::kan$ for analysis of the lipid-released oligosaccharides. A 250mL culture of the resulting strain was grown at 30 °C and induced when the optical density reached an ABS_{600} around ~2.0. Cells were harvested after ~20 hours for isolation of lipid-linked oligosaccharides by the method of Gao and Lehrman. Briefly, pellet was resuspended in 10 mL methanol and lysed by sonication. Material was dried at 60 °C and subsequently resuspended in 1mL 2:1 chloroform: methanol (v/v, CM) via sonication and material was washed two times in CM. The pellet was then washed in water then lipids were extracted with 10:10:3 chloroform: methanol: water (v/v/v, CMW) followed by methanol. The CMW and methanol extracts were combined and loaded onto a DEAE cellulose column. CMW was used to wash the column and lipid-linked oligosaccharides were eluted with 300mM NH_4OAc in CMW. The lipid-linked oligosaccharides were extracted with chloroform and dried.

[00334] To release the glycans from the lipids, the material was resuspended in 1.5mL 0.1N HCl in 1:1 isopropanol: water (v/v). The solution was heated at 50 °C for 2 hours and then dried at 75 °C. Residue was suspended in water saturated butanol and the aqueous phase containing the glycans was dried, resuspended in water, and purified with AG50W-H8 (hydrogen atom) cation exchange resin followed by Ag1-X8 (formate form) anion exchange resin.

[00335] Purified oligosaccharides solubilized in water were subjected to incubation with $\alpha 1,2$ fucosidase (NEB) treatment) or a buffer only control and analyzed on an AB SCIEX TOF/TOF mass spectrometer using dihydroxybenzoic acid (DHB) as the matrix (**Figure 17a**). In the buffer control (top panel), two major peaks present (m/z 755) and (m/z 609) are consistent with the expected (m/z) of the fucosylated product (Fuc Hex HexNAc₂) and the T antigen glycan (Hex Hex NAc₂) respectively. Following fucosidase treatment (bottom panel), the peak at (m/z 755) is greatly reduced while the peak at (m/z 609) is relatively larger. The difference between these peaks (146) is consistent with the size of a fucose residue (deoxyhexose).

EXAMPLE 16

[00336] Improving relative fucosylation through expression of GDP-Fucose biosynthetic genes

[00337] In order to improve conversion from the T antigen glycan to the fucosylated product, a system was devised in order to allow for expression of additional copies of the biosynthetic machinery for GDP-Fucose, UDP-Gal, and UDP-GalNAc. To accomplish this, the following genes were cloned as a synthetic operon under control of the pBAD promoter in pMQ70: *galE* (*C. jejuni*), *galE*, *gmd*, *fcl*, *gmm*, *cpsB*, *cpsG* (*E. coli* K12) to make plasmid pGNF-70 using homologous recombination in yeast. The sequences of the *E. coli* genes cloned in pGNF-70 are provided as SEQ ID NOs: 21-26.

[00338] Strain LPS1 $\Delta waaL::kan$ was transformed with plasmids pJK-07 and pGNF-70. The resulting strain was cultured in 250mL LB medium under ampicillin and chloramphenicol selection and expression of both plasmids was induced at an optical density of approximately 2.0 and induction continued at 30 °C for approximately 16 hours. Pellets were harvested and LLOs were purified as previously described by the method of Gao and Lehrman.

[00339] Purified oligosaccharides were analyzed by Mass Spectrometry as described above (**Figure 17b**). The major peak identified following this treatment (m/z 755) is consistent with the desired fucosylated glycan (dHex Hex HexNAc₂) suggesting efficient fucosylation. An additional peak is present at (m/z 609) which is consistent with the glycan (Hex HexNAc₂).

EXAMPLE 17

[00340] Generating a fucosylated glycoprotein *in vivo* in *E. coli*

[00341] Following analysis of the fucosylated glycan, it is necessary to confirm that the glycan is amenable to use in the glycosylation reaction. The TNF α Fab was selected as an initial target for glycosylation. A codon optimized version of the Fab including signal peptide sequences for each chain was obtained from DNA 2.0 and cloned into pTRCY using homologous recombination in *S. cerevisiae* to append a 4X GlycTag and hexahistidine tag to the heavy chain. The resulting plasmid is designated pTnf α Fab4X-Y. The sequence of the modified TNF α Fab light and heavy chains are supplied as SEQ ID NOs: 27 and 28.

[00342] pTnf α Fab4X-Y was used to transform strain LPS1 bearing glycosylation plasmid pJK-07 or empty vector pMW07 and the resulting strains were used to inoculate a 50mL culture of LB and grown under selection of ampicillin and chloramphenicol. At an optical density of ABS₆₀₀ of 1.5, expression of both plasmids was induced with the addition of 0.2%

arabinose and 0.1mM IPTG and cultures were maintained at 30 °C for approximately 16 hours. Protein was purified using nickel affinity chromatography was subjected to SDS PAGE followed by Western blot with anti Histidine antibody. A mobility shift was apparent for the Fab heavy chain grown in the presence of glycosylation plasmid pJK-07 but not vector pMW07 consistent with glycosylation (**Figure 18**).

EXAMPLE 18

[00343] Generating a fucosylated glycopeptide in vivo in *E. coli* modified with the blood group H-antigen

[00344] Experiments described above indicated the potential for increasing the relative amount of fucosylated product as determined by Mass spectrometry through expression of additional copies of the GDP-Fucose biosynthetic pathway. A plasmid pMG1X-Y encoding the glycosylation acceptor peptide is modified using yeast homologous recombination to also include the following genes: *galE* (*C. jejuni*), *galE* (*E. coli*), *gmd*, *fcl*, *gmm*, *cpsB*, and *cpsG* to make plasmid pMG1X-GNF-Y. A similar plasmid was cloned in the same manner with the following genes in addition to the glucagon construct: *wbnK*, *galE* (*E. coli*), *gmd*, *fcl*, *gmm*, *cpsB*, and *cpsG* termed pMG1X-KGF-Y.

[00345] In preparation for glycosylation, strain LPS1 is transformed with plasmid pDisJK-07. To this, plasmids encoding the glycosylation acceptor protein (pMG1X-Y) or the acceptor protein with the GDP-Fucose biosynthetic machinery were added (pMG1X-GNF-Y, pMG1X-KGF-Y). Resulting strains were grown at 30 °C in 50mL cultures in LB medium with ampicillin and chloramphenicol. Both plasmids were induced with the addition of 0.2% arabinose and 0.1mM IPTG when the culture reached an approximate optical density of ABS_{600} 1.5. After 16 hours, pellets were harvested and proteins purified by nickel affinity chromatography. Eluate was exchanged into 50mM Tris, 100mM NaCl and 30 μ L of the concentrated protein was treated with TEV protease for 3 hours to release the glycopeptide.

[00346] Glycopeptide was analyzed on an AB SCIEX TOF/TOF mass spectrometer using dihydroxybenzoic acid (DHB) as the matrix (**Figure 19**). Peaks consistent with the expected sizes of the fucosylated glycopeptide (dHex Hex HexNAc₂, m/z 6103) and galactosylated glycopeptide (Hex HexNAc₂, m/z 5957) are present in glycopeptide prepared from the strain with plasmid pMG1X (left). Side product is marked with an asterick. Glycopeptide from the strain harboring pMG1X GNF-Y exhibited one major peak consistent with the expected m/z of the H-antigen glycopeptide (dHex Hex HExNAc₂, m/z 6105). An additional smaller peak

at (m/z 5960) is also present likely representing remaining unfucoylated glycopeptide containing the T antigen glycan (Hex HexNAc₂).

[00347] Glycopeptide prepared from strain LPS1 pJK-07 pMG1X KGF-Y was divided and subjected to treatment with α 1,2 fucosidase (NEB) or a buffer control for 8 hours at 37 degrees prior to analysis on an AB SCIEX TOF/TOF mass spectrometer using DHB as the matrix (**Figure 20**). The major peak present in the buffer-only sample (m/z 6107) is consistent with the expected size of the H-antigen containing glycan (dHex Hex HexNAc₂). The sample treated with fucosidase has a major peak at (m/z 5963) consistent with the expected size of the gal terminal T antigen glycan (Hex HexNAc₂).

EXAMPLE 19

[00348] Glycosylation of GH2 with the H antigen glycan

[00349] In order to examine glycosylation of a recombinant human protein, human growth hormone placental variant (GH2) was adapted for expression in this *E. coli* platform.

Homologous recombination in yeast was used to fuse the *malE* gene sequence with a 3' TEV protease cleavage site to the gene encoding GH2 bearing a c-terminal hexahistidine motif in pTrcY. In addition, the sequence surrounding the native glycosylation site of this protein was modified to encode a DQNAT. The genes found to improve generation of the H antigen glycan (*galE Cj*, *galE Ec*, *gmd*, *fcl*, *gmm*, *cpsB*, *cpsG*) were inserted after the 3' end of sequence encoding the GH2 fusion protein to make plasmid pG4-His-GNF-Y. The DNA sequence for the GH2 fusion protein is provided as SEQ ID NO 29.

[00350] pG4-HisGNF-Y was used to transform *E. coli* strain LPS1 for optimal fucosylation with the H antigen glycan. The resulting strain was made electrocompetent and transformed with a second plasmid containing the genes for expression of the glycosyltransferases required to produce the H antigen glycan and oligosaccharyltransferase PglB (pJK-07). To express GH2 and GH2-H antigen, one liter of LB was inoculated with 50mL of overnight culture and grown at 30 °C under selection with 100 μ g/mL ampicillin or 100 μ g/mL ampicillin and 25 μ g/mL chloramphenicol respectively. Cultures were incubated on a shaking platform at 30 °C until an approximate ABS₆₀₀ of 3.0 was reached. The culture containing pG4-His-GNF-Y plasmid alone was induced with 0.1 mM IPTG while the culture containing G4-His-GNF-Y and pJK-07 plasmids was induced with 0.1 mM IPTG and 0.2% v/v arabinose for 16 hr at 30° C.

[00351] Purification and determination of glycoylation

[00352] The cells were pelleted and then resuspended in Ni-NTA column binding buffer with lysozyme at 1mg/mL. Cells were incubated on ice for 30 minutes then disrupted by sonication with five 10second pulses. After sonication, the clarified lysate was then filtered through a 5µm filter and then a 0.45µm filter and Ni-affinity purified using 5 mL HisTrap FF column (GE Healthcare). The protein was buffer exchanged into DEAE loading buffer (20 mM Tris, pH 6.8) and purified using a 5mL DEAE HiTrap FF column (GE Healthcare) and eluted with a NaCl gradient (0-500 mM) in 20mM Tris pH 6.8. Eluted protein was pooled, concentrated, and exchanged into Ni-NTA column binding buffer containing 17mM β Mercaptoethanol (BME) then treated with ~1000U TEV protease and incubated overnight at room temperature. The success of cleavage from MBP was assessed by coomassie-stained SDS-PAGE. Once the cleavage was completed, the protein was then brought up in 20mL of Ni-NTA binding buffer and purified by Ni-NTA resin as described above. Fractions were concentrated down to 1 mL and loaded onto a GST/Amylose mixed resin gravity flow column to remove MPB and TEV. hGH was found in the column wash buffer (50 mMTris, 1 mM EDTA, 200 mM NaCl, pH 7.4). Each protein was then concentrated to ~1 mg/mL and stored at -80°C.

[00353] The purified GH2-H antigen was analyzed to assess glycosylation. Purified protein was separated by SDS PAGE and transferred to PVDF for Western blot with detection by with αhGH antibody (Abcam AB9821) (**Figure 21**, left panel). The appearance of a doublet is consistent with glycosylation. GH2-H antigen was further analyzed by MALDI TOF mass spectrometry. Analysis revealed the major peak (m/z 23047.8) is consistent with the expected size of the GH2 protein modified with the fucosylated H antigen (**Figure 21**, right) suggesting efficient glycosylation. An additional peak noted at m/z 22328.5 is consistent with the expected size of the aglycosylated protein.

EXAMPLE 20

[00354] **Assessing *in vitro* and *in vivo* properties of GH2-H antigen**

[00355] To determine if glycosylation benefits protein stability, resistance to agitation was compared for glycosylated and aglycosylated forms of GH2. A 0.5 mg/mL solution of each GH2 form was vortexed for various lengths of time over a range of 0-10 minutes. The ABS₄₀₅ nm was immediately recorded to determine the turbidity of the solution. Additionally, the ABS₂₈₀ nm was also monitored to determine if the increase in turbidity was due to aggregation, where loss of soluble protein would result in a decrease in the ABS₂₈₀ nm reading. Each dataset was plotted and the rate of denaturation was determined (**Figure 22a**).

The increase in turbidity was slower and less severe for the GH2-H antigen protein compared to GH2 alone consistent with a greater resistance to agitation induced denaturation.

[00356] To determine if glycosylation of GH2 affects receptor binding, both the aglycosylated and glycosylated forms of GH2 were subjected to an ELISA-based receptor binding assay. MaxiSorp ELISA plates were incubated with 2 µg/mL of the ectodomain of the hGH receptor fused to IgG (hGHR) (R and D systems) for 2 hours at room temperature. The plates were subsequently blocked for 1 hour with blocking buffer (5% BSA w/v, 0.1% Tween-20 v/v in PBS) then washed with PBS. A concentration range of 0-500 nM of each GH2 form was incubated in the hGHR-coated ELISA plate for 1 hour at room temperature and subsequently washed with blocking buffer. Each well was incubated with either an anti-HisTag-HRP antibody or an anti-hGH antibody for 1 hour at room temperature and subsequently washed with blocking buffer. In the case of the anti-hGH antibody, a mouse HRP-conjugated 2° antibody was incubated in each well for 45 minutes at room temperature. The wells were then washed and developed with 1-Step Ultra TMB-ELISA (Thermo) and the reaction was stopped with 2 M HCL and read at 450 nm. Each K_d value was determined by plotting the values on GraphPad Prism software. Calculated dissociation constants (K_d) were $1.5 \pm 0.7 \times 10^1$ nM for the GH2 and $1.6 \pm 0.9 \times 10^1$ nM for the GH2-H antigen suggesting that the effect of glycosylation on receptor binding is negligible (**Figure 22b**).

[00357] In vivo half-life

[00358] To determine if glycosylation has any effect to the half-life of GH2, each form was studied in a rat model *in vivo* half-life assay. A single intravenous bolus dose of 300 µg of a 1 mg/mL GH2 or GH2-H antigen solution was administered to a group of 4 Sprague Dawley male rats. Blood samples were drawn at various time points over a 24 hr period and the serum was separated by centrifugation. An anti-hGH antibody (Abcam) was diluted into a carbonate buffer pH 9.6 to a concentration of 2 µg/mL and used to coat MaxiSorp ELISA plates for 2 hours at room temperature. The plates were then blocked, washed and incubated with the serum of each time point for 30 min at room temperature. An anti-HisTag-HRP conjugate antibody was added to the wells for 30 min at room temperature and subsequently developed with 1-Step Ultra TMB-ELISA (Thermo). The reaction was stopped by adding 2 M HCl and absorbance recorded at 450 nm. Detection of the GH2-H antigen protein was more robust than that of the aglycosylated version, most notably in the last three time points suggesting glycosylation with the H antigen glycan enhanced serum persistence (**Figure 23**).

EXAMPLE 21

[00359] Production of glycans with related structures

[00360] An important consideration in the design of strategies for *N*-glycoprotein production in *E. coli* is the efficiency of the glycosylation reaction which is likely to be determined by a number of factors. One such factor is any selectivity the OST may display for particular glycan structures. The glycoforms outlined in the preceding examples are all elaborations of an efficiently transferred “base” glycan consisting of a GalNAc α 1,3 GlcNAc glycan from which structures including the T antigen, (2,3) sialyl T antigen, H antigen, and (2,6) sialylated T antigen were made. These examples have served in part as tools to illustrate the ability to generate *E. coli* glycoproteins containing galactosylated, sialylated, and fucosylated glycans, however are not meant to be an exclusive representation of the glycan structures that can be built from the basic structure.

[00361] The Lewis x glycan (Gal β 1,4[Fuc α 1-3]GlcNAc) for example in addition to related structures, could be built from the T antigen glycan. To do this, genes encoding glycosyltransferases such as those from *Haemophilus influenzae* with β 1,3 GlcNAc transferase (LsgE) and β 1,4 Gal transferase (LsgD) activities, along with an α 1,3 fucosyltransferase such as *Helicobacter pylori* FucT [23] would be inserted into the pdisJ-07 plasmid. This plasmid, when coexpressed with pGNF-70 or pMG1X-GNF-Y to allow sufficient accumulation of required sugar nucleotides, would be expected to result in production of a glycan with the structure Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3Gal β 1,3GlcNAc. The sequences of the LsgE, LsgD, and FucT proteins are included as SEQ ID NOs: 30-32.

[00362] Similarly, the H antigen glycan discussed in Examples 15-19 could be further built upon to generate additional related structures. The human blood group determinants AB and O are interrelated structures based on the blood group O glycan (H antigen). *E. coli* O86 naturally makes an oligosaccharide similar to the human blood group B glycan and thus is a potential source of the galactosyltransferase activity required to extend the H antigen glycan. By inserting the gene encoding the α 1,3 galactosyltransferase WbnI[24] into the existing pJK-07 plasmid and expressing it under similar conditions used to generate the H antigen glycan, is expected to result in production of a glycan with the structure Gal α 1,3[Fuc α 1,2] Gal β 1,3 GalNAc α 1,3GlcNAc. Similarly, an α 1,3 GalNAc transferase such as BgtA from *Helicobacter mutelae* [25] could be used to generate a glycan containing the A antigen with the structure GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 GalNAc α 1,3GlcNAc. The amino acid sequences of WbnI and BgtA are included as SEQ ID NOs: 33-34.

[00363] It is further expected that the the oligosaccharides described herein could be assembled on an alternate UndPP-linked sugar. Alternatives may include GalNAc which can

be attached to UndP by GNE from *E. coli* O157 [26] (SEQ ID NO: 35) or Bacillosamine through the activity of *C. jejuni* glycosyltransferase PglC (SEQ ID NO: 36) and sugar nucleotide synthesis proteins PglFED[27] (SEQ ID NOs: 37-39).

INFORMAL SEQUENCE LISTINGS

Sequence ID No 1**pMW07: vector****7610 bp ds-DNA**

1 gatttatctt cgtttcctgc aggttttgt tctgtgcagt tgggttaaga atactgggca
61 atttcatgtt tcttcaacac tacatatgcg tatatatacc aatctaagtc tgtgctcctt
121 ccttcgttct tctttctgtt cggagattac cgaatcaaaa aaatttcaaa gaaaccgaaa
181 tcaaaaaaaaa gaataaaaaa aaaaatgatga attgaattga aaagctgtgg tatggtgcac
241 tctcagtaca atctgctctg atgccgcata gttaagccag ccccgacacc cgccaacacc
301 cgtgacgcg ccctgacggg ctgtctgct cccggcatcc gcttacagac aagctgtgac
361 cgtctccggg agctgcatgt gtcagagggt ttaccgtca tcaccgaaac gcgcgagacg
421 aaagggcctc gtgatacgcc tattttata ggtaatgtc atgataataa tggtttctta
481 ggacggatcg ctgcctgta acttacacgc gcctcgtatc ttttaatgat ggaataattt
541 gggaatttac tctgtgttta tttatttta tgtttgtat ttggatttta gaaagtaaat
601 aaagaaggta gaagagttac ggaatgaaga aaaaaaata aacaaagggt taaaaaattt
661 caacaaaaag cgtactttac atatataatt attagacaag aaaagcagat taaatagata
721 tacattcgat taacgataag taaaatgtaa aatcacagga tttcgtgtg tggcttctta
781 cacagacaag atgaacaat tcggcattaa tacctgagag caggaagagc aagataaaag
841 gtagtatttg ttggcgatcc ccctagagtc tttacatct tcggaaaaca aaaactattt
901 tttcttaat ttctttttt actttctatt ttaatttat atatttatat taaaaaattt
961 aaattataat tattttata gcacgtgatg aaaaggacc aggtggcact ttcggggaa
1021 atgtgcgcgg aacccttatt tgtttattt tctaaataca tcaaatatg tatccgctca
1081 tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc
1141 aacatttcg tgctgccctt attcccttt ttgcggcatt ttgccttct gttttgctc
1201 acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gtttaagggc accaataact
1261 gccttaaaaa aattacgccc cgccctgcca ctcatcgag tactgttgta attcattaag
1321 cattctgccg acatggaagc catcacagac ggcatgatga acctgaatcg ccagcggcat
1381 cagcaccttg tcgcttgcg tataatattt gccatgggtg aaaacggggg cgaagaagtt
1441 gtccatattg gccacgttta aatcaaaact ggtgaaactc acccagggat tggctgagac
1501 gaaaaacata ttctcaataa acccttagg gaaataggcc aggtttcac cgtaacacgc
1561 cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt cactccagag
1621 cgatgaaaac gtttcagttt gctcatggaa aacggtgtaa caagggtgaa cactatccca
1681 taccaccagc tcaccgtctt tcattgcat acggaattcc ggatgagcat tcacaggcg
1741 ggcaagaatg tgaataaagg ccggataaaa ctgtgctta ttttcttta cgtctttta
1801 aaaggccgta atatccagct gaacggtctg gttataggtta cattgagcaa ctgactgaaa
1861 tgctcaaaa tgttctttac gatgccattg ggatatatca acggtggtat atccagtgt
1921 tttttctcc atttagctt ccttagctcc tgaaaatctc gataactcaa aaaatacgcc
1981 cggtagtgat cttatttcat tatggtgaaa gttggaacct ctacgtgcc gatcaacgtc
2041 tcattttcgc caaaagtgg cccagggtt cccggtatca acagggacac caggatttat
2101 ttattctgcg aagtgtatct ccgtcacagg tatttatctg gcgcaaagtg cgtcgggtga
2161 tgctgccaac ttactgattt agtgtatgat ggtgttttg aggtgtccta gtggttctg
2221 tttctatcag ctgtccctcc tgttcageta ctgacggggt ggtgcgtaac ggcaaaagca
2281 ccgcccggaca tcagcgctag cggagtgtat actggcttac tatgttgga ctgatgaggg
2341 tgcagtgaat gtgcttcatg tggcaggaga aaaaaggctg caccggtgcg tcagcagaag
2401 atgtgataca ggatatattc cgttctctcg ctactgact cgctacgctc ggtcgttga
2461 ctgcggcgag cggaatggc ttacgaacgg ggccggagatt tcttgaaga tgccaggaag
2521 atacttaaca ggggaagtga agggccgcgg caaagccgtt ttccatagg ctccgcccc

2581 ctgacaagca tcacgaaatc tgacgctcaa atcagtgggtg gcgaaacccg acaggactat
2641 aaagatacca ggcgtttccc cctggcggct cctcgtgcg ctctcctgtt cctgccttfc
2701 gggttacccg tgctattccg ctgttatggc cgcgtttgtc tcattccacg cctgacactc
2761 agttccgggt aggcagttcg ctccaagctg gactgtatgc acgaaccccc cgttcagtc
2821 gaccgctgcg cttatccgg taactatcgt cttgagtcca acccggaag acatgcaaaa
2881 gcaccactgg cagcagccac tggttaattga ttagaggag ttagcttga agtcatgcgc
2941 cggtaaggc taaactgaaa ggacaagttt tggtagctgc gctcctcaa gccagttacc
3001 tcggttcaaa gagttgtag ctgagagaac ctgcgaaaaa ccgccctgca aggcggtttt
3061 ttcgtttca gagcaagaga ttacgcgcag accaaaacga tctcaagaag atcatcttat
3121 taatcagata aaatatttgc tcatgagccc gaagtggcga gcccgatctt ccccatcggt
3181 gatgtcggcg atataggcg cagcaaccgc acctgtggcg ccggtgatgc cggccacgat
3241 gcgtccggcg tagaggatct gctcatgttt gacagcttat catcgatgca taatgtgcct
3301 gtcaaatgga cgaagcaggg attctgcaa ccctatgcta ctccgtcaag cgtcaattg
3361 tctgattcgt taccaattat gacaactga cggtacatc attcactttt tctcacaac
3421 cggcacggaa ctgcctcggg ctggccccgg tgcatTTTT aaataccgc gagaaataga
3481 gttgatcgtc aaaaccaaca ttgcgaccga cgggtggcgat aggcacccgg gtggtgctca
3541 aaagcagctt cgcctggctg atacgttgg cctcgcgcca gcttaagacg ctaatcccta
3601 actgctggcg gaaaagatgt gacagacgcg acggcgacaa gcaaacatgc tgtgcgacgc
3661 tggcgatcgc aaaattgctg tctgccaggt gatcgtgat gtactgaaa gcctcgcgta
3721 cccgattatc catcggtgga tggagcgact cgtaatcgc ttccatgcgc cgcagtaaca
3781 attgctcaag cagatttate gccagcagct ccgaatagcg ccttccct tggcggcggt
3841 taatgatttg cccaaacagg tcgctgaaat gcggctgggt cgcttcatcc gggcgaaaga
3901 acccgtatt ggcaaatatt gacggccagt taagccattc atgccagtag gcgcgaggac
3961 gaaagtaaac cactggtga taccattcgc gagcctccgg atgacgaccg tagtgatgaa
4021 tctctctgg cggaacagc aaaatatcac ccggtcggca acaaatct ctgcctgat
4081 tttcaccac ccccgaccg cgaatggtga gattgagaat ataacttc attcccagc
4141 gtcggtgat aaaaaaatc agataaccgt tggcctcaat cggcgtaaa cccgccacca
4201 gatgggcatt aaacgagat cccggcagca ggggatcatt ttgcgttca gccatactt
4261 tcatactccc gccattcaga gaagaaacca attgtcata ttgcatcaga cattgccgtc
4321 actgctctt ttactggctc ttctcgctaa ccaaaccggg aaccccgctt attaaaagca
4381 ttctgtaaca aagcgggacc aaagccatga caaaacgcg taacaaaagt gtctataatc
4441 acggcagaaa agtccacatt gattatttgc acggcgctac actttgctat gccatagcat
4501 tttatccat aagattagcg gatcctacct gacgttttt atcgcaactc tctactgtt
4561 ctccataccc gttttttgg gctagcgaat tcgagctcgg taccgggga tctctagag
4621 tcgacctgca ggcatgcaag cttggctgtt ttggcggtg agagaagatt ttcagcctga
4681 tacagattaa atcagaacgc agaagcggtc tgataaaaca gaattgcct ggcggcagta
4741 gcgcggtggt cccacctgac cccatgccga actcagaagt gaaacgccgt agcgcgatg
4801 gtagtgtggg gtctcccat gcgagagtag ggaactgcca ggcatcaaat aaaacgaaag
4861 gctcagtcga aagactgggc cttcgtttt atctgtgtt tgcggtgaa cgctctcctg
4921 agtaggacaa atccgccggg agcggatttg aacgttgca agcaacggcc cggagggtgg
4981 cgggcaggac gcccgccata aactgccagg catccttga gcacatccc ctttgcag
5041 ctggcgtaat agcgaagagg ccgcaccga tcgcccttc caacagttgc gcagcctgaa
5101 aggcaggccg ggccgtggtg gccacggcct ctaggccaga tccagcgga tctgggttag
5161 tcgagcgcg gccgcttccc atgtctacc agggcgagcc tgttcgca tctcagcatc
5221 tgaaatctc ccggccttgc gcttcgttg ggccttacc accgcttgg cgggcttct
5281 cgggtcaaaa ctgaacaaca gatgtgtgac cttgcgccc gtcttctgct gcgcccactc
5341 cactgtagc gggctgtgct cgttgatctg cgtcacggct ggatcaagca ctgcgaactt
5401 gaagtccttg atcgaggat accggccttc cagttgaaac cactttcgca gctggtcaat
5461 ttctatttgc cgtggccga tgctgtccca ttgcatgagc agctcgtaa gcctgatcgc
5521 gtgggtgctg tcatcttgg ccacgtcagc caaggcgtat ttggtgaact gtttggtag

5581 ttccgtcagg tacggcagca tgtctttggt gaacctgagt tctacacggc cctcaccctc
 5641 ccggtagatg attgtttgca ccagccgggt aatcatcaca ctcgggtctt tccccttgcc
 5701 attgggctct tgggttaacc ggacttcccg ccgtttcagg cgcagggccg cttctttgag
 5761 ctggtttagt gaagattcga tagggacacc cgccatcgtc gctatgtcct ccgccgtcac
 5821 tgaatacatc acttcatcgg tgacaggctc gctcctcttc acctggctaa tacaggccag
 5881 aacgatccgc tgttctgaa cactgaggcg atacgcggcc tcgaccaggg cattgtcttt
 5941 gtaaaccatt gggggtagg ccacgttcga cattccttgt gtataagggg aactgtatc
 6001 tgcgtcccac aatacaaca atccgtccct ttacaacaac aaatccgtcc cttcttaaca
 6061 acaaatccgt ccctaattgg caacaaatcc gtcccttttt aaactctaca ggccacggat
 6121 tacttgccct gtagacgtcc taaaagggtt aaaagggaaa aggaagaaaa ggggtgaaac
 6181 gcaaaaaacg caccactacg tggccccgtt ggggccgcgt ttgtcccct gaaggggagg
 6241 gggaggcgct tgggcaatcc ccgtttacc agtcccctat cgccgcctga gagggcgag
 6301 gaagcgagta atcagggtat cgaggcggtat tcacccttgg cgtccaacca gcggcaccag
 6361 cggctcgaca acccttaata taacttcgta taatgtatgc tatacgaagt tattaggtct
 6421 agagatctgt ttgcttgcc tegtccccgc cgggtcagcc ggcggttaag gtatacttc
 6481 cgctgcataa cctgtctcg gggtcattat agcgattttt tcggtatata catcctttt
 6541 cgcacgatat acaggatttt gccaaagggt tcgtgtagac ttctcttgggt gtatccaacg
 6601 gcgtcagccg ggcaggatag gtgaagtagg cccacccgcg agcgggtgtt ccttctcac
 6661 tgcccttat tcgcacctgg cgggtgctcaa cgggaatcct gctctgcgag gctggccgat
 6721 aagctccacg tgaataactg atataattaa attgaagctc taattgtga gtttagtata
 6781 catgcattta cttataatac agtttttag tttgctggc cgcattctct caaatatgct
 6841 tccagcctg ctttctgta acgttcaccc tctaccttag catcccttcc ctttgcaat
 6901 agtctcttc caacaataat aatgtcagat cctgtagaga ccacatcac cacggttcta
 6961 tactgttgac ccaatgcgc tccctgtca tctaaacca caccgggtgt cataatcaac
 7021 caatcgtaac cttcatctct tccacccatg tctctttgag caataaagcc gataacaaaa
 7081 tctttgtcgc tcttcgcaat gtcaacagta ccttagtat attctccagt agatagggag
 7141 ccttgcagtg acaattctgc taacatcaaa aggcctctag gttccttgt tacttctct
 7201 gccgctgct tcaaaccgct aacaatacct gggcccacca caccgtgtgc attcgtaagt
 7261 tctgccatt ctgctattct gtatacccc gcagagtact gcaatttgac tgtattacca
 7321 atgtcagcaa atttctgtc ttcgaagagt aaaaattgt acttgccgga taatgcctt
 7381 agcggcttaa ctgtccctc catggaaaaa tcagtcaaga tatccacatg tgttttagt
 7441 aaacaaattt tgggacctaa tgcttcaact aactccagta attccttgggt ggtacgaaca
 7501 tccaatgaag cacacaagtt tgtttgctt tcgtgcatga tattaaatag cttggcagca
 7561 acaggactag gatgagtagc agcacgttcc ttatatgtag ctttcgacat

//

SEQ ID NO 2***galE*: epimerase, *C. jejuni*****EC 5.1.3.2****987 bp ds-DNA**

1 atgaaaattc ttattagcgg tgggtcaggt tatatagggt ctcatacttt aagacaattt
 61 ttaaaaacag atcatgaaat ttgtgtttta gataatcttt ctaagggttc taaaatcgca
 121 atagaagatt tgcaaaaaat aagaactttt aaatttttg aacaagattt aagtgattt
 181 caaggcgtaa aagcattgtt tgagagagaa aaatttgacg ctattgtgca tttgcagcg
 241 agcattgaag ttttgaaag tatgcaaaac ctttaaagt attatatgaa taacactgtt
 301 aatacgacaa atctcatcga aactgtttg caaactggag tgaataaatt tataatttct
 361 tcaacggcag ccacttatgg cgaaccacaa actcccgttg tgagcgaaac aagtccttta
 421 gcacctatta atccttatgg gcgtagtaag cttatgagcg aagaggtttt gcgtgatgca
 481 agtatggcaa atcctgaatt taagcattgt attttaagat attttaatgt tgcagggtct
 541 tgcattggatt atactttagg acaacgctat ccaaagcgca ctttgcttat aaaagttgca

601 gctgaatgtg ccgcaggaaa acgtaataaa cttttcatat ttggcgatga ttatgataca
 661 aaagatggca ctgcataag agattttatc catgtggatg atatttcaag tgcgcattta
 721 tcggctttgg attatttaaa agagaatgaa agcaatgttt ttaatgtagg ttatggacat
 781 ggtttttagcg taaaagaagt gattgaagcg atgaaaaaag ttagcggagt ggattttaaa
 841 gtagaacttg cccacgccc tgcgggtgat cctagtgtat tgatttctga tgcaagtaaa
 901 atcagaaatc ttacttcttg gcagcctaaa tatgatgatt tagggcttat ttgtaaatct
 961 gcttttgatt gggaaaaaca gtgctaa

//

SEQ ID NO 3***pglB*: OST, *C. jejuni*****EC 2.4.1.119****2142 bp ds-DNA**

1 atgttgaaaa aagagtattt aaaaaaccct tatttagttt tgtttgcgat gattatatta
 61 gcttatgttt ttatgttatt ttgcagggtt tattgggttt ggtgggcaag tgagttaaat
 121 gagtattttt tcaataatca gtaaatgac atttcaaag atggctatgc ttttctgag
 181 ggcgcaagag atatgatagc aggttttcat cagcctaag atttgagttt ttatggatct
 241 tctttatccg cgcttactta ttggctttat aaaatcacac ctttttcttt tgaaagtatc
 301 attttatata tgagtacttt ttatcttct ttggtggtga ttctactat ttgctagct
 361 aacgaataca aacgtccttt aatgggcttt gtatctgctc ttttagcaag tatagcaaac
 421 agttattata atcgactat gagtgggtat tatgatacgg atatgctggt aattgttttg
 481 cctatgttta tttattttt tatggttaaga atgattttta aaaaagactt ttttcattg
 541 attgccttgc cgttatttat aggaatttat ctttgggtgt atccttcaag ttatacttta
 601 aatgtagctt taattggact tttttaatt tatacactta ttttcatag aaaagaaaag
 661 atttttata tagctgtgat ttgtcttct ctactcttt caaatatagc atggttttat
 721 caaagtgcc aatagtaat actttttgct ttattcgcct tagagcaaaa acgcttaaat
 781 tttatgatta taggaatttt aggtagtga actttgatat ttttgatttt aagtgggtgg
 841 gttgactcta tactttatca gcttaaattt tatattttta gaagtgtatg aagtgcgaat
 901 ttaacgcagg gctttatgta ttttaatgtc aatcaaacca tacaagaagt tgaaaatgta
 961 gatcttagcg aatttatgcg aagaattagt gtagtgtaa ttgtttttt gtttctttg
 1021 tttggttttg tatggctttt gagaaaacat aaaagtatga ttatggcttt acctatattg
 1081 gtgcttgggt ttttagcctt aaaagggggg cttagattta ccatttattc tgtacctgta
 1141 atggccttag gatttggttt ttattgagc gagttaagg ctataatggt taaaaaatat
 1201 agccaattaa ctcaaatgt ttgtattgtt ttgcaacta tttgacttt agctccagta
 1261 tttatccata ttacaacta taaagcgcca acagttttt ctcaaatga agcatcatta
 1321 ttaaatcaat taaaaaatat agccaataga gaagattatg tggttaactg gtgggattat
 1381 ggttatcctg tgcgttatta tagcgatgtg aaaactttag tagatggtgg aaagcattta
 1441 ggtaaggata attttttccc ttcttttgc ttaagcaaag atgaacaagc tgcagctaat
 1501 atggcaagac ttagtgtaga atatacagaa aaaagctttt atgctccgca aaatgatatt
 1561 ttaaaaacag acattttgca agccatgatg aaagattata atcaaagcaa tgtggatttg
 1621 tttctagctt cattatcaaa acctgatttt aaaatcgata cgccaaaaac tcgtgatatt
 1681 tatctttata tgcccgttag aatgtctttg atttttcta cgggtggctag ttttcttt
 1741 attaatntag atacaggagt ttggataaa cttttacct ttgacacgc ttatccactt
 1801 gatgttaaaa atggagaaat ttatcttagc aacggagtgg ttttaagcga tgattttaga
 1861 agttttaaaa tagtgataaa tgtggtttct gtaaatagta tcgtagagat taattctatt
 1921 aaacaagggt aatacaaat cactccaatt gatgataagg ctcaatttta tttttttat
 1981 ttaaaggata gtgctattcc ttacgcacaa tttattttaa tggataaaac catgtttaat
 2041 agtgcttatg tgcaaatgtt ttttttagga aattatgata agaatttatt tgacttggtg
 2101 attaatctta gagatgctaa ggtttttaaa cttaaaattt aa

//

SEQ ID NO 4***pglA*: α 1,3-N-acetylgalactosamine transferase****EC 2.4.1.-****1131 bp ds-DNA**

```
1 atgagaatag gattttatc acatgcagga gcaagtattt atcattttag aatgcctatt
61 ataaaagcat taaaagatag aaaagatgaa gtttttgta tagtgccgca agatgaatac
121 acgcaaaaac ttagagatct tggtttaaaa gtaattgttt atgagtttc aagagctagt
181 ttaaatcctt tttagtttt aaagaatttt tttatcttg ctaaggtttt aaaaaattta
241 aatcttgatc ttattcaaag tgcggcacac aaaagcaata ctttggaat tttagcggca
301 aaatgggcaa aaattcctta tcgttttgct ttggtagaag gcttgggac ttttatata
361 gatcaagggt ttaaggcaaa tttagtacgt ttgttatta ataacttta taaattaagt
421 tttaaatttg cacaccaatt tttttgtc aatgaaagta atgccgagtt tatgcggaat
481 ttaggactta aggaaaataa aatttgtgtg ataaaatccg tagggatcaa tttaaaaaaa
541 ttttctcta tttataga atcggaaaaa aaagagcttt ttggagaaa tttaaatata
601 gataaaaaac ctattgttct tatgatagca agagctttat ggcataaagg tgtaaaagaa
661 tttatgaaa gtgctactat gctaaaagac aaagcaaatt ttgtttagt tgggtgaaga
721 gatgaaaatc cttctgtgc gagtttgag ttttaaaact cgggtgtggt gcattatttg
781 ggtgctagaa gtgatagat cgagctttg caaattgtg atattttgt ttaccaagc
841 tataaagaag gcttctctgt aagtgtttg gaggcaaaag cttgtggcaa ggctatagtg
901 gtgagtgatt gtgaagggtg ttagaggct atttctaag cttatgatgg actttgggca
961 aaaacaaaa atgctaagga ttaagcgaa aaaattcac tttattaga agatgaaaaa
1021 ttaagattaa attagctaa aaatgctgcc caagatgctt tacaatacga tgaataaat
1081 atcgcacagc gttatttaaa actttatgat agggtaatta agaattgatg a
```

//

SEQ ID NO 5***whnJ*: β 1,3 galactosyl transferase****EC 2.4.1.307****765 bp ds-DNA**

```
1 atgtcattga gaatattaga tatgattca gtaataatgg ctgtacaccg atatgataaa
61 tatgttgata ttcaattga tagtatctta aatcagacat actctgactt tgagttaata
121 ataattgcaa atggagggga ttgttcgag atagcaaac agctgaagca ttatacagag
181 ctggataaca gagttaaaat ttatacatta gaaatagggc agttatcgtt tgcattaaat
241 tacgcagtaa ctaagtgtaa atactctatt attgccagaa tggattccga cgatgtttca
301 ctgccgttac gtctagaaaa acaatatatg tatatgttg agaattgatt agaaatgggtg
361 gggactggga tcagacttat caatgaaaac ggtgagtta taaagaatt aaaatatcca
421 aatcataata agataaataa gatacttct tttaaaatt gtttgcgca tctactttg
481 atgttcaaga aagatgttat actaaagcag cgaggttatt gtggtggtt taattcagaa
541 gattatgac tatggctcag aatcttaaat gaatgtccga atatacgtg ggataatcta
601 agtgagtgtt tgctaaatta tcgaattcat aacaaatcta cgaaaaatc agcactcgca
661 tattatgaat gtgctagtta ttctctgca gaattcttaa aaaaaagaac tattacgaat
721 tttcttctt gcctctatca ttttgtaaa gcactataa aataa
```

//

SEQ ID NO 6**pTRC99Y****6866 bp ds-DNA**

```
1 gttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc
```

61 ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc
121 gcaactcccg tctggataat gtttttgcg ccgacatcat aacggttctg gcaaatattc
181 tgaatgagc gtgtgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcgga
241 taacaatttc acacaggaaa cagaccatgg aattcgagct cggtaaccgg ggatcctcta
301 gagtcgacct gcaggcatgc aagcttggtt gttttggcgg atgagagaag attttcagcc
361 tgatacagat taaatcagaa cgcagaagcg gtctgataaa acagaatttg cctggcggca
421 gtacgcgggt ggtccacct gaccccatgc cgaactcaga agtgaaacgc cgtagcggcg
481 atggtagtgt ggggtctccc catgcgagag tagggaactg ccaggcatca aataaaacga
541 aaggctcagt cgaaagactg ggctttcgt ttatctgtt gttgtcggg gaacgctctc
601 ctgagtagga caaatccgcc gggagcggat ttgaacgtt cgaagcaacg gcccgagggg
661 tggcggggcag gacggccgcc ataaactgcc aggcatacaa ttaagcagaa ggccatctg
721 acggatggcc ttttgcgtt tctacaaact cttttgttt atttttctaa atacattcaa
781 atatgtatcc gctcatgaga caataacct gataaatgt tcaataatat tgaaaaagga
841 agagtatgag tattcaacat ttcgtgtcg ccctattcc ctttttgcg gcattttgcc
901 ttctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg
961 gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt gagagttttc
1021 gcccgaaga acgttttcca atgatgagca cttttaaagt tctgctatgt ggcgcggat
1081 taccctgtt tgacggcggg caagagcaac tcggtcgccg catacactat tctcagaatg
1141 acttggttga gtactacca gtcacagaaa agcatcttac ggatggcatg acagtaagag
1201 aattatgcag tctgccata accatgagt ataactgc ggccaactta cttctgacaa
1261 cgatcggagg accgaaggag ctaaccgctt tttgcacaa catgggggat catgtaactc
1321 gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag cgtgacacca
1381 cgatgcctac agcaatggca acaacgttgc gaaactatt aactggcgaa ctacttactc
1441 tagcttcccg gcaacaatta atagactgga tggagggcga taaagttgca ggaccattc
1501 tgcgctcggc cttccgggt ggctggttta ttgctgataa atctggagcc ggtgagcgtg
1561 ggtctcggg tatcattgca gactggggc cagatggtaa gccctccgt atcgtagtta
1621 tctacagac ggggagtcag gcaactatgg atgaacgaaa tagacagatc gctgagatag
1681 gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat atactttaga
1741 ttgatttaa acttcattt taatttaaa ggatctaggt gaagatcctt ttgataatc
1801 tcatgaccaa aatcccttaa cgtgagttt cgtccactg agcgtcagac cccgtagaaa
1861 agatcaaagg atctcttga gatectttt ttctgcgct aatctgctg ttgcaaaca
1921 aaaaaccacc gtaccagcg gtggtttgt ttccggatca agagctacca actcttttc
1981 cgaaggtaac tggcttcagc agagcgcaga taccaatac tgccttcta gttagccgt
2041 agttaggcca ccactcaag aactctgtag caccgctac atacctcgt ctgctaatcc
2101 tgttaccagt ggtgctgcc agtggcgata agtcgtgtt taccgggtt gactcaagac
2161 gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca
2221 gcttgagcgc aacgacctac accgaactga gatactaca gcgtgagcta tgagaaagcg
2281 ccacgcttcc cgaagggaga aaggcggaca ggtatccgt aagcggcagg gtcggaacag
2341 gagagcgcac gagggagctt ccagggggaa acgectgta tctttatagt cctgtcgggt
2401 ttgccacct ctgacttgag cgtcgattt ttgtatgctc gtcagggggg cggagcctat
2461 ggaaaaacgc cagcaacgcg gccttttac ggttctggc cttttgctgg cttttgctc
2521 acatgttctt tctgcgtta tcccgtgatt ctgtggataa ccgtattacc gcctttgagt
2581 gagtgatac cgctcgccg agccgaacga ccgagcgcag cgagtcagt agcgaggaag
2641 cggaagagcg cctgatcgg tattttctcc ttacgcatc gtgcggtatt tcacaccgca
2701 tatgttgaag ctctaattg tgagtttagt atacatgcat ttacttataa tacagtttt
2761 tagtttctt ggccgcatc tctcaaatat gcttccagc ctgctttct gtaacgttca
2821 cctctacct tagcatcct tcccttgca aatagctc ttccaacaat aataatgtca
2881 gatcctgtag agaccacatc atccacggt ctatactgtt gacccaatgc gtctccctg
2941 tcatctaac ccacaccggg tgcataatc aaccaatcgt aacctcacc tcttccacc
3001 atgtctctt gagcaataaa gccgataaca aaatctttgt cgctcttcgc aatgtcaaca

3061 gtacccttag tatattctcc agtagatagg gagcccttgc atgacaatc tgctaacatc
3121 aaaaggcctc taggttctt tgttacttct tctgccgcct gcttcaaacc gctaacaata
3181 cctgggcccac ccacaccgtg tgcattcgta atgtctgccc attctgctat tctgtataca
3241 cccgcagagt actgcaattt gactgtatta ccaatgtcag caaattttct gtcttcgaag
3301 agtaaaaaat tgtacttggc ggataatgcc tttagcggct taactgtgcc ctccatggaa
3361 aaatcagtca agatatccac atgtgtttt agtaaacaaa tttgggacc taatgcttca
3421 actaactcca gtaattcctt ggtggtacga acatccaatg aagcacacaa gtttgtttgc
3481 ttttcgtgca tgatattaaa tagcttggca gcaacaggac taggatgagt agcagcacgt
3541 tccttatatg tagctttcga catgatttat ctctgttcc tgcaggttt tgttctgtgc
3601 agttgggfta agaatactgg gcaatttcat gtttctcaa cactacatat gcgtatatat
3661 accaatctaa gtctgtgctc ctctctctgt tcttcttct gtctggagat taccgaatca
3721 aaaaaatttc aaagaaaccg aaatcaaaaa aaagaataaa aaaaaaatga tgaattgaat
3781 tgaaaaactg tggatggtg cactctcagt acaatctgct ctgatgccgc atagttaagc
3841 cagcccccac acccgccaac acccgctgac gcgccctgac gggttgtct gctcccggca
3901 tccgcttaca gacaagctgt gaccgtctcc gggagctgca tgtgtcagag gtttcaccg
3961 tcataccga aacgcgcgag acgaaagggc ctctgtatgc gcctatttt ataggftaat
4021 gtcatgataa taatggttct ttagtatgat ccaatatcaa aggaaatgat agcattgaag
4081 gatgagacta atccaattga ggagtggcag catatagaac agctaaaggg tagtgctgaa
4141 ggaagcatat gatccccgc atggaatggg ataatatcac aggaggtact agactacct
4201 tcactctaca taaatagacg catataagta cgcatttaag cataaacacg cactatgccg
4261 ttcttctcat gtatatatat atacaggcaa cacgcagata taggtgcgac gtgaacagt
4321 agctgtatgt gcgcagctcg cgttgcattt tcggaagcgc tcgttttcgg aaacgcttg
4381 aagttctcat tccgaagtc ctattctcta gaaagtatag gaactcaga gcgctttga
4441 aaacccaaag cgtctgaag acgcatttc aaaaaacca aaacgcaccg gactgtaacg
4501 agctactaaa atattgcgaa taccgcttcc acaaacattg ctcaaaagta tctcttgc
4561 atatatctct gtgctatate cctatataac ctacccatcc accttcgct ccttgaactt
4621 gcatctaaac tcgacctcta cattttttat gtttatctct agtattactc ttagacaaa
4681 aaaattgtag taagaactat tcataagagt aatcgaaaac aatacgaaaa tgtaaactat
4741 tcctatacgt agtatataga gacaaaatag aagaaaccgt tcataatttt ctgaccaatg
4801 aagaatcatc aacgctatca ctttctgttc acaaagtatg cgcaatccac atcggtatag
4861 aatataatcg gggatgcctt tatcttgaaa aaatgcaccc gcagcttcgc tagtaatcag
4921 taaacgcggg aagtggagtc aggcttttt tatggaagag aaaatagaca ccaaagtagc
4981 ctcttcttaa ccttaacgga cctacagtgc aaaaagtat caagagactg cattatagag
5041 cgcacaaaag agaaaaaag taatctaaga tgccttgta gaaaaatagc gctctcgga
5101 tgcatttttg tagaacaaaa aagaagtata gattcttgt tggtaaaata gcgctctgc
5161 gttgcatttc tgttctgtaa aaatgcagct cagattcttt gttgaaaaa ttagcgctct
5221 cgcgttgcatt tttgtttta caaaaatgaa gcacagattc ttcgttggtg aaatagcgt
5281 ttcgcttgc atttctgttc tgtaaaaatg cagctcagat tcttgtttg aaaaattagc
5341 gctctcgcgt tgcattttg ttctacaaaa tgaagcacag atgcttcgtt cagggtgcac
5401 tctcagtaca atctgtctg atgccgata gttagccag tataactcc gctatcgta
5461 cgtgactggg tcatggctgc gcccgcacac ccgccaacac ccgtgacgc gcctgacgg
5521 gcttgtctgc tcccggcatc cgttacaga caagctgtga ccgtctccg gagctgcatg
5581 tgcagaggt ttcaccgtc ataccgaaa cgcgcgaggc agcagatcaa ttcgcgcgcg
5641 aaggcgaagc ggcattgcat tacgttgaca ccatcgaatg gtgcaaaacc tttcgcggtg
5701 tggcatgata gcgcccggaa gagagtcaat tcagggtggt gaatgtgaaa ccagtaacgt
5761 tatacatgt cgcagagtat gccggtgtct ctatcagac cgttcccg gcgttgaaac
5821 aggccagcca cgttctgctg aaaacgcggg aaaaagtga agcggcgatg gcggagctga
5881 attacattcc caaccgctg gcacaacaac tggcgggcaa acagtcgttg ctgattggcg
5941 ttgccacctc cagtctggcc ctgcacgcgc cgtcgcaaat tgcgcggcg attaaatctc
6001 gcgccgatca actgggtgcc agcgtggtg tgcgatggt agaacgaagc ggcgtcgaag

6061 cctgtaaagc ggcggtgcac aatcttctcg cgcaacgcgt cagtgggctg atcattaact
6121 atccgctgga tgaccaggat gccattgctg tggaagctgc ctgcactaat gttccggcgt
6181 tatttctga tgtctctgac cagacacca tcaacagtat tatttctcc catgaagacg
6241 gtacgcgact gggcgtggag catctggtcg cattgggtca ccagcaaate gcgctgttag
6301 cgggcccatt aagttctgtc tcggcgctgc tgcgtctggc tggctggcat aaatatctca
6361 ctgcgaatca aattcagccg atagcggaaac gggaaggcga ctggagtgcc atgtccgggt
6421 ttcaacaaac catgcaaatg ctgaatgagg gcatcgttcc cactgcgatg ctggttgcca
6481 acgatcagat ggcgctgggc gcaatgcgcg ccattaccga gtccgggctg cgcgttggtg
6541 cggatatctc ggtagtggga tacgacgata ccgaagacag ctcatgttat atcccgcgt
6601 caaccacat caaacaggat ttccgctgc tggggcaaac cagcgtggac cgttctgctc
6661 aactctctca gggccaggcg gtgaagggca atcagctgtt gccctctca ctggtgaaa
6721 gaaaaaccac cctggcgccc aatacgcaaa ccgectctcc ccgcgcgttg gccgattcat
6781 taatgcagct ggcacgacag gttcccgac tggaaagcgg gcagtgcgcg caacgaatt
6841 aatgtgagtt agcgcgaatt gatctg

//

SEQ ID NO 7**MBP-hGH****1791 bp ds-DNA**

1 atgaaaaaga ttggctggc gctggctggt ttagtttag cgttagcgc atcggcgtct
61 agaaaaatcg aagaaggtaa actggtaatc tggattaacg gcgataaagg ctataacggt
121 ctgctgaag tcgtaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat
181 ccgataaac tggaaagaaa attccacag gttgcggcaa ctggcgatgg ccctgacatt
241 atcttctggg cacacgaccg ctttgggtgc tacgtcaat ctggcctgtt ggctgaaatc
301 accccggaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac
361 aacggcaagc tgattgtta cccgatcgt gttgaagcgt taccgtgat ttataacaaa
421 gatctgtcgc cgaacccgcc aaaaacctgg gaagagatcc cggcgctgga taaagaactg
481 aaagcgaaag gtaagagcgc gctgatgttc aacctgcaag aaccgtactt cacctggccg
541 ctgattgctg ctgacggggg ttatcgcttc aagtatgaaa acggcaagta cgacattaaa
601 gacgtgggcg tggataacgc tggcgcgaaa gcgggtctga cttcctggt tgacctgatt
661 aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc cttaataaaa
721 ggcgaaacag cgatgacat caacggcccg tgggcatggt ccaacatcga caccagcaaa
781 gtgaattatg gtgtaacggt actgccgacc tcaaggggtc aaccatccaa accgttcgtt
841 ggcgtgctga gcgcaggtat taacgccgcc agtccgaaca aagagctggc gaaagagttc
901 ctgaaaact atctctgac tgatgaagg tggaaagcgg ttaataaaga caaacgctg
961 ggtgccgtag cgctgaagtc ttacaggaa gagttggcga aagatccacg tattgccgcc
1021 accatgaaa acgcccagaa aggtgaaatc atgccgaaca tcccgcagat gtccgtttc
1081 tggatgccg tgcgtactgc ggtgatcaac gccgccagcg gtcgtcagac tgcgatgaa
1141 gccctgaaag acgcgcagac tcgtatcacc aagcatcacc atcatcaca tggcgaaaac
1201 ctgtatttcc agggcttccc aaccattccc ttatccaggc ttttgacaa cgctatgctc
1261 cgcgcccgtc gcctgtacca gctggcatat gacacctatc aggagttga agaagcctat
1321 atcctgaagg agcagaagta ttattctctg cagaaccccc agacctccct ctgcttctca
1381 gagtctatc caacacctc caacagggtg aaaacgcagc agaaatctaa cctagagctg
1441 ctccgatct cctgctgct catccagtca tggctggagc cgtgcagct cctcaggagc
1501 gtcttcgcca acagcctggt gtatggcgcc tcggacagca acgtctatcg ccacctgaag
1561 gacctagagg aaggcatcca aacgctgatg tggaggctgg aagatggcag ccccgaggat
1621 gggcaggatc agaacgccac gtacagcaag ttgacacaa aatcgacaaa cgatgacgca
1681 ctgtcaaga actacgggct gctctactgc ttcagggaagg acatggacaa ggtcgagaca
1741 ttctgcgca tcgtgcagtg ccgtctgtg gagggcagct gtggcttcta a

//

SEQ ID NO 8

neuB: N-acetylneuraminate synthase**EC 2.5.1.56****1041 bp ds-DNA**

```
1 atgagtaata tatatategt tgctgaaatt ggttgcaacc ataatggtag tgttgatatt
61 gcaagagaaa tgatattaaa agccaaagag gccggtgtta atgcagtaaa attccaaaca
121 tttaaagctg ataaattaat ttcagctatt gcacctaagg cagagtatca aataaaaaac
181 acaggagaaat tagaatctca gttagaaatg acaaaaaagc ttgaaatgaa gtatgacgat
241 tatctccatc taatggaata tgcagtcagt ttaaatttag atgtttttc taccctttt
301 gacgaagact ctattgattt ttagcatctt ttgaaacaaa aaatatggaa aatcccttca
361 ggtgagttat tgaatttacc gtatcttgaa aaaatagcca agcttccgat cctgataag
421 aaaataatca tatcaacagg aatggctact attgatgaga taaaacagtc tgtttctatt
481 ttataaata ataaagtcc ggttgataat attacaatat tacattgcaa tactgaatat
541 ccaacgccct ttgaggatgt aaaccttaat gctattaatg attgaaaaa acacttcct
601 aagaataaca taggtctc tgatcattct agcgggtttt atgcagctat tgcggcgggtg
661 ccttatggaa taactttat tgaaaaacat ttcactttag ataaatctat gtctggccca
721 gatcatttgg cctcaataga acctgatgaa ctgaaacatc tatgtattgg ggtcaggtgt
781 gttgaaaaat cttaggttc aaatagtaaa gtggttacag cttcagaaaag gaagaataaa
841 atcgtagcaa gaaagtctat tatagctaaa acagagataa aaaaaggatga ggtttttca
901 gaaaaaata taacaacaaa aagacctggt aatggtatca gtccgatgga gtggtataat
961 ttattgggta aaattgcaga gcaagacttt attccagatg aattaataat tcatagcgaa
1021 ttcaaaaatc agggggaata a
```

//

SEQ ID NO 9**neuA: N-acetylneuraminate cytidyltransferase****EC 2.7.7.43****1257 bp ds-DNA**

```
1 atgagaacaa aaattattgc gataattcca gcccgtagtg gatctaaagg gttgagaaat
61 aaaaatgctt tgatgctgat agataaacct cttcttgctt atacaattga agctgccttg
121 cagtcagaaa tgtttgagaa agtaattgtg acaactgact ccgaacagta tggagcaata
181 gcagagtcat atgggtgctga tttttgctg agaccggaag aactagcaac tgataaagca
241 tcatcatttg aattataaaa acatgcgtta agtatatata ctgattatga gaactttgct
301 ttattacaac caacttcacc ctttagagat tcgaccata ttattgaggc tgtaaagtta
361 tatcaaaact tagaaaaata ccaatgtgtt gtttctgtta ctagaagcaa taagccatca
421 caaataatta gaccattaga tgattactcg aactgtctt ttttgacct tgattatag
481 aaatataatc gaaactcaat agtagaatat catccgaatg gagctatatt tatagcta
541 aagcagcatt atcttcatac aaagcatttt ttggtcgtc attcactagc ttatattatg
601 gataaggaaa gctctttaga tatagatgat agaattggatt tcgaacttgc aattaccatt
661 cagcaaaaaa aaaatagaca aaaatactt tatcaaaaca tacataatag aatcaatgag
721 aaacgaaatg aatttgatag tgtaagtgat ataacttta ttggacactc gctgtttgat
781 tattgggacg taataaaaaa aatgatata gaagttaata acttaggtat cgctgggata
841 aactcgaagg agtactatga atatattatt gagaagagc ggattgttaa ttcggagag
901 tttgttttca tcttttttgg aactaatgat atagttgtta gtgattggaa aaaagaagac
961 acattgtggt atttgaagaa aacatgccag tatataaaga agaaaaatgc tgcatacaaa
1021 atttatttat tgcggttcc tctgtttt gggcgtattg atcgagataa tagaataatt
1081 aatgatttaa attctatct tcgagagaat gtagattttg cgaagttat tagcttgat
1141 cagtttttaa aagactctta tggcaatcta aataaaatgt atacttatga tggcttacat
1201 tttaatagta atgggtatc agtattagaa aacgaaatag cggagattgt taaatga
```

//

SEQ ID NO 10**neuC: UDP-N-acetylglucosamine 2-epimerase****EC 5.1.3.14****1176 bp ds-DNA**

```
1 atgaaaaaaa tattatacgt aactggatct agagctgaat atggaatagt tcggagactt
61 ttgacaatgc taagagaaac tccagaaata cagcttgatt tggcagttac aggaatgcat
121 tgtgataatg cgtatggaaa tacaatacat attatagaac aagataattt taatattatc
181 aagggtgtgg atataaatat caatacaact tcacatactc acattctcca ttcaatgagt
241 gtttgccctc attcgtttgg tgatttttt tcaataaca catatgatgc ggttatgggt
301 ttaggcgata gatatgaaat atttcagtc gctatcgag catcaatgca taatattcca
361 ttaattcata ttcattggtg tgaaaagaca ttgctaatt atgatgagtt tattaggcat
421 tcaattacta aatgagtaa actccatctt acttctacag aagagtataa aaaacgagta
481 attcaactag gtgaaaagcc tggtagtgtg ttaatatgt gttctcttgg tgcagaaaat
541 gctctttcat tgcatttacc aaataagcag gagttggaac taaaatatgg ttcactgtta
601 aaacggctact ttgtgtagt attccatcct gaaacacttt ccacgcagtc ggttaatgat
661 caaatagatg agttattgtc agcgatttct tttttaaaa atactcacga cttattttt
721 attggcagta acgctgacac tggttctgat ataattcaga gaaaagtaaa atatttttgc
781 aaagagtata agttcagata ttgattttct attcgttcag aagattattt ggcaatgatt
841 aaatgctctt gtgggcta at tgggaactcc tcctctggtt taattgaggt tccatcttta
901 aaagtgtcaa caattaacat tggatagagg cagaaaggcc gtgttcgtgg agccagtgtg
961 atagatgtac ccgttgaaaa aaatgcaatc gtcagaggga taaatatatc tcaagatgaa
1021 aaatttatta gtgtgtaca gtcactagt aatccttatt ttaaagaaaa tgctttaatt
1081 aatgctgtta gaattattaa ggattttatt aaatcaaaaa ataaagatta caaagatttt
1141 tatgacatcc cggaatgtac caccagttat gactag
```

//

SEQ ID NO 11**neuD: sialic acid biosynthesis protein, acetyltransferase family****EC 2.3.1.45****624 bp ds-DNA**

```
1 atgagtaaaa aattaataat atttggcg ggtggtttt caaaatctat aattgacagc
61 ttaaatcata aacattacga gtaatagga ttatcgata aatataaaag tggttatcat
121 caatcatatc caatattagg taatgatatt gcagacatcg agaataagga taattattat
181 tattttattg ggataggcaa accatcaact aggaagcact atttaaacad cataagaaaa
241 cataatctac gcttaattaa cattatagat aaaactgcta ttctatcacc aatattata
301 ctgggtgatg gaattttat tggtaaaatg tgtatactta accgtgatac tagaatacat
361 gatgccgttg taataaatac taggagtta attgaacatg gtaatgaaat aggctgctgt
421 agcaatatct ctactaatgt tgtactta at ggtgatgttt ctgttgaga agaaactttt
481 gttggtagct gtactgttgt aaatggccag ttgaagctag gctcaaagag tattattggt
541 tctgggtcgg ttgtaattag aaatatacca agtaatgttg tagttgctgg gactccaaca
601 agattaatta gggggaatga atga
```

//

SEQ ID NO 12**lst: 2,3 NeuNAc transferase*****N. meningitidis*****EC 2.4.99.4****1116 bp ds-DNA**

```
1 atgggcttga aaaaggcttg ttgaccgtg ttgtgttga ttgtttttg tttcgggata
61 ttttatacat ttgaccgggt aaatcagggg gaaaggaatg cggttccct gctgaaggag
```

121 aaacttttca atgaagaggg ggaaccgggc aatctgattt tctgttatac catattgcag
181 atgaaggtgg cggaaaggat tatggcgag catccgggag agcgggttta tgtggtgctg
241 atgtctgaaa acaggaatga aaaatacgat tattatttca atcagataaa ggataaggcg
301 gagcgggagc actttttcca cctgccctac gggttgaaca aatcggttaa ttcatccg
361 acgatggcgg agctgaaggt aaagtcgatg ctgctgccga aagcaagcg gattatttg
421 gcaagtttgg aaaaagtcag cattgccgcc ttttgagca cttaccgga tgcggaaatc
481 aaaaccttg acgacgggac aggcaattta attcaaagca gcagctattt gggcgatgag
541 tttctgtaa acgggacgat caagcggaaat ttgcccggga tgatgatcgg agattggagc
601 atcgccaaaa ccgcaatgc tccgacgag cattacacga tattcaaggg ttgaaaaac
661 attatggacg acggccgccc caagatgact tacctgccgc tgttcgatgc gtccgaactg
721 aagacggggg acgaaacggg cggcacggtg cggatacttt tgggttcgcc cgacaaagag
781 atgaaggaaa ttccgaaaaa ggccggcaaaa aactcaaaa tacaatatgt cgcgccgat
841 ccccgccaaa cctacgggct tccggcgta accacattaa attcgcccta tgtcatgaa
901 gactatattt tgcgcgagat taagaaaaac ccgcatacga ggtatgaaat ttatacttt
961 ttcagcggcg cggcgttgac gatgaaggat ttcccaatg tgcacgttta cgcattgaaa
1021 ccggcttccc tccggaaga ttattggctc aagccggtgt atgccctgtt tacccaatcc
1081 ggcacccga tttgacatt tgacgataaa aattaa

//

SEQ ID NO 13**MBP-3TEV-Glucagon-4XGlycTag-6X-His****1416 bp ds-DNA**

1 atgaaaaaga tttggctggc gctggctggt ttagtttag cgttagcgc atcggcgtc
61 agaaaaatcg aagaaggtaa actggtaatc tggattaacg gcgataaagg ctataacggt
121 ctcgctgaag tcgtaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat
181 ccggataaac tggaagagaa attccacag gttgcggcaa ctggcgatgg ccctgacatt
241 atcttctggg cacacgaccg ctttggtggc tacgtcaat ctggcctgtt ggctgaaatc
301 acccgggaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac
361 aacggcaagc tgattgctta ccgatcgtt gttgaagcgt tatcgtgat ttatacaaa
421 gatctgtcgc cgaacccgcc aaaaacctgg gaagagatcc cggcgtgga taaagaactg
481 aaagcgaaag gtaagagcgc gctgatgtc aacctgcaag aaccgtactt cacctggcgc
541 ctgattgctg ctgacggggg ttatgcgttc aagtatgaaa acggcaagta cgacattaaa
601 gacgtgggagc tggataacgc tggcgcgaaa gcgggtctga cttcctggt tgacctgatt
661 aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc cttaataaa
721 ggcgaaacag cgatgacat caacggcccgc tgggcatggt ccaacatcga caccagcaaa
781 gtgaattatg gtgaacggg actgccgacc tcaagggc aaccatcaa accgttcgtt
841 ggctgtctga gcgcaggtat taacggccgc agtcgaaca aagagctggc gaaagagttc
901 ctcgaaaact atctgtgac tgatgaaggt ctggaagcgg ttaataaaga caaacctg
961 ggtgccgtag cgctgaagtc ttacaggaa gagttggcga aagatccacg tattgccgc
1021 accatggaaa acgcccagaa aggtgaaatc atgccgaaca tcccgcagat gtccgtttc
1081 tggatatccg tgcgtactgc ggtgatcaac gccgccagcg gtcgtcagac tgcgatgaa
1141 gccctgaaag acgcgacagc tcgtatcacc aaggaaaacc tgtatttca gggcgaaaa
1201 ctgtatttcc agggcgaaaa cctgtattt cagggccact cacaggcacc attcaccagt
1261 gactacagca agtacctgga ctccaggcgt gccaggatt tcgtgcagtg gctgatgaat
1321 accaagagag atcagaacgc gaccgatcag aacgcgaccg atcagaacgc gaccgatcag
1381 aacgcgaccg tcgaccatca ccatcatcac cattaa

//

SEQ ID NO 14

MBP-3TEV-Glucagon-1XGlycTag-6X-His**1371 bp ds-DNA**

1 atgaaaaaga tttggctggc gctggctggt ttagtttag cgttagcgc atcggcgtct
 61 agaaaaatcg aagaaggtaa actggtaac tggattaacg gcgataaagg ctataacggt
 121 ctcgctgaag tcggaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat
 181 ccggataaac tggaagagaa attccacag gttgcggcaa ctggcgatgg ccctgacatt
 241 atcttctggg cacacgaccg ctttgggtgc tacgtcaat ctggcctgtt ggctgaaatc
 301 accccggaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac
 361 aacggcaagc tgattgctta cccgatcgt gttgaagcgt tatcgtgat ttataacaaa
 421 gatctgtgc cgaaccgcc aaaaacctgg gaagagatcc cggcgctgga taaagaactg
 481 aaagcgaaag gtaagagcgc gctgatgttc aacctgcaag aaccgtactt cacctggccg
 541 ctgattgctg ctgacggggg ttatgcgttc aagtatgaaa acggcaagta cgacattaaa
 601 gacgtgggcg tggataacgc tggcgcgaaa gcgggtctga cttcctggt tgacctgatt
 661 aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc cttaataaaa
 721 ggcgaaacag cgatgacat caacggcccc tgggcatggt ccaacatcga caccagcaaa
 781 gtgaattatg gtgaacggt actgccgacc ttcaagggtc aaccatcaa accgttcgtt
 841 ggctgtctga gcgcaggtat taacgccgcc agtccgaaca aagagctggc gaaagagttc
 901 ctcgaaaact atctgtgac tgatgaaggt ctggaagcgg ttaataaaga caaacctgtg
 961 ggtgccgtag cgctgaagtc ttacaggaa gagttggcga aagatccacg tattgccgcc
 1021 accatggaaa acgccagaa aggtgaaatc atgccgaaca tcccgcagat gtccgtttc
 1081 tggatgccg tgcgtactgc ggtgatcaac gccgccagcg gtcgtcagac tgcgatgaa
 1141 gccctgaaag acgcgcagac tcgtatcacc aaggaaaacc tgtattttca gggcgaaaa
 1201 ctgtatttc agggcgaaaa cctgtattt cagggccact cacaggcac attcaccagt
 1261 gactacagca agtacctgga ctccaggcgt gccaggatt tcgtgcagt gctgatgaat
 1321 accaagagag atcagaacgc gaccgtcgc catcaccatc atcaccatta a

//

SEQ ID NO 15*Pdst6: α2,6 sialyltransferase**P. damsela***EC 2.4.99.1****2028 bp ds-DNA linear**

1 atgaagaaaa tactgacagt tctatctatt ttattcttt cagcgtgtaa tagcgacaat
 61 accagcctga aagagactgt tagcagcaat tcagcggatg ttgtggaaac cgaaacttat
 121 caactgacgc cgatcgatgc tccttcttcg ttctgagcc attcttggga acagacctgt
 181 ggtacaccaa ttctgaacga gtccgacaaa caggccattt cttcgtattt tgttgccccg
 241 gaactgaaac aagacgagaa atattgcttc acctcaaag gcattaccgg tgatcatcgt
 301 tatacacga acaccactct gactgtcgtg gcaccgacac tggaagtgtg tatcgacct
 361 gccagcctgc ctagtctgca gcaactgatc catattatcc aggcgaaaga cgaatatccg
 421 agcaaccagc gttttgtgag ctggaaacgt gttactgtgg atcccgacaa cgccaataaa
 481 ctgaacattc acacctatcc tctgaaaggc aataacacca gccctgagat ggtagcggcg
 541 attgatgagt atgcccagag caaaaaccgt ctgaacattg agttctatac caatacggcc
 601 cacgtgttta ataacctgcc gccaatcatt caacctctgt ataacaacga gaaagtgaaa
 661 atcagccaca ttctgctgta tgatgatggc agtagcgagt atgttagcct gtatcagtgg
 721 aaagacaccc cgaataaaat cgagactctg gaggggtgaag ttctctgct ggccaactat
 781 ctggccggtg caagtctga tgcctcgaaa gggatgggta accgctataa ttggcacaaa
 841 ctgtatgaca ccgactatta tttctgcgc gaggattatc tggacgtgga agccaatctg
 901 catgatctgc gcgattatct gggttctagc gccaaacaaa tgccgtggga tgaatttgc
 961 aaactgtccg attctcagca aacctgttc ctggacatcg ttggctttga taaagagcag

1021 ctgcaacagc agtatagcca gtcaccgctg ccgaacttca ttttactgg caccaccaca
 1081 tgggcagggg gtgagacaaa agagtattat gctcaacaac aggtgaacgt catcaacaat
 1141 gccattaacg aaacctcccc atattatctg ggtaaagact atgacctgtt ctttaaaggc
 1201 catccggctg gaggagtgtat taatgatatt atcctgggct ccttctctga catgattaac
 1261 attccggcga aaatctcatt tgaggtgctg atgatgactg atatgctgcc ggataccgtt
 1321 gctggaattg cctcttcctt gtatttcacc attcctgccg acaaagtga cttcatcgtg
 1381 ttcaccagca gtgataccat tacagaccgt gaagaagcgc tgaaatctcc tctggttcag
 1441 gtgatgctga cactgggtat cgtgaaagaa aaagacgtcc tgtttgggc cgaccataaa
 1501 gtgaatagca tggaggtggc catcgacgaa gcgtgtactc gtattatcgc caaacgtcag
 1561 cctaccgctt cagatctcgc tctggttate gccattatca aaacgatcac ccatctggag
 1621 cgtattggag atgttgccga aagcattgcc aaagtggccc tggagagctt ttctaacaaa
 1681 cagtataatc tgctggtcag cctggaatct ctgggtcaac acaccgttcg tatgctgcat
 1741 gaagtgtctg atgcttttgc ccgtatggat gtgaaagcag ccattgaagt ctatcaggag
 1801 gatgaccgta tcgatcagga atatgagagc attgtccgct aactgatggc ccatatgatg
 1861 gaagatccgt ctagcattcc gaatgtgatg aaagtgatgt gggcagctcg tagtattgaa
 1921 cgtgtgggtg accgctgcca gaacattgtg gagtatatca tctatttcgt aaaaggcaaa
 1981 gatgttcgcc acaccaaacc ggatgacttc ggtactatgc tggactga

//

SEQ ID NO 16*cstII: bifunctional 2,3 2,8 sialyltransferase***EC 2.4.99.4, 2.4.99.8***Campylobacter jejuni***876 bp ds-DNA**

1 atgaagaaag tcatcattgc aggcaatggg ccaagtctga aagagatcga ctattctcgc
 61 ctcccaaatg attttgatgt ctttcgctgc aaccagtctt atttgaaga taaatattat
 121 ctggggaaaa aatgtaaagc ggtattctac aaccgatcc tttctttga acagtattat
 181 acctgaaac acctattca aaaccaggaa tatgaaaccg aactattat gtgtagcaat
 241 tacaatcagg cgcacctgga aatgaaaat ttgtcaaaa cgttctatga ttacttcca
 301 gatgcccatt taggttacga ttttttaaa cagcttaaag acttcaacgc ttactttaaa
 361 tttcacgaaa tttattttaa tcagcgcatt acctcaggtg tatacatgtg cgccgttgcg
 421 attgcattgg gctacaaaga aatctatttg tcaggcatcg atttctacca aaacggtagc
 481 agttacgcat ttgatacaca acagaagaac ctcttaaat tagcgcctaa ttttaaaaac
 541 gacaactcac attacatcgg ccacagcaag aatacagata ttaaagcgtt ggaattcctg
 601 gaaaaaacat ataagatcaa actgtattgc ttatgccga attctctgtt ggcgaacttt
 661 attgagcttg ccccaaatct gaacagcaat ttatcatcc aggagaagaa caactatacg
 721 aaagacatte tcatcccgtc cagegaagcg tatggtaaat tcagtaaaaa catthaatttt
 781 aagaaaatca agattaaaga aaatatctat tataaactga ttaaagatct gctgcgctta
 841 ccgtccgaca tcaagcatta tttcaaaggc aatatga

//

SEQ ID NO 17*lic3b: bifunctional 2,3 2,8 sialyltransferase***EC 2.4.99.4, 2.4.99.8***H. influenzae***999 bp ds-DNA**

1 atgcccaatc aatcaatcaa tcaatcaatc aatcaatcaa tcaatcaatc aatcaatcaa
 61 tcaatcaatc aatcaatcaa tcaatcaatc aatcaatcaa tcaatcaatc aaagcctgtc
 121 attattgcag gtaattggaac aagtttaaaa tcaattgact atagtttatt acctaaagat
 181 tatgatgttt tccgttgcaa tcaattttat ttgaggatc attattttct tggttaagaaa

241 ataaaaaagg tattttttaa ttgttctgta atttttgaac aatactatac gtttatgcaa
301 ttaattaaata ataatagaata tgaatatgct gatattatc tatcatcttt tctgaattta
361 ggggattcag aattaaagaa aatccagcgt ttagaaaaat tactaccaca aatcgatctt
421 ggcatagct atttaaaaaa actacgagct ttgatgctc atttacaata tcacgaacta
481 tatgagaata agaggattac atcaggcgtt tatatgtgtg cagtggcaac tgctatgggt
541 tataaagatc ttatttgac aggcattgat tttatcaag aaaaaggga tccttacgca
601 tttcatcatc aaaaagaaaa tattattaaa ttattacctt cttttcaca aaataaaagt
661 caaatgata tccattctat ggaatatgat ttaaatgcac ttatttctt acaaaaacat
721 tatgggtgaa atatttactg tatttcgcca gaaagtcctc tatgtaatta ttttcctta
781 tcaccactga ataaccatt tacttttatt cccgaagaaa agaaaaatta cacacaagat
841 attttaattc cgccagagtc agtgataaaa aaaattggtat tatattccaa accaagaatt
901 taccaaaatc tgggttttcg gttgatctgg gatataattac gtttacctaa tgatataaaa
961 aaagctttga aagcaaagaa atgagacta cgcaaatga

//

SEQ ID NO 18***neuS: 2,8 polysialyltransferase*****EC 2.4.99.8*****E. coli K1*****1230 bp ds-DNA**

1 atgatatttg atgctagttt aaagaagttg aggaaattat ttgtaaatcc aattggggtt
61 ttccgtgact catggttttt taattctaaa aataagactg aaaagctatt gtcaccttta
121 aaaataaaaa acaaaaatat ttttattgtt gttcatttag ggcaattaaa gaaagcagag
181 ctttttatac aaaaatttag taagcgtagt aattttctta tcgtcttggc aactaaaaaa
241 aacactgaaa tgccaagatt agttcttgag caaatgaata aaaagttgtt ttctcatat
301 aaactactat ttataccaac agagccaaat acattttcgc ttaaaaaagt tatatgggtt
361 tataatgtat ataaatatat agttttaaat tcaaaagcta aagatgctta tttatgagc
421 tatgcacaac attatgcaat cttcatatgg ttgttcaaaa aaaacaatat aagatgttca
481 ttaattgaag aggggacagc gacgtataaa acagagaaaa aaaacacact agtaaatatt
541 aatttttatt cgtggatcat taattcaatt atctgttcc attatccaga tttaaattt
601 gaaaatgtat acggcacctt tccaaatttg ttaaaagaaa aatttgatgc aaaaaattt
661 ttgagttta aaactattcc attagttaaa tcgtcaacaa gaatggataa tctcatatc
721 aaatatcgta tctagagaga tgatattata tatgtaagtc aaagatattg gattgacaac
781 gaattgtatg cgcattcatt aatatctacc ttgatgagaa tagataaatc tgataacgca
841 agagttttta taaaacctca ccctaaagaa actaaaaaac atattaatgc aattcaaggt
901 gcaataaata aagcaaagcg tcgtgatata attattattg tagaaaaaga cttttaata
961 gagtcaataa taaaaaatg caaaataaaa cacttgattg gattaacatc atcttctttg
1021 gtatacgcat ctttagttta taaagagtgt aagacatatt caatagcacc tattattata
1081 aaattgtgta ataataaaaa atcccaaaaa gggactaata cgctgcgtct ccatttcgat
1141 attttaaaaga attttgataa tgtaaaaata ttatcggatg atatatcatc tcctctttg
1201 cagataaaaa ggattttctt gggggagtaa

//

SEQ ID NO 19***siaD: 2,8 polysialyltransferase*****EC 2.4.99.8*****N. meningitidis*****1431 bp ds-DNA**

1 atgtgggtga caacatctcc attttatctt accccccccac gtaacaattt atttgtcata
61 tctaatttag gtcagcttaa ccaagtccaa agcctaatta aaatacaaaa attaaccaat
121 aatttactag taattttata tacttctaaa aacttaaaaa tgcctaagtt agttcatcaa

181 tcagctaaca agaatctatt tgaatctatt tatctatttg agcttcctag aagcccta
241 aatataactc ctaaaaaatt actttatatt tatagaagtt acaaaaaaat ccttaatt
301 atacagcctg ctcactccta tatgctgtct ttacaggcc actactccta tctgattagt
361 attgcaaaaa aaaagaatat tacgactcat ttaattgatg aagggactgg aacatatgct
421 cctttattag aatcatttcc atatcatcca aaaaattag aacgttattt gattggaaat
481 aatcttaata ttaaaggata tatagatcat ttgacatat tgcattgccc ctttctgaa
541 tatgctaaaa aaatatttaa tgcaaaaaaa tataaccggt tttttgcgca tgctggagga
601 ataagcatta ataataacat tgcaaaccta cagaaaaaat atcaaatatc taaaaatgac
661 tatatttttg ttagtcaacg ctaccccatc tcagatgatt tgtattataa gagtatagta
721 gaaatcttaa acagcataag ttacaaaatt aaaggaaaga tatttattaa actacacca
781 aaagagatgg gcaacaacta tgtaattgtct ttatttctaa atattgtaga aataaacct
841 cggctggttag ttattaatga acctcctttt ctaattgagc ccctaataa ctaacaat
901 cctaaaggaa ttataggcct ggctcttagt tctttaattt atacaccatt actctaccc
961 tcaacccaat gtctttctat tggagagtta attattaact taattcaaaa atattcaatg
1021 gtggaaaaca ctgaaatgat ccaagaacac ttagagatta ttaagaaatt taattttatt
1081 aatatactaa atgatttaa tggggtaata agtaaccctc tctttaaac agaagaaca
1141 ttgaaacac ttctaaatc tgcagaattc gcataataa ctaaaaacta ctttcaggct
1201 atttttact ggcaacttgc cagcaaaaac aatattacct tattagggca taaagcatta
1261 tggactaca atgcacttta taatgtaaaa caaattata agatggaata ttcagatatt
1321 ttttatatcg ataatactc cgtagacttt catagtaaag ataaattgac atgggaaaaa
1381 attaacatt attactattc cgccgacaat agaattggta gagatagata a

//

SEQ ID NO 20***whnK*: α 1,2 fucosyltransferase*****E. coli* O86****EC 2.4.1.69****909 bp ds-DNA**

1 atgtatagtt gttgtctgg tgggttaggt aatcaaatgt ttcagtatgc tgcggcatat
61 atcttacaga gaaagcttaa acaaatgatca ttggttttag acgatagcta ttttttagat
121 tgctcaaatc gtgatacag tagaagattt gaattgaatc aatttaacat atgttatgat
181 cgtctgacta caagtaagga aaaaaagag atatccataa tacgacatgt aaatagatat
241 cgtttgccct tatttggtac aaattctata ttggagttc tactaaaaaa aaactatttg
301 cctgaagcaa aattttatga attttgaac aactgtaaat tacagggtta aaatggttat
361 tgtctatttt cttatttcca gtagtctaca ttgatagata gtcacgtga tatgattctc
421 ccattattcc agattaatga agatttgctc aatttatgta atgacttgc tatttcaaaa
481 aaagtatat gtgagaatgc taacacaact tcactacata tcaggcgtgg agactacatc
541 accaaccctc acgcctctaa atttcattggg gtgttgccca tggattacta tgaaaaggct
601 attcgttata ttgaggatgt tcaaggagaa cagggtgatta tcgtatttcc agatgatgtg
661 aaatgggctg agaatacatt tgctaatcaa cctaattatt acgttggtta taattctgaa
721 tgcgagtaca gtgcgattga tatgttttta atgtcaaagt gtaaaaacaa tataatagcc
781 aatagtacat atagtgtgtg gggggcatgg taaataactt tcgaagataa aatagtgtt
841 tcccctcgta agtggtttgc tggaaataat aaatctaagt tgaccatgga tagttggatt
901 aatctttga

//

SEQ ID NO 21***galE******E. coli* K12****EC 5.1.3.2****1017 bp ds-DNA**

1 atgagagttc tgggtaccgg tggtagcggg tacattggaa gtcataacctg tgtgcaatta
61 ctgcaaaacg gtcgatgtgt catcattctt gataacctct gtaacagtaa gcgcagcgta
121 ctgcctgtta tcgagcggtt aggcggcaaa catccaacgt ttgtgaagg cgatattcgt
181 aacgaagcgt tgatgaccga gatcctgcac gatcacgcta tcgacaccgt gatccacttc
241 gccgggctga aagccgtggg cgaatcggta caaaaaccgc tggaatatta cgacaacaat
301 gtcaacggca ctctgcgcct gattagcgcc atgcgcgccg ctaacgtcaa aaactttatt
361 tttagctcct ccgccaccgt ttatggcgat cagcccaaaa ttccatacgt tgaaagcttc
421 ccgaccggca caccgcaaag cccttacggc aaaagcaagc tgatgggtgga acagatecctc
481 accgatctgc aaaaagccca gccggactgg agcattgccc tgctgcgcta cttcaaccgg
541 gttggcgcgc atccgtcggg cgatatgggc gaagatccgc aaggcattcc gaataacctg
601 atgccatata tcgcccaggt tgctgtaggc cgctcgcgact cgctggcgat ttttggtaac
661 gattatccga ccgaagatgg tactggcgta cgcgattaca tccacgtaat ggatctggcg
721 gacggtcacg tcgtggcgat ggaaaaactg gcgaacaagc caggcgta caatctacaac
781 ctcggcgctg gcgtaggcaa cagcgtgctg gacgtggta atgccttcag caaagcctgc
841 ggcaaacccg ttaattatca tttgcaccg cgctcgcgagg gcgaccttc gccctactgg
901 gcggacgcca gcaaagccga ccgtgaactg aactggcgcg taacgcgcac actcgatgaa
961 atggcgcagg acacctggca ctggcagtc cgcacccac agggatatcc cgattaa

//

SEQ ID NO 22*gmd:GDP –mannose4,6 dehydratase**E. coli K12***EC 4.2.1.47****1122 bp ds-DNA**

1 atgtcaaaag tcgtctcat caccggtgta accggacaag acggttctta cctggcagag
61 tttctgctgg aaaaagggtta cgaggtgcat ggtattaagc gtcgcgcac gtcattcaac
121 accgagcgcg tggatcacat ttatcaggat ccgcacacct gcaaccgaa attccatctg
181 cattatggcg acctgagtga tacctctaac ctgacgcgca ttttgcgtga agtacagccg
241 gatgaagtgt acaacctggg cgcaatgagc cacgttgccg tctcttttga gtcaccagaa
301 tataccgctg acgtcgacgc gatgggtacg ctgcgcctgc tggaggcgat ccgttcctc
361 ggtctggaaa agaaaactcg ttctatcag gttccacct ctgaactgta tggctgggtg
421 caggaaatc cgcagaaaga gaccacgccg ttctaccgc gatctccgta tgcggtcgcc
481 aaactgtacg cctactggat caccgttaac taccgtgaat cctacggcat gtacgcctgt
541 aacggaatc tctcaacca tgaatccccg cgccgcggcg aaaccttcgt taccgcaaa
601 atcaccgcg caatcgccaa catcgcccag gggctggagt cgtgcctgta cctcggcaat
661 atggattccc tgcgtgactg gggccacgcc aaagactacg taaaaatgca gtggatgatg
721 ctgcagcagg aacagccgga agatttcgtt atcgcgaccg gcgttcagta ctccgtgcgt
781 cagttcgtgg aaatggcggc agcacagctg ggcatcaaac tgcgcttga aggcacgggc
841 gttgaagaga agggcattgt ggtttccgtc accgggcatg acgcgccggg cgtaaaccg
901 ggtgatgtga ttatcgctgt tgaccgcgt tacttccgtc cggctgaagt tgaaacgctg
961 ctcggcgacc cgaccaaagc gcacgaaaaa ctgggctgga aaccggaaat caccctcaga
1021 gagatgggtg ctgaaatggt ggctaatac ctgaaagcgg cgaaaaaaca ctctctgctg
1081 aaatctcac gctacgacgt ggcgatcgcg ctggagtcat aa

//

SEQ ID NO 23*Fcl: GDP-fucose synthetase**E. coli K12***EC 1.1.1.271****966 bp ds-DNA**

1 atgagtaaac aacgagtttt tattgctggg catcgcgga tggcgggtc cgccatcagg
61 cggcagctcg aacagcgagg tgatgtggaa ctggtattac gcacccgca cgagctgaac
121 ctgctggaca gccgcgccgt gcatgatttc ttgccagcg aacgtattga ccaggtctat
181 ctggcgccgg cgaaagtggg cggcattgtt gccacaaca cctatccggc ggatttcate
241 taccagaaca tgatgattga gagcaacatc attcacgccg cgcatcagaa cgacgtgaac
301 aaactgctgt ttctcgatc gtctgcac tacccgaaac tggcaaaaca gccgatggca
361 gaaagcgagt tggtcaggg cagctggag ccgactaac agccttatgc tattgcaaa
421 atcgccggga tcaaactgtg cgaatcatic aaccgccagt acggacgca ttaccgtca
481 gtcatgccga ccaacctgta cgggccacac gacaacttc acccgagtaa ttcgatgtg
541 atcccagcat tgctgcgctg ctccacgag gcgacggcac agaagcgcc ggacgtggg
601 gtatggggca gcggtacac gatgcgcgaa ttctgcacg tcgatgatat ggccggcgccg
661 agcattcatg tcatggagct ggcgcatgaa gtctggctgg agaaccacca gccgatgtg
721 tcgcacatta acgtcggcac gggcgtgac tgactatcc gcgagctggc gcaaacatc
781 gccaaagtgg tgggttaca aggcgggtg gttttgatg ccagcaaac ggatggcacg
841 ccgcgcaaac tgctggatgt gacgcgctg catcagctg gctggtatca cgaaatctca
901 ctggaagcgg ggctgccag cactaccag tggctcctg agaatcaaga ccgcttcgg
961 gggtaa

//

SEQ ID NO 24*gmm:GDP-mannose mannosyl hydrolase**E. coli* K12**EC 3.2.1.42****480 bp ds-DNA**

1 atgtttttac gtcaggaaga ctttgccacg gtagtgcgct cactccgct tgtctctc
61 gactttattg tcgagaacag tcgcggcgag ttctgcttg gcaaaagaac caaccgccc
121 gcgcagggtt actggtttgt gccgggaggg cgctgcaga aagacgaac gctggaagcc
181 gcatttgacg ggctgacgat ggcggaactg gggctgcgtt tgccgataac agcaggccag
241 tttacgggtg tctggcagca ctttatgac gataactct ctggcacgga ttcaccact
301 cactatgtgg tgctcggttt tcgcttcaga gtatcggaag aagagctgtt actgccgat
361 gagcagcatg acgattaccg ctggctgacg tcggacgcgc tgctcgccag tgataatgt
421 catgctaaca gccgcgccta tttctcgtg gagaagcgta ccggagtacc cggattatga

//

SEQ ID NO 25*cpsB: mannose-1-phosphate guanyltransferase**E. coli* K12**EC 2.7.7.13****1437 bp ds-DNA**

1 atggcgagc cgaaactcta tccagttgtg atggcaggtg gctccggtag ccgcttatgg
61 ccgctttccc gcgtacttta tccaagcag ttttatgcc tgaaggcga tctcaccatg
121 ctgcaaacca ccatctgccg cctgaacggc gtggagtgcg aaagcccggg ggtgatttgc
181 aatgagcagc accgctttat tgctgcggaa cagctgcgtc aactgaaca actaccgag
241 aacattatc tcgaaccggc agggcgaaac acggcacctg ccattgcgct ggccggcgctg
301 gcggcaaaac gtcatagccc ggagagcgac ccgttaatgc tggatttggc ggccgatcat
361 gtgattgccg atgaagacgc gtccgtgcc gccgtgcgta atgcatgcc atatgccgaa
421 gcgggcaagc tggtagctt cggcattgtg ccgatctac cagaaaccgg ttatggctat
481 attcgtcgcg gtgaagtgtc tgcgggtgag caggatatgg tggccttga agtggcgag
541 tttgtcga aaaccgaatct ggaaaccgct caggcctatg tggcaagcgg cgaatattac
601 tggaaacagc gtatgttctt gttccgcgcc ggacgctatc tcgaagaact gaaaaaatat

661 cgcccgata tctcgtatgc ctgtgaaaaa gcatgagcgc ccgtcgtacc ggatctcaat
721 ttattcgcg tggatgaaga agcgtttctc gcctgcccgc aagatcggt ggattacgcg
781 gtcattgaac gtacggcaga tgctgttggt gtgccgatgg atcggggctg gagcgatgtt
841 ggctcctggt ctctattatg ggagatcagc gccacaccg ccgagggcaa cgtttgccac
901 ggcatgtga ttaatcaca aactgaaaac agctatgtgt atgctgaatc tggcctggc
961 accaccgtcg ggggtgaaaga tctggtagt gtgcagacca aagatcggt gctgattgcc
1021 gaccgtaacg cgttacagga tgtgaaaaa gtggtcgagc agatcaaagc cgtggtcgc
1081 catgagcatc ggggtgcatc cgaagtgtat cgtccgtggg gcaaatatga ctctatcgac
1141 gcgggcgacc gctaccaggt gaaacgcac accgtgaaac cgggagagg cttgtcgga
1201 cagatgcacc ataccgcgc ggaacactgg gtggtgtcg cggaacggc aaaagtcacc
1261 attgatggtg atataaaact gcttggtgaa aacgagcca ttatattcc gctggggcg
1321 acgattgcc tggaaaacc ggggaaaatt ccgtcgtatt taattgaagt gcgctccgc
1381 tctatctcg aagaggatga tgtggtcgt ttcgcgcatc gctacggac ggtgtaa

//

SEQ ID NO 26*cpsG: phosphomannomutase**E. coli K12***EC 5.4.2.8****1371 bp ds-DNA**

1 atgaaaaaat taactgctt taaagcctat gatattcgcg ggaaattagg cgaagaactg
61 aatgaagata tcgctggcg cattggcgc gcctatggcg aatttctca accgaaaacc
121 attgtgttag gcggtgatgt ccgcctcacc agcgaaacct taaaactggc gctggcgaaa
181 ggtttacagg atcggggcgt tgacgtgctg gatattgta tgcggcac cgaagatc
241 tatttcgcca cgttccatct cggcgtggat ggcggcattg aagtaccgc cagccataat
301 ccgatggatt ataacggcat gaagctggtt cgcgaggggg ctcgccgat cagcggagat
361 accggactgc gcgacgtcca gcgtctggct gaagccaacg acttctctc cgtcgtgaa
421 accaaacgcg gtcgctatca gcaaatcaac ctgcgtgacg ctacgttga tcactgttc
481 ggttatatca atgtcaaaaa cctcagccg ctcaagctgg tgatcaactc cggaacggc
541 gcagcgggtc cgggtgtgga cgccattgaa gcccgctta aagccctcg cgcgcccg
601 gaattaaatc aagtgcacaa cagcccgac ggcaattcc ccaacggtat tctaacca
661 ctactgccg aatgccgca cgacaccgc aatgcggtca tcaaacacg cgcggatag
721 ggcattgctt ttgatggcga tttgaccgc tgttctctg ttgacgaaa agggcagtt
781 attgagggt actacattg cggcctgtg gcagaagcat tctcgaaa aaatccgcg
841 gcgaagatc tccacgatc acgtctctc tggaaacgc ttgatgtgt gactccgca
901 ggtggcacgc cgtaatgtc gaaaaccgga cagccctta taaagaacg tatgcgaag
961 gaagacgcca tctatgtgt cgaaatgagc gccaccatt actccgtga ttctgttac
1021 tgcgacagc gcatgatcc gtggtgctg gtcgccgac tgggtgcct gaaagataa
1081 acgtgggcg aactggtacg cgaccgatg gcggcgttc cggaacgg tgagatcaac
1141 agcaaaactg cgcaaccgt tgaggcgatt aaccgctgg aacagcatt tagccgtgag
1201 gcgctggcg tgatgcac cgatggcatc agcatgacct ttccgactg gcgtttaac
1261 ctgcgcacct ccaataccg accggtgtg ccctgaatg tggaatcg cggtgatgt
1321 ccgctgatg aagcggaac ggaactctg ctgacgttc tgaacgagta a

//

SEQ ID NO 27*TNFα light chain***708 bp ds-DNA**

1 atgaacaaa gcactattgc actggcactc ttaccgttac tgtttaccc ttgacaaaa
61 gccgatattc agatgacca gtccccgagc agcctgagc caagcgtcg cgaccgtgt
121 actattacct gtaaagccag ccaaatgtg ggcacgaatg tcgctggta tcaacagaag
181 ccgggcaaag cgccgaaggc gctgatctat tcggcgagct tttgtacag cgcggttcg

241 taccgtttta gcggcagcgg ttccggccacc gactttacgc tgaccattag ctggtgcag
301 ccggaagatt ttgcgacctt ctactgccag cagtataaca tctaccgct gacgttcggt
361 caaggtacga aagtcgaaat caagcgcacc gttgccgcac cgagcgtgtt catctttccg
421 ccaagcgatg aacaactgaa gtccggcact gcaagcgtgg ttgtctgct gaataacttt
481 tatccacgcg aagccaaagt gcagtgaag gtggacaatg cactgcaaag cggcaactct
541 caggaaagcg ttactgagca ggacagcaaa gactccacct acagcttgc ctctacgttg
601 accctgtcca aagcagacta cgagaagcac aaagtttatg cctgtgaggt caccaccaa
661 ggtctgagca gcccggtcac gaagtccttc aatcgcgcg agtgctaa

//

SEQ ID NO 28*TNFα heavy chain-4X dqnat- hexahistidine***876 bp ds-DNA**

1 aacattaaga aggaggtaaa acatatgaaa aaactgctgt tcgcgattcc gctgggtgtg
61 ccgttctata gccatagcga agtcagctg gttgagagcg gcggtggtct ggtgcagccg
121 ggtggtagcc tgcgtctgag ctgcgctgcg agcggttacg tttcaccga ctatggcatg
181 aactgggttc gccagcacc gggttaagggt ctggagtga tgggttgat caatacctac
241 attggtgaac cgatctacgc tgattccgtg aagggtcgtt tcacctctc tctggatacg
301 tctaagagca cggcctatct gcaaatgaac agcctgcgtg cagaggatac ggcggtctac
361 tattgtgcgc gtggttaccg ttctacgcg atggattact ggggtcaagg caccctggtg
421 accgttagca gcgcgagcac caagggtccg tctgtgtcc cgttgccgc tagctccaag
481 agcacctcgg gtgttacggc tgcgtgggc tcctgtgca aagactattt cccggagccg
541 gtcacggtta gctggaacag cgggtccctg accagcgggt ttcacacctt tccggcggtc
601 ttgcagtcta gcggtctgta tagcctgagc agcgtcgta cggttccgag cagcagcctg
661 ggcacgcaga cctacatctg caacgtgaac cacaaccaa gcaatacaa agtagacaag
721 aaagtcgagc cgaagtcctg cgataagacc catacctgtg ctgcgctcga ggatcagaac
781 gcgaccggcg gtgacaaaaa tgccacaggt ggcgatcaaa acgccaccgg cggtgaccag
841 aatgcgacag tcgacatca ccatcatcac catta

//

SEQ ID NO 29*malE-tev-gh2-hexahistidine***1797 bp ds-DNA**

1 atgaaaaaga ttggctggc gctggctggt ttagttag cgtttagcgc atcggcgtct
61 agaaaaatcg aagaaggtaa actggtatc tggattaacg gcgataaagg ctataacggt
121 ctgcgtgaag tcggaagaa attcgagaaa gataccggaa ttaaagtcac cgttgagcat
181 ccggataaac tggaagagaa attccacag gttgcggcaa ctggcgatgg ccctgacatt
241 atcttctggg cacacgaccg ctttgggtgc tacgtcaat ctggcctgtt ggctgaaatc
301 accccggaca aagcgttcca ggacaagctg tatccgttta cctgggatgc cgtacgttac
361 aacggcaagc tgattgctta cccgatcgt gttgaagcgt tatcgtgat ttataacaaa
421 gatctgtgc cgaacccgcc aaaaacctgg gaagagatcc cggcgtgga taaagaactg
481 aaagcgaaag gtaagagcgc gctgatgttc aacctgcaag aaccgtactt cacctggccg
541 ctgattgctg ctgacggggg ttatgcgttc aagtatgaaa acggcaagta cgacattaaa
601 gacgtgggcg tggataacgc tggcgcaaaa gcgggtctga cttcctggt tgacctgatt
661 aaaaacaaac acatgaatgc agacaccgat tactccatcg cagaagctgc cttaataaaa
721 ggcgaaacag cgatgacat caacggcccc tgggcatggt ccaacatcga caccagcaaa
781 gtgaattatg gtgaacggt actgccgacc tcaagggtc aaccatcaa accgttcgtt
841 ggcgtgctga gcgcaggtat taacggccc agtcgaaca aagagctggc gaaagagttc
901 ctgaaaact atctgctgac tgatgaaggt ctggaagcgg ttaataaaga caaacgctg
961 ggtgccgtag cgctgaagtc ttacgaggaa gagttggcga aagatccacg tattgccgcc

1021 accatggaaa acgcccagaa aggtgaaatc atgccgaaca tcccgcagat gtccgctttc
1081 tggatatgccg tgcgtactgc ggtgatcaac gccgccagcg gtcgtcagac tgcgatgaa
1141 gccctgaaag acgcgcagac tcgtatcacc aagggcgaaa acctgtattt tcagggttc
1201 ccaaccattc cttatccag gcttttgac aacgctatgc tccgcgcccgc tgcctgtac
1261 cagctggcat atgacaccta tcaggagttt gaagaagcct atatcctgaa ggagcagaag
1321 tattcattcc tgcagaaccc ccagacctcc ctctgcttct cagagtctat tccaacacct
1381 tccaacaggg tgaaaacgca gcagaaatct aacctagagc tgctccgcat ctccctgctg
1441 ctcatccagt catggctgga gcccgctgcag ctctcagga gcgtcttcgc caacagcctg
1501 gtgtatggcg cctcggacag caacgtctat cgccacctga aggacctaga ggaaggcatc
1561 caaacgctga tgtggaggct ggaagatggc agccccgga ctgggcagga tcagaacgcc
1621 acgtacagca agtttgacac aaaatcgcac aacgatgacg cactgctcaa gaactacggg
1681 ctgctctact gcttcaggaa ggacatggac aaggtcgaga cattctgcg catcgtgcag
1741 tgccgctctg tggagggcag ctgtggcttc gtcgaccatc accatcatca ccattaa

//

SEQ ID NO 30**LsgE: N-acetylglucosaminyltransferase***Haemophilus influenzae***EC 2.4.1-****294 a.a.**

1 MLSIIVPSYNRKAIEVPALLE
21 SLTQQTSSNFEVIIVDDYSK
41 ERVVVEQRYSPVTVIRNET
61 NQGAAESRNIGARASKGDWL
81 LFLDDDDRFMPEKCEKILQV
101 IEQNPDIINFIYHPAKCEMVN
121 EGFTYVTQPIEPQEISTERI
141 LLANKIGGMPMVAIKKEMFL
161 KIGGLSTALRSLEDYDFLLK
181 LLQESSFTPYKINEPLTYCT
201 FHTKRSSVSTDNTQKAID
221 YIREHYVKTVEQARNFDINA
241 SYILAYPHIMNLSRKAAYY
261 FDIFKKTKSIKQFIITLVIL
281 ISPKLAINLKRLGK*

//

SEQ ID NO 31**LsgD: galactosyltransferase***Haemophilus influenzae***EC 2.4.1-**

1 MLKKYLISLDKDIQRRKLFF
21 SQKNTEDFQIFSAINTMQKD
41 WDELASIFNIEQFKAHYFRN
61 VTKGEIGCTLSHLSVYQKIV
81 EDNDIAEDSYALVCEDDALF
101 HLDFFQQNL TALLSEKLEAEI
121 ILLGQSNINNFNDDTDLEINY
141 PTTFSFLCKKTGNVNYAFPY
161 KSYFAGTVGYLIKSAARRF

181 IQQISQNKPFWLADDFLLFE
201 QNFNIRNKVVRPLMVIENPV
221 LISNLESVRGSLSNLLKKL
241 MKYPLKKIFAIKKNLAN*

SEQ ID NO 32**FucT: fucosyltransferase***Helicobacter pylori***2.4.1.152**

1 MFQPLLDAYV ESASIEKMAS KSPPLKIAV
31 ANWWGDEEIK EFKNSVLYFI LSQRYTITLH
61 QNPNEFSDLV FGNPLGSARK ILSYQNAKRV
91 FYTGENESPN FNLFDYAIGF DELDFNDRYL
121 RMPLYDYDR LH HKAESVNDTT APYKLKDNSL
151 YALKKPSHCF KEKHPNLC AV VNDESDPLKR
181 GFASFVASNP NAPIRNAFYD ALNSIEPVTG
211 GGSVRNTLGY NVKNKNEFLS QYKFNLCFEN
241 TQGYGYVTEK IIDAYFSHTI PIYWGSPSVA
271 KDFNPKS FVN VHDFKNFDEA IDYIKYLH TH
301 KNAYLDM LYE NPLNTLDGKA YFYQNLSFKK
331 ILAFFKTILE NDTIYHDNPF IFCRDLNEPL
361 VTIDDLRVNY DDLRVNYDDL RINYDDL RVN
391 YDDL RINYDD LRVNYDDL RV NYDDL RINYD
421 DLRVNYDDL R VNYERLLSKA TPLLELSQNT
451 TSKIYRKAYQ KSLPLLRAIR RWVKKLGL*

SEQ ID NO 33**WbnI***E. coli* 086**EC 2.4.1.37**

1 MVINIFYICTGEYKRFFDKF
21 YLSCEDKFIPEFGKKYYVFT
41 DSDRIYFSKYLNVEVINVEK
61 NCWPLNTLLRFSYFLKVIDK
81 LQTN SYTFFFNANAVIVKEI
101 PFSTFMESDLIGVIHPGYKN
121 RISILYPWERRKNATCYLGY
141 LKKGIYYQGC FNGGKTASFK
161 RLIQICNMMTMADLKKNLIA
181 KVHDESYLNYYYYY NKPLLL
201 SELYSWPEKYGENKDAKIIM
221 RDKERESWYGN IKK*

SEQ ID NO 34**BgtA: α -N-acetylgalactosaminyl transferase***Helicobacter mustelae***EC 2.4.1.40**

1 MQSTAQNTQQNTHFAGSSQT
21 TPQAAQSVQQASLALPKSSP
41 TCYKIAILYICTGAYSIFWQ
61 DFYDSAKVHLLPAHRLTYFV
81 FTDADSLYAEAEASDVRKIYQ
101 ENLGWPFNTLKRFEMLGQE
121 EALREFDFVFFFNANCLFFQ
141 HIGDEFLPIEEDILVTQHYG
161 FRDASPECFTYERNPKSLAY
181 VPFKGKAYVYGSTNGGKAG
201 AFLALARTLQERIQEDLSRG
221 IIAIWHDESHLNAYIIDHPN
241 YKMLDYG YGFPEGYGRVPGG
261 GVIYIFLRDKSRVIDVNAIKG
281 MGSPANRRLKNALRKLKHFS
301 KRLLGR*

SEQ ID NO 35**GNE: aka z3206 UDP-N-acetylglucosamine 4-epimerase***E. coli***EC 5.1.3.c**

1 MNDNVLLIGASGFVGTRLLE
21 TAIADFNKNLKDQQSHFYP
41 EITQIGDVRDQQALDQALAG
61 FDTVVLLAAEHRDDVSPTSL
81 YYDVNVQGTRNVLAAMEKNG
101 VKNIIFTSSVAVYGLNKHNP
121 DENHPHDPFNHYGKSKWQAE
141 EVLREWYNKAPTERSLTIIR
161 PTVIFGERNRGNVYNLLKQI
181 AGGKFMMVGAGTNYKSMAYV
201 GNIVEFIKYKLKNVAAGYEV
221 YNYVDKPDLMNQLVAEVEQ
241 SLNKKIPSMHLPYPLGMLGG
261 YCFDILSKITGKKYAVSSVR
281 VKKFCATTQFDATKVHSSGF
301 VAPYTLSQGLDRTLQYEFVH
321 AKKDDITFVSE*

SEQ ID NO 36**PglC: Bacillosamine transferase***Campylobacter jejuni***EC 2.7.8.6**

1 MYEKVFKRIFDFILALVLLV
21 LFSPVILITALLKITQGSV
41 IFTQNRPGLDKIFKIYKFK
61 TMSDERDEKGELLSDELRLK
81 AFGKIVRSLSLDELLQLFNV
101 LKGDMSFVGPRPLLVEYLSL

121 YNEEQKLRHKVRPGITGWAQ
141 VNGRNAISWQKKFELDVYYV
161 KNISFLDLKIMFLTALKVL
181 KRSGVSKEGHVTTEKFNGKN
201 *

SEQ ID NO 37**PglD: UDP-N-acetylbacillosamine N-acetyltransferase***Campylobacter jejuni***EC 2.3.1.203**

1 MARTEKIYIYGASGHGLVCE
21 DVAKNMGYKECIFLDDFKGM
41 KFENTLPKYDFFIAIGNNEI
61 RKKIYQKISENGFKIVNLIH
81 KSALISPSASVEENAGILIM
101 PYVVINAKAKIEKGVILNTS
121 SVIEHECVIGEFSHVSVGAK
141 CAGNVKIGKNCFLGINS CVL
161 PNLSLADDSILGGGATLVKS
181 QNEKGVFVGVPKRKI*

SEQ ID NO 38**PglE: aminotransferase***Campylobacter jejuni***EC 2.6.1.34**

1 MRFFLSPPHMGGNELKYIEE
21 VFKSNIYIAPLGEFVNRFEQS
41 VKDYSKSENALALNSATAAL
61 HLALRVAGVKQDDIVLASSF
81 TFIASVAPICYLKAKPVFID
101 CDETYNIDVDLLKLAIKECE
121 KKPKALILTHLYGNAAKMDE
141 IVEICKENEIVLIEDAAEAL
161 GSFYKNKALGTFGEFGAYS Y
181 NGNKIITTS GGGMLIGKNKE
201 KIEKARFYSTQARENCLHYE
221 HLDYGYNYRLSNVLGAIGVA
241 QMEVLEQRVLKKREIYEWYK
261 EFLGEYFSFLDELENSRSNR
281 WLSTALIDFDKNELNACQKD
301 INISQKNITLHPKISKLIED
321 LKNEQIETRPLWKAMHTQEV
341 FKGTKAYLNGNSELFFQKGI
361 CLPSGTAMSKDDVYEISKLI
381 LKSIKA*

SEQ ID NO 39**PglF: dehydratase***Campylobacter jejuni*

EC 4.2.1.135

1 MIFYKSKRLAFLTSDIVLI
21 LLSVYLAFSLRFSGDIPSIF
41 YHGMMVSAIILLVLKLSFLF
61 VFRIYKVAWRFFSLNEARKI
81 FIALLLAEFCFFLIFYFFSD
101 FFNPFPRSAIVIDFVLSYMF
121 IGTLRISKRMLVDFKPSKMK
141 EEETPCIVVGATSKALHLLK
161 GAKEGSLGLFPVGVVDARKE
181 LIGTYCDKFVVEEKEKIKSY
201 VEQGVKTAIIALRLEQEELK
221 KLFEELVAYGICDVKIFSFT
241 RNEARDISIEDLLARKPKDL
261 DDSAVAAFLKDKVVLVSGAG
281 GTIGSELCKQCIFGAKHLI
301 MVDHSEYNLYKINDDLNLYK
321 EKITPILLSILDKQSLDEVL
341 KTYKPELILHAAAYKHVPLC
361 EQNPHSAVINNILGTKILCD
381 SAKENKVAKFVMISTDKAVR
401 PTNIMGCTKRVCELYTLSMS
421 DENFEVACVRFGNVLGSSGS
441 VIPKFKAQIANNEPLTLTHP
461 DIVRYFMLVAEAVQLVLQAG
481 AIAKGGELFVLDMGKPVKII
501 DLAKKMLLLSNRNDLEIKIT
521 GLRKGEKLYEELLIDENDAK
541 TQYESIFVAKNEKVDLDWLN
561 KEIENLQICEDISEALLKIV
581 PEFKHNKEGI*

SEQ ID NO 40**WbnH: α 1,3 GalNAc transferase*****E. coli***

EC 2.4.1.306

1017 bp ds-DNA

1 atgaaaaatg ttggtttat tgttacaaaa tcagaaattg gtggtgcaca aacatgggta
61 aatgaaatat ctaacctat taaagaggaa tgtaatatat ttcttattac atctgaagaa
121 ggatggctca cacataaaga tgtcttggcc ggagttttg tcataccagg tattaataaaa
181 tattttgact tccttacatt gtttaaattg agaaaaattt taaaagaaaa taacatttca
241 acgttaatat caagttctgc taatgccgga gtttatgcca ggtagttcg attactagtc
301 gactttaaat gtatttatgt ttgcatgga tggcttgtt tatataatgg tggcgccta
361 aaatcaattt ttgcattgt tgaaaaatac ctttcttat taactgatgt tatatgggtg
421 gtttccaaaa gtgatgaaaa aaaggcaatt gagaatattg gtataaaga accaaagata
481 atcacagat cgaattcagt gcctcagatg ccgagatgta ataataaaca actccagat
541 aaggttctgt ttgttgtag gttaacacac cctaagcgc ccgaattgtt agcgaatgta
601 atatcgaata agccccagta tagcctccat atcgtaggag ggggggaaag gttagaatca
661 ttgaagaac aattcagtga atgtgaaaat attcattttt tgggtgaggt caataattt
721 tataactatc atgagtatga ttattttca ctgatatccg atagtgaagg ttgcctatg

781 tcaggccttg aggctcacac agctgcaata ccactcctgt taagtgatgt gggcggatgt
841 ttgaattaa ttgagggtaa tgggttactt gtggaaaata ctgaagacga cattggatat
901 aaattggata aaatattcga tgactatgaa aattatcggg aacaggcaat tcgtgcctcc
961 gggaaatttg ttatcgagaa ctatgcttca gcatataaaa gcattatttt aggttga

//

References

1. Harris JM, Chess RB: **Effect of pegylation on pharmaceuticals.** *Nat Rev Drug Discov* 2003, **2**(3):214-221.
2. Veronese FM, Pasut G: **PEGylation, successful approach to drug delivery.** *Drug Discov Today* 2005, **10**(21):1451-1458.
3. Graham ML: **Pegaspargase: a review of clinical studies.** *Adv Drug Deliv Rev* 2003, **55**(10):1293-1302.
4. Levy Y, Hershfield MS, Fernandez-Mejia C, Polmar SH, Scudiero D, Berger M, Sorensen RU: **Adenosine deaminase deficiency with late onset of recurrent infections: response to treatment with polyethylene glycol-modified adenosine deaminase.** *J Pediatr* 1988, **113**(2):312-317.
5. DeFrees S, Wang ZG, Xing R, Scott AE, Wang J, Zopf D, Gouty DL, Sjoberg ER, Panneerselvam K, Brinkman-Van der Linden EC *et al*: **GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli.** *Glycobiology* 2006, **16**(9):833-843.
6. Caliceti P, Veronese FM: **Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates.** *Adv Drug Deliv Rev* 2003, **55**(10):1261-1277.
7. Whelan J: **Beyond PEGylation.** *Drug Discov Today* 2005, **10**(5):301.
8. Vutskits L, Djebbara-Hannas Z, Zhang H, Paccaud JP, Durbec P, Rougon G, Muller D, Kiss JZ: **PSA-NCAM modulates BDNF-dependent survival and differentiation of cortical neurons.** *Eur J Neurosci* 2001, **13**(7):1391-1402.
9. Muhlenhoff M, Eckhardt M, Gerardy-Schahn R: **Polysialic acid: three-dimensional structure, biosynthesis and function.** *Curr Opin Struct Biol* 1998, **8**(5):558-564.
10. Stockert RJ: **The asialoglycoprotein receptor: relationships between structure, function, and expression.** *Physiol Rev* 1995, **75**(3):591-609.
11. Steenbergen SM, Vimr ER: **Functional relationships of the sialyltransferases involved in expression of the polysialic acid capsules of Escherichia coli K1 and K92 and Neisseria meningitidis groups B or C.** *J Biol Chem* 2003, **278**(17):15349-15359.
12. Fernandes AI, Gregoriadis G: **The effect of polysialylation on the immunogenicity and antigenicity of asparaginase: implication in its pharmacokinetics.** *Int J Pharm* 2001, **217**(1-2):215-224.
13. Jain S, Hreczuk-Hirst DH, McCormack B, Mital M, Epenetos A, Laing P, Gregoriadis G: **Polysialylated insulin: synthesis, characterization and biological activity in vivo.** *Biochim Biophys Acta* 2003, **1622**(1):42-49.
14. Constantinou A, Epenetos AA, Hreczuk-Hirst D, Jain S, Wright M, Chester KA, Deonarain MP: **Site-Specific Polysialylation of an Antitumor Single-Chain Fv Fragment.** *Bioconjug Chem* 2009.
15. Constantinou A, Epenetos AA, Hreczuk-Hirst D, Jain S, Deonarain MP: **Modulation of antibody pharmacokinetics by chemical polysialylation.** *Bioconjug Chem* 2008, **19**(3):643-650.
16. Gao N, Lehrman M: **Non-radioactive analysis of lipid-linked oligosaccharide compositions by fluorophore-assisted carbohydrate electrophoresis.** *Methods Enzymol* 2006, **415**:3-20.
17. Pabst M, Altmann F: **Glycan analysis by modern instrumental methods.** *PROTEOMICS* 2011, **11**(4):631-643.
18. Mechref Y, Hu Y, Desantos-Garcia JL, Hussein A, Tang H: **Quantitative Glycomics Strategies.** *Molecular & Cellular Proteomics* 2013, **12**(4):874-884.
19. Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, Heiss C, Azadi P, Aebi M, DeLisa MP: **An engineered eukaryotic protein glycosylation pathway in Escherichia coli.** *Nat Chem Biol* 2012, **8**(5):434-436.

20. Fisher AC, Haitjema CH, Guarino C, Çelik E, Endicott CE, Reading CA, Merritt JH, Ptak AC, Zhang S, DeLisa MP: **Production of Secretory and Extracellular N-Linked Glycoproteins in Escherichia coli**. *Applied and Environmental Microbiology* 2011, **77**(3):871-881.
21. Springer GF: **Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy**. *J Mol Med (Berl)* 1997, **75**(8):594-602.
22. Heimbarg-Molinaro J, Lum M, Vijay G, Jain M, Almogren A, Rittenhouse-Olson K: **Cancer vaccines and carbohydrate epitopes**. *Vaccine* 2011, **29**(48):8802-8826.
23. Hug I, Zheng B, Reiz B, Whittall RM, Fentabil MA, Klassen JS, Feldman MF: **Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein**. *The Journal of biological chemistry* 2011, **286**(43):37887-37894.
24. Yi W, Shao J, Zhu L, Li M, Singh M, Lu Y, Lin S, Li H, Ryu K, Shen J *et al*: **Escherichia coli O86 O-Antigen Biosynthetic Gene Cluster and Stepwise Enzymatic Synthesis of Human Blood Group B Antigen Tetrasaccharide**. *Journal of the American Chemical Society* 2005, **127**(7):2040-2041.
25. Yi W, Shen J, Zhou G, Li J, Wang PG: **Bacterial Homologue of Human Blood Group A Transferase**. *Journal of the American Chemical Society* 2008, **130**(44):14420-14421.
26. Rush JS, Alaimo C, Robbiani R, Wacker M, Waechter CJ: **A Novel Epimerase That Converts GlcNAc-P-P-undecaprenol to GalNAc-P-P-undecaprenol in Escherichia coli O157**. *Journal of Biological Chemistry* 2010, **285**(3):1671-1680.
27. Glover KJ, Weerapana E, Numao S, Imperiali B: **Chemoenzymatic Synthesis of Glycopeptides with PglB, a Bacterial Oligosaccharyl Transferase from Campylobacter jejuni**. *Chemistry & biology* 2005, **12**(12):1311-1316.

What is claimed is:

1. A recombinant host cell comprising GalNAc transferase activity and galactosyltransferase activity; wherein the host cell produces an oligosaccharide composition comprising one or more GalNAc, galactose or galactose-GalNAc residues linked to a lipid carrier.
2. The host cell of claim 1, wherein the host cell further comprises one or more enzyme activities selected from fucosyltransferase, sialyltransferase and N-acetylglucosaminyl transferase; wherein the host cell produces an oligosaccharide composition comprising at least one fucose, sialic acid or GlcNAc residues linked to a lipid carrier.
3. The host cell of claim 1, comprising one or more activities selected from UndP N-acetylglucosaminyl transferase, UndPP GalNAc epimerase and UndP bacillosamine transferase.
4. The host cell of claim 1, wherein the GalNAc transferase activity comprises α 1,3-N-acetylgalactosamine transferase activity.
5. The host cell of claim 1, wherein the galactosyltransferase activity comprises one or more activities selected from β 1,3 galactosyltransferase, β 1,4 galactosyltransferase and α 1,3 galactosyl transferase activity.
6. The host cell of claim 2, wherein the fucosyltransferase activity comprises one or more activities selected from α 1,2 fucosyltransferase, α 1,3 fucosyltransferase and α 1,3/1,4 fucosyltransferase.
7. The host cell of claim 2, wherein the N-acetylglucosaminyl transferase activity comprises β 1,3N-acetylglucosaminyl transferase activity.
8. The host cell of claim 2, wherein the sialyltransferase activity comprises one or more activities selected from α 2,3 NeuNAc transferase, α 2,6 NeuNAc transferase, bifunctional α 2,3 α 2,8 NeuNAc transferase and α 2,8 polysialyltransferase activity.
9. The host cell of claim 3, wherein the UndP N-acetylglucosaminyl transferase activity comprises undecaprenyl-phosphate α -N-acetylglucosaminyltransferase activity.
10. The host cell of claim 3, wherein the UndPP GalNAc epimerase activity comprises N-acetyl- α -D-glucosaminyl-diphospho-ditrans, octacis-undecaprenol 4- epimerase activity.

11. The host cell of claim 3, wherein the UndP bacillosamine transferase activity comprises undecaprenyl phosphate N,N'-diacetylbacillosamine 1-phosphate transferase activity.
12. The host cell of claim 1, wherein the host cell further comprises an attenuation in at least one of the enzyme activities selected from N-acetylneuraminate lyase, undecaprenyl-phosphate glucose phosphotransferase and O-antigen ligase activity.
13. The host cell of claim 1, wherein the host cell further comprises one or more activities selected from N-acetylneuraminate synthase, N-acetylneuraminate cytidyltransferase, UDP-N-acetylglucosamine 2-epimerase and N-acetylneuraminate acetyltransferase.
14. The host cell of claim 1, wherein the host cell further comprises GalNAc epimerase activity.
15. The host cell of claim 1, wherein the host cell further comprises Gal epimerase activity.
16. The host cell of claim 1, wherein the host cell further comprises one or more enzyme activities selected from GDP-mannose 4,6 dehydratase, GDP-fucose synthetase, GDP-mannose mannosyl hydrolase, mannose-1-phosphate guanyltransferase and phosphomannomutase.
17. The host cell of claim 1, wherein the host cell further comprises one or more enzyme activities selected from UDP-N-acetylbacillosamine N-acetyltransferase, UDP-N-acetylglucosamine 4,6 dehydratase and UDP-N-acetylbacillosamine transaminase.
18. The host cell of claim 1, wherein the host cell produces one or more oligosaccharide composition characterized as human or human-like antigens selected from A antigen, H antigen, B antigen, T antigen, sialyl T antigen, Lewis X antigen and polysialylated antigen.
19. The host cell of claim 1 or 2, wherein the host cell produces one or more oligosaccharide compositions selected from:
 - a. $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal} \beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc} \beta 1-$;
 - b. $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal} \beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GlcNAc} \beta 1-$;
 - c. $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal} \beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{GalNAc} \alpha 1-$;
 - d. $(\text{Sia } \alpha 2,8)_n - \text{Sia } \alpha 2,8 - \text{Sia } \alpha 2,3 - \text{Gal} \beta 1,3 - \text{GalNAc } \alpha 1,3 - \text{Bac} \alpha 1-$;
 - e. $\text{Sia} \alpha 2,8 - \text{Sia} \alpha 2,3 - \text{Gal} \beta 1,3 - \text{GalNAc} \alpha 1,3 - \text{GlcNAc} \beta 1-$;

- f. Sia α 2,8 – Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- g. Sia α 2,8 – Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- h. Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- i. Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – GalNAc1-;
- j. Sia α 2,3 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- k. Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- l. Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- m. Sia α 2,6 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- n. Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- o. Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- p. Fuc α 1,2 – Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- q. Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- r. Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- s. Gal α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- t. GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- u. GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – GalNAc α 1-;
- v. GalNAc α 1,3[Fuc α 1,2] Gal β 1,3 – GalNAc α 1,3 – Bac α 1-;
- w. Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – GlcNAc β 1-;
- x. Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – GalNAc α 1-;
- y. Gal β 1,4[Fuc α 1-3]GlcNAc β 1,3 – Gal β 1,3 – Bac α 1-;
- z. Gal β 1,3 – GalNAc α 1,3 – GlcNAc β 1-;
- aa. Gal β 1,3 – GalNAc α 1,3 – Bac α 1-; and
- bb. Gal β 1,3 – GalNAc α 1,3 – GalNAc 1-.

- 20. The host cell of claim 1 or 2, wherein the host cell further comprises a gene encoding a heterologous protein of interest.
- 21. The host cell of claim 20, wherein the protein of interest comprises the oligosaccharide composition.
- 22. The host cell of claim 20, wherein the host cell further comprises an oligosaccharyl transferase activity capable of transferring the oligosaccharide composition onto an N-glycosylation acceptor site of the protein of interest.
- 23. An oligosaccharide composition produced by the host cell of claim 1 or 2.
- 24. A glycoprotein composition produced by the host cell of claim 20.
- 25. A cell culture comprising the host cell of claim 1 or 2.

26. A method for producing an oligosaccharide composition comprising culturing a recombinant host cell of claim 1 or 2.
27. A method for producing a glycoprotein composition comprising culturing a recombinant host cell of claim 20.