

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 October 2007 (25.10.2007)

PCT

(10) International Publication Number
WO 2007/120932 A2

(51) International Patent Classification:
C12P 21/06 (2006.01)

(21) International Application Number:
PCT/US2007/009782

(22) International Filing Date: 19 April 2007 (19.04.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/793,531 19 April 2006 (19.04.2006) US
60/842,926 6 September 2006 (06.09.2006) US

(71) Applicant (for all designated States except US): **NEOSE TECHNOLOGIES, INC.** [US/US]; 102 Witmer Road, Horsham, Pennsylvania 19044 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DEFREES, Shawn** [US/US]; 126 Filly Drive, North Wales, Pennsylvania

19454 (US). **SCHWARTZ, Marc F.** [US/US]; 324 Oak Lane, West Windsor, New Jersey 08555 (US). **JOHNSON, Karl** [US/US]; 5320 Ivystream Road, Hatboro, Pennsylvania 19040 (US).

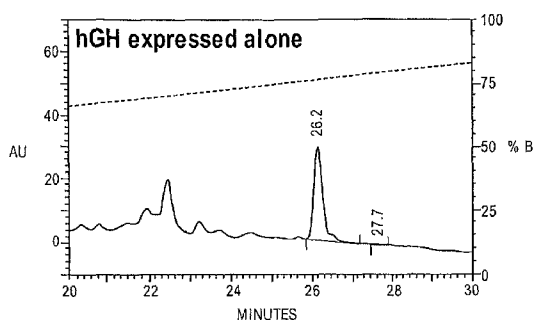
(74) Agents: **KELLY, Beth L.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111 (US).

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

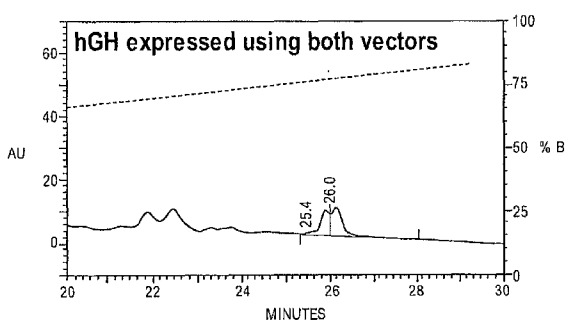
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: EXPRESSION OF O-GLYCOSYLATED THERAPEUTIC PROTEINS IN PROKARYOTIC MICROORGANISMS



A



B

(57) Abstract: The invention relates to methods of producing an O-glycosylated soluble therapeutic protein in a prokaryotic microorganism by co-expressing the therapeutic protein and a heterologous glycosyltransferase that transfers a sugar moiety to an amino acid acceptor on the therapeutic protein.

WO 2007/120932 A2



GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:

13 December 2007

(15) Information about Correction:

see PCT Gazette No. 50/2007 of 13 December 2007

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

EXPRESSION OF O-GLYCOSYLATED THERAPEUTIC PROTEINS IN PROKARYOTIC MICROORGANISMS

CROSS-REFERENCES TO RELATED APPLICATIONS[0001] This application claims the benefit of U.S. Provisional Application No. 60/793,531, filed April 19, 2006 and U.S. Provisional Application No. 60/842,926, filed September 6, 2006; both of which are herein incorporated by reference for all purposes.

FIELD OF THE INVENTION

[0002] The invention relates to methods of producing an O-glycosylated soluble therapeutic protein in a prokaryotic microorganism by expressing the therapeutic protein and at least one heterologous glycosyltransferase that transfers a sugar moiety to an amino acid acceptor on the therapeutic protein. The therapeutic protein and the heterologous glycosyltransferase can be expressed in separate microorganisms and combined after cell lysis or the therapeutic protein and the heterologous glycosyltransferase can be expressed in the same microorganism.

BACKGROUND OF THE INVENTION

[0003] One of the most efficient hosts for production of recombinant proteins are prokaryotic host cells. Many therapeutic proteins are being used to treat humans and other higher mammals. Yet in spite of the need for efficient and economic methods to produce such proteins, therapeutic proteins are produced in costly eukaryotic cell systems, *e.g.*, mammalian tissue culture cells, such as CHO cells; insect cells, and yeast cells. The main obstacles to use of prokaryotic production systems are production of insoluble therapeutic proteins in many prokaryotic cells, *e.g.*, *E. coli*, and failure of prokaryotic cells to provide appropriate post-translation modification of eukaryotic proteins, *e.g.*, glycosylation of eukaryotic proteins. Thus, at present production of therapeutic proteins in prokaryotic hosts must include labor intensive and expensive steps of refolding misfolded proteins, purification of refolded therapeutic proteins and purification of refolded glycosyltransferases. The present invention solves these and other needs.

BRIEF SUMMARY OF THE INVENTION

[0004] In one aspect the present invention provides a method of producing an O-glycosylated soluble therapeutic protein in a prokaryotic microorganism by expressing a

soluble therapeutic protein and a heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase protein in the prokaryotic microorganism, and growing the microorganism under conditions that allow intracellular transfer of a first sugar moiety from a first donor substrate to an amino acid acceptor substrate on the therapeutic protein catalyzed
5 by the heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase protein, to produce the O-glycosylated soluble therapeutic protein.

[0005] In one embodiment, the prokaryotic microorganism has an intracellular oxidizing environment. Examples of prokaryotic microorganisms include, *e.g.*, *E. coli* and *Pseudomonas* bacterium. In one embodiment, the prokaryotic microorganism is genetically
10 modified to have the intracellular oxidizing environment. Genetic manipulation of *E. coli*, by mutating an endogenous reductase nucleic acid is known to affect the redox state of the cell and produce an intracellular oxidizing environment.

[0006] In one embodiment, the heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase is a soluble active eukaryotic N-acetyl galactosaminyl transferase
15 (GalNAcT) protein.

[0007] In another embodiment, the method has an additional step of expressing a first heterologous soluble active glycosyltransferase in the prokaryotic microorganism and allowing it to catalyze intracellular transfer of a sugar moiety from a donor substrate to O-linked sugar on the therapeutic protein. Exemplary heterologous soluble active
20 glycosyltransferase proteins include, *e.g.*, a eukaryotic core I galactosyltransferase (Core 1 GalT1) protein and a ST6 GalNAc 1 protein. In a further embodiment, the method has an additional step of expressing a second heterologous soluble active glycosyltransferase in the prokaryotic microorganism and allowing it to catalyze intracellular transfer of a sugar moiety from a donor substrate to a second acceptor substrate, *e.g.*, part of the O-linked glycan, on the
25 therapeutic protein. Exemplary second heterologous soluble active glycosyltransferase protein include, *e.g.*, a eukaryotic $\alpha(2,3)$ sialyltransferase (ST3Gal1) protein, and a bacterial $\alpha(2,3)$ sialyltransferase protein.

[0008] In another embodiment, the microorganism is grown in a medium that comprises a precursor of the donor substrate, *e.g.*, GalNAc, GlcNAc, glucose, or sialic acid. The growth
30 medium can be a rich medium, *e.g.*, LB, or a minimal medium used for growth of microorganisms. The microorganism, *e.g.*, *E. coli*, can also be genetically modified to

enhance production of the donor substrate. Examples of such genetic modifications are included herein.

[0009] In still another embodiment, the N-glycosylated soluble therapeutic protein is isolated from the microorganism. The N-glycosylated soluble therapeutic protein can be produced on a commercial scale. The N-glycosylated therapeutic protein can be further modified in vitro by, *e.g.*, addition of a PEG moiety.

[0010] In another embodiment, the prokaryotic microorganism also expresses an accessory enzyme that has a role in synthesis of any donor substrate needed. The accessory enzyme can be *e.g.*, a UDP-glucose 4' epimerase protein, a UDP-GlcNAc 4' epimerase protein or a dual function UDP-glucose 4' epimerase protein/ UDP-GlcNAc 4' epimerase protein.

[0011] In another aspect the invention provides a prokaryotic microorganism that expresses both the O-glycosylated soluble therapeutic protein and the heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase. In some embodiments, both the O-glycosylated soluble therapeutic protein and the heterologous soluble active nucleotide sugar: polypeptide glycosyltransferase are expressed in the intracellular space of the microorganism. In one embodiment, the heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase is expressed in the intracellular space and glycosylates a therapeutic protein in the intracellular space. The O-glycosylated therapeutic protein is then transported through the bacterial membrane to the periplasm where refolding occurs. In some embodiments, the prokaryotic microorganism has an intracellular oxidizing environment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 demonstrates glycosylation of mutant human growth hormone (hGH) either as purified material (left) or in a crude cell lysate containing soluble hGH (right). hGH was produced in a *trxB*, *gor*, *supp E. coli* strain. GalNAc was transferred to the hGH protein after addition of reaction substrates and a crude cell lysate from a *trxB*, *gor*, *supp E. coli* strain that expressed human GalNAc-T2 $\Delta 51 \Delta 445$ protein and an epimerase protein, GalE. Reaction products were analyzed by MALDI TOF mass spectrometry and compared to unglycosylated hGH. The expected mass due to addition of GalNAc (expected +203.2, observed +207 or +209) to hGH from either source was observed.

[0013] Figure 2 demonstrates further glycosylation of hGH-GalNAc. The starting material originated from purified hGH (left) or a crude cell lysate containing hGH (right). hGH was produced in a *trxB*, *gor*, *supp E. coli* strain. GalNAc was first transferred to hGH as

described in Figure 1. Galactose was transferred to the hGH-GalNAc protein after addition of reaction substrates and a crude cell lysate from a *trxB*, *gor*, *supp E. coli* strain that expressed MBP-Core-1-GalT1 protein. Reaction products were analyzed by MALDI TOF mass spectrometry and compared to unglycosylated hGH. The expected mass due to addition of GalNAc-Gal (expected +365.6, observed +350 or +356) to hGH was observed.

[0014] Figure 3 provides analysis of a hGH GlycoPEGylation reaction utilizing crude lysates of hGH, GalNAcT2 $\Delta 51 \Delta 445$, GalE, MBP-Core-1-GalT1 and MBP-ST3Gal1 produced in *trxB*, *gor*, *supp E. coli*. hGH-GalNAc-Gal was prepared as described in Figures 1-2 was subsequently glycoPEGylated after addition of reaction substrates and a crude cell lysate from a *trxB*, *gor*, *supp E. coli* strain that expressed MBP-ST3Gal1 protein. Figure 3A is a SDS-PAGE analysis; colloidal blue stain. Lane 1: molecular weight markers; Lane 2: GlycoPEGylation reaction at 0 min; Lane 3: GlycoPEGylation reaction at 19 hours; Lane 4: GlycoPEGylation reaction at 46 hours; Lane 5: GlycoPEGylation reaction at 62 hours. The hGH-GalNAc-Gal-SA-cys-PEG-40 kDa product, boxed, is observed in Lanes 3, 4 and 5. The unmodified hGH band is observed in Lanes 2, 3, 4, and 5 at the 22 kDa band. Figure 3B provides RP-HPLC analysis of the GlycoPEGylation reaction at 62 hours. The retention time of the product hGH-GalNAc-Gal-SA-cys-PEG-40 kDa is 23.2 minutes. The retention time of the hGH starting material is 25.7 minutes. Zorbax 300SB-C3 150 x 2.1 mm, 5 micron column.

[0015] Figure 4 demonstrates glycoPEGylation of hGH-GalNAc-Gal produced in crude lysates by purified ST3Gal1. The reaction scheme is presented in Figure 4A. The glycosylated starting material originated from a crude cell lysate of hGH produced in a *trxB*, *gor*, *supp E. coli* strain. hGH-GalNAc-Gal protein prepared using crude lysate mixing (eg Figures 1-2) was glycoPEGylated with sialic acid-cys-PEG-40kDa after addition of reaction substrates and a purified ST3Gal1 protein that was produced in a baculoviral system. Reactions were incubated for up to eighty hours at 20°C or 32°C. Reaction products were analyzed by SDS-PAGE; the final timepoint is shown in Figure 4B. The hGH-GalNAc-Gal-SA-cys-PEG-40 kDa product is boxed. RP-HPLC analysis of the final 32°C glycoPEGylation reaction is shown in Figure 4C. The retention time of the product hGH-GalNAc-Gal-SA-cys-PEG-40 kDa is 23.0 minutes. The retention time of the hGH starting material is 25.5 minutes. Zorbax 300SB-C3 150 x 2.1 mm, 5 micron column. A timecourse of hGH glycoPEGylation is provided in Figure 4D. In the best reaction, up to seventy percent of the starting material was converted to hGH-GalNAc-Gal-SA-cys-PEG-40 kDa.

[0016] Figure 5 shows the results of reverse phase HPLC (RP-HPLC) analysis of hGH-GalNAc-Gal-SA-cys-PEG-40 kDa after purification by Q-sepharose fast flow chromatography followed by size exclusion chromatography.

5 [0017] Figure 6 shows a reaction scheme for production of glycoPEGylated hGH using lysates from two fermentations. Soluble, active hGH protein is expressed at high levels in a trxB, gor, supp *E. coli* strain. Separately, necessary glycosyltransferases, eg GalNAc-T2 $\Delta 445$, Core-1-Gal-T1, ST3Gal1, and accessory enzymes, eg GalE or GNE, are all coexpressed in the same trxB, gor, supp *E. coli* strain. After fermentation, the hGH cells and enzyme cells are lysed and mixed together in the presence of appropriate reaction substrates
10 to form the glycoPEGylated product.

[0018] Figure 7 demonstrates *in vivo* glycosylation of hGH by GalNAc-T2 $\Delta 51 \Delta 445$ co-expressed in a trxB, gor, supp *E. coli* strain. Epimerases GalE or GNE were also expressed in the cells, in order to compare their *in vivo* activities. Figure 7 shows the results of the co-expression as analyzed by RP-HPLC. Figure 7A is a control RP-HPLC of crude lysate from
15 trxB, gor, supp *E. coli* expressing hGH alone. Figure 7B shows RP-HPLC of crude lysate from trxB, gor, supp *E. coli* coexpressing hGH with GalNAc-T2 $\Delta 51 \Delta 445$ and GalE. The peak shift in the lower graph, as compared to the control, indicates that the hGH was glycosylated *in vivo*.

[0019] Figure 8 summarizes the expression levels and percent yield of hGH and hGH-GalNAc in crude lysates produced by the coexpression of hGH with GalNAcT2 and GalE or GNE.
20

[0020] Figure 9 shows LC/MS of the *in vivo* glycosylation of hGH that verifies the glycosylation reaction occurred. Crude cell lysates from a trxB, gor, supp *E. coli* strain expressing hGH, GalNAc-T2 $\Delta 51 \Delta 445$, and either GNE (Figure 9A) or GalE (Figure 9B)
25 were analyzed. The expected mass due to addition of GalNAc (expected +203.2, observed +203 for both) to hGH was observed.

[0021] Figure 10 shows a reaction scheme for production of glycosylated hGH using a single fermentation. The hGH protein is coexpressed in a trxB, gor, supp *E. coli* strain with, here, GalE, GalNAc-T2 $\Delta 445$, and Core-1-Gal-T1. After fermentation, the cells are lysed
30 and, if desired, mixed with an ST3Gal1 protein and CMP-sialic acid-cys-PEG-40kDa to form the hGH-GalNAc-Gal-SA-cys-PEG-40 kDa product.

[0022] Figure 11 shows the yields of bacterially produced hGH and hGH-GalNAc. The yields of hGH and glycosylated hGH in crude lysate for two polycistronic expression cassettes expressing hGH, GalNAcT2 $\Delta 53 \Delta 445$, Core-1-GalT1 $\Delta 50$, and GNE in *trxB*, *gor*, *supp E. coli* were determined by RP-HPLC and LC/MS and compared.

5 [0023] Figure 12 demonstrates the soluble crude lysate activity levels of GalNAcT2 and Core-1-GalT1 expressed from polycistronic expression cassettes in *trxB*, *gor*, *supp E. coli*. The mock control is from cells handled in parallel bearing an empty expression vector.

[0024] Figure 13 provides the distribution of total hGH as nonglycosylated hGH (clear), hGH-GalNAc (light gray), and hGH-GalNAc-Gal (dark gray) when crude coexpression
10 lysate was reacted in vitro with additional sugar nucleotides. hGH was coexpressed with glycosyltransferases and accessory enzymes in *trxB*, *gor*, *supp E. coli* using a polycistronic expression cassette. The mock treated sample was handled in parallel but did not receive exogenous sugar nucleotide.

[0025] Figure 14 provides an analysis of the progression of glycosylation *in vivo* when
15 produced from a polycistronic expression cassette in *trxB gor supp mutant E. coli*. hGH was coexpressed with glycosyltransferases and accessory enzymes in *trxB*, *gor*, *supp E. coli* using a polycistronic expression cassette. Glycosylation of hGH in cell lysate at each timepoint was monitored by LC/MS. Figure 14A shows the distribution of *in vivo* glycosylated hGH product over an induction time course from 16 to 40 hours. Figure 14B shows distribution of
20 *in vivo* glycosylated hGH product after 66 hours of induction.

[0026] Figure 15 provides the comparison of soluble lysate activity levels of catalytic domain GalNAc-T2 substitution mutants relative to non-mutated GalNAc-T2 when expressed in *trxB gor supp mutant E. coli*.

[0027] Figure 16 provides examples of some naturally-occurring O-glycan structures.

25 [0028] Figure 17 provides examples of glycosyltransferases that can be used to conjugate sugars to amino acids, *e.g.*, Serine or Threonine. O-glycan structures are also provided.

[0029] Figure 18 provides a list of exemplary nucleotide sugar : polypeptide glycosyltransferase proteins from CAZy family 27.

DEFINITIONS

[0030] The recombinant glycosyltransferase proteins produced by the methods of the invention are useful for transferring a saccharide from a donor substrate to an acceptor substrate. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include, but are not limited to, biologically significant molecules such as carbohydrates, proteins (*e.g.*, glycoproteins), and lipids (*e.g.*, glycolipids, phospholipids, sphingolipids and gangliosides).

The following abbreviations are used herein:

Ara = arabinosyl;

Fru = fructosyl;

Fuc = fucosyl;

Gal = galactosyl;

GalNAc = N-acetylgalactosylamino;

Glc = glucosyl;

GlcNAc = N-acetylglucosylamino;

Man = mannosyl; and

NeuAc = sialyl (N-acetylneuraminyl)

FT or FucT = fucosyltransferase*

ST = sialyltransferase*

GalT = galactosyltransferase*

[0031] Arabic or Roman numerals are used interchangeably herein according to the naming convention used in the art to indicate the identity of a specific glycosyltransferase (*e.g.*, FTVII and FT7 refer to the same fucosyltransferase).

[0032] Oligosaccharides are considered to have a reducing end and a non-reducing end, whether or not the saccharide at the reducing end is in fact a reducing sugar. In accordance with accepted nomenclature, oligosaccharides are depicted herein with the non-reducing end on the left and the reducing end on the right.

[0033] The term "sialic acid" refers to any member of a family of nine-carbon carboxylated sugars. The most common member of the sialic acid family is N-acetyl-neuraminic acid (2-keto-5-acetamido-3,5-dideoxy-D-glycero-D-galactononulopyranos-1-onic acid (often abbreviated as Neu5Ac, NeuAc, or NANA). A second member of the family is N-glycolyl-neuraminic acid (Neu5Gc or NeuGc), in which the N-acetyl group of NeuAc is hydroxylated.

A third sialic acid family member is 2-keto-3-deoxy-nonulosonic acid (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* 261: 11550-11557; Kanamori *et al.*, *J. Biol. Chem.* 265: 21811-21819 (1990)). Also included are 9-substituted sialic acids such as a 9-O-C₁-C₆ acyl-Neu5Ac like 9-O-lactyl-Neu5Ac or 9-O-acetyl-Neu5Ac, 9-deoxy-9-fluoro-Neu5Ac and 9-azido-9-deoxy-
5 Neu5Ac. For review of the sialic acid family, *see, e.g.*, Varki, *Glycobiology* 2: 25-40 (1992); *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, New York (1992)). The synthesis and use of sialic acid compounds in a sialylation procedure is disclosed in international application WO 92/16640, published October 1, 1992.

[0034] An "acceptor substrate" for a glycosyltransferase is an oligosaccharide moiety that
10 can act as an acceptor for a particular glycosyltransferase. When the acceptor substrate is contacted with the corresponding glycosyltransferase and sugar donor substrate, and other necessary reaction mixture components, and the reaction mixture is incubated for a sufficient period of time, the glycosyltransferase transfers sugar residues from the sugar donor substrate to the acceptor substrate. The acceptor substrate will often vary for different types of a
15 particular glycosyltransferase. For example, the acceptor substrate for a mammalian galactoside 2-L-fucosyltransferase (α 1,2-fucosyltransferase) will include a Gal β 1,4-GlcNAc-R at a non-reducing terminus of an oligosaccharide; this fucosyltransferase attaches a fucose residue to the Gal via an α 1,2 linkage. Terminal Gal β 1,4-GlcNAc-R and Gal β 1,3-GlcNAc-R and sialylated analogs thereof are acceptor substrates for α 1,3 and α 1,4-fucosyltransferases,
20 respectively. These enzymes, however, attach the fucose residue to the GlcNAc residue of the acceptor substrate. Accordingly, the term "acceptor substrate" is taken in context with the particular glycosyltransferase of interest for a particular application. Acceptor substrates for additional glycosyltransferases, are described herein. Acceptor substrates also include *e.g.*, glycolipids, peptides, proteins, glycopeptides, glycoproteins and therapeutic proteins.

25 [0035] An "amino acid acceptor substrate" is an amino acid in a protein or peptide that is conjugated to a sugar moiety in a reaction catalyzed by a nucleotide sugar : polypeptide glycosyltransferase protein. Amino acid acceptor substrates for nucleotide sugar : polypeptide glycosyltransferase protein include, *e.g.*, threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids.

30 [0036] A "donor substrate" for glycosyltransferases is an activated nucleotide sugar. Such activated sugars generally consist of uridine, guanosine, and cytidine monophosphate derivatives of the sugars (UMP, GMP and CMP, respectively) or diphosphate derivatives of

the sugars (UDP, GDP and CDP, respectively) in which the nucleoside monophosphate or diphosphate serves as a leaving group. For example, a donor substrate for fucosyltransferases is GDP-fucose. Donor substrates for sialyltransferases, for example, are activated sugar nucleotides comprising the desired sialic acid. For instance, in the case of NeuAc, the activated sugar is CMP-NeuAc. Other donor substrates include *e.g.*, GDP mannose, UDP-galactose, UDP-*N*-acetylgalactosamine, CMP-NeuAc-PEG (also referred to as CMP-sialic acid-PEG), UDP-*N*-acetylglucosamine, UDP-glucose, UDP-glucuronic acid, and UDP-xylose. Sugars include, *e.g.*, NeuAc, mannose, galactose, *N*-acetylgalactosamine, *N*-acetylglucosamine, glucose, glucuronic acid, and xylose. Bacterial, plant, and fungal systems can sometimes use other activated nucleotide sugars.

[0037] A "method of remodeling a protein, a peptide, a glycoprotein, or a glycopeptide" as used herein, refers to addition of a sugar residue to a protein, a peptide, a glycoprotein, or a glycopeptide using a glycosyltransferase. In a preferred embodiment, the sugar residue is covalently attached to a PEG molecule.

[0038] A "glycosyltransferase" as used herein refers to an enzyme that catalyzes transfer of a sugar residue from a donor substrate, *i.e.*, from an activated nucleotide sugar, to an acceptor substrate, *e.g.*, an oligosaccharide, a glycolipid, a peptide, a protein, a glycopeptide, or a glycoprotein. In preferred embodiments, a glycosyltransferase transfers a sugar from a donor substrate, *i.e.*, a nucleotide sugar, to a peptide, a protein, a glycopeptide, or a glycoprotein. In another preferred embodiment, a eukaryotic glycosyltransferase is a type II transmembrane glycosyltransferase. Unmodified type II transmembrane glycosyltransferases typically include an amino terminal cytoplasmic domain, a signal-anchor or transmembrane domain, a stem region, and a catalytic domain. *See, e.g.*, Paulson and Colley, *J. Biol. Chem.* 264:17615-17618 (1989). Many unmodified type II transmembrane glycosyltransferases are associated with cellular membranes and, therefore, fractionate with insoluble material when expressed in eukaryotic cells. Removal of the amino terminal cytoplasmic domain, signal-anchor or transmembrane domain, and the stem region of a eukaryotic glycosyltransferase results in an active, soluble enzyme that does not associate with the membrane fraction of a eukaryotic cell. *See, e.g.*, Paulson *et al.*, U.S. Patent No. 5,032,519, filed October 24, 1989; which is herein incorporated by reference for all purposes. A glycosyltransferase can be derived from an eukaryotic organism, *e.g.*, a unicellular or multicellular eukaryotic organism, a plant, an invertebrate animal, such as *Drosophila* or *C. elegans*, a vertebrate animal, an amphibian or reptile, a mammal, a rodent, a primate, a human, a rabbit, a rat, a mouse, a cow,

or a pig and so on. Examples of many eukaryotic glycosyltransferases are at, *e.g.*, *See, e.g.*, U.S. Provisional Application No. 60/665,396, filed March 24, 2005; U.S. Provisional Application No. 60/668,899, filed April 5, 2005; U.S. Provisional Application No. 60/732,409, filed October 31, 2005; and International Application PCT/US06/11065, filed March 24, 2006; each of which are herein incorporated by reference for all purposes. Prokaryotic glycosyltransferases can also be used in the methods of the invention.

[0039] A "sialyltransferase" as used herein, refers to an enzyme that catalyzes the transfer of a sialic acid moiety from a CMP-sialic acid donor to an acceptor molecule. Eukaryotic sialyltransferases can also be recognized by the presence of conserved structural motifs, *e.g.*, a sialyl motif L and a sialyl motif S as described in Tsuji, *J. Biochem.* 120:1-13 (1996), which is herein incorporated by reference for all purposes. Additional sialyltransferase motifs, *e.g.*, the very small (VS) motif and motif III, are described in Patel and Balaji, *Glycobiology*, 16:108-116 (2006), e-published October 5, 2005, which is herein incorporated by reference for all purposes. Sialyltransferases include enzymes that form a variety of linkages including $\alpha 2 \rightarrow 3$, $\alpha 2 \rightarrow 6$, $\alpha 2 \rightarrow 8$. Sialyltransferases transfer the sialic acid moiety to different acceptor sugars on an acceptor molecule, *e.g.*, galactose, GalNAc, and another sialic acid molecule. Eukaryotic sialyltransferases that catalyze specific reaction, *i.e.*, that are members of the ST3Gal, ST6Gal, ST6GalNAc, or ST8Sia families can be identified by the presence of amino acid residues conserved within those families. Such family-based conserved amino acid residues are disclosed at Patel and Balaji, *Glycobiology*, 16:108-116 (2006), e-published October 5, 2005, which is herein incorporated by reference for all purposes. Prokaryotic sialyltransferases are known and can also be used in the disclosed methods.

[0040] An $\alpha(2,3)$ sialyltransferase is a sialyltransferase that transfers sialic acid to an acceptor substrate by forming an $\alpha(2,3)$ linkage. One example of a eukaryotic $\alpha(2,3)$ sialyltransferase is the ST3Gal3 protein. This enzyme catalyzes the transfer of sialic acid to the Gal of a Gal β 1,3GlcNAc, Gal β 1,3GalNAc or Gal β 1,4GlcNAc glycoside (*see, e.g.*, Wen *et al.* (1992) *J. Biol. Chem.* 267: 21011; Van den Eijnden *et al.* (1991) *J. Biol. Chem.* 256: 3159). The sialic acid is linked to a Gal with the formation of an α -linkage between the two saccharides. Bonding (linkage) between the saccharides is between the 2-position of NeuAc and the 3-position of Gal. Like other eukaryotic glycosyltransferases, ST3Gal3 enzymes have a transmembrane domain, a stem region, and a catalytic domain. This particular enzyme can be isolated from rat liver (Weinstein *et al.* (1982) *J. Biol. Chem.* 257:

13845); the human cDNA (Sasaki *et al.* (1993) *J. Biol. Chem.* 268: 22782-22787; Kitagawa & Paulson (1994) *J. Biol. Chem.* 269: 1394-1401) and genomic (Kitagawa *et al.* (1996) *J. Biol. Chem.* 271: 931-938) DNA sequences are known, facilitating production of this enzyme by recombinant expression. Rat ST3Gal3 has been cloned and the sequence is known. See,
5 *e.g.*, Wen *et al.*, *J. Biol. Chem.* 267:21011-21019 (1992) and Accession number M97754, each of which are herein incorporated by reference. Prokaryotic $\alpha(2,3)$ sialyltransferases are also known *e.g.*, enzymes from *Neisseria* species and CstI, CstII, and CstIII enzymes from *Campylobacter*, and can be used in the methods of the invention.

[0041] A "eukaryotic α -N-acetylgalactosaminide α -2,6-sialyltransferase I (ST6GalNAcT1 or ST6GalNAc-1) as used herein, refers to an $\alpha(2,6)$ sialyltransferase isolated from a
10 eukaryotic organism. The enzyme catalyzes the transfer of sialic acid from a CMP-sialic acid donor to an acceptor molecule. The transfer is an $\alpha 2,6$ -linkage to N-acetylgalactosamine-O-Thr/Ser. Like other eukaryotic glycosyltransferases, ST6GalNAcT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of ST6GalNAcT1
15 enzymes have been isolated and characterized, *e.g.*, the full length mouse sequence, Kurosawa *et al.*, *J. Biochem.* 127:845-854 (2000) and accession number JC7248, each of which is herein incorporated by reference. Prokaryotic sialyltransferases that catalyze α -2,6 linkages are also known and can be used in the methods of the invention.

[0042] A "eukaryotic Gal $\beta 1,3$ GalNAc $\alpha 2,3$ -sialyltransferase (ST3GalI or ST3Gal-1)" as
20 used herein, refers to a Gal $\beta 1,3$ GalNAc $\alpha 2,3$ -sialyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of sialic acid from a CMP-sialic acid donor to an acceptor molecule. The transfer is an $\alpha 2,3$ -linkage to N-acetylgalactosamine-O-Thr/Ser. Like other eukaryotic glycosyltransferases, ST3GalI enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of ST3GalI enzymes have been isolated and
25 characterized, *e.g.*, the full length porcine sequence, Gillespie *et al.*, *J. Biol. Chem.* 267:21004-21010 (1992) and accession number A45073, each of which are herein incorporated by reference.

[0043] Other sialyltransferases that can be used in the present invention include, *e.g.*, "eukaryotic beta galactoside alpha 2, 6- sialyltransferase (ST6Gal I) proteins.

[0044] Eukaryotic sialyltransferase proteins used in the invention also include alpha 2,8
30 sialyltransferase proteins, *e.g.*, ST8Sia I, ST8Sia II, ST8Sia III, and ST8Sia IV.

[0045] Prokaryotic $\alpha(2,3)$ sialyltransferases are also known, *e.g.*, enzymes from *Neisseria* species and CstI, CstII, and CstIII enzymes from *Campylobacter*, and can be used in the methods of the invention. Exemplary *Neisseria* sialyltransferase are disclosed at, *e.g.*, U.S. Patent No. 6,096,529, issued August 1, 2000; which is herein incorporated by reference for all purposes. Exemplary *Campylobacter* sialyltransferase are disclosed at, *e.g.*, U.S. Patent No. 6,689,604, issued February 10, 2004, U.S. Patent No. 6,503,744, issued January 7, 2003, and U.S. Patent No. 6,699,705, issued March 2, 2004; each of which is herein incorporated by reference for all purposes. Other exemplary bacterial sialyltransferases are disclosed at International Application PCT/CA2005/001432, filed September 16, 2005; which is herein incorporated by reference for all purposes. Prokaryotic $\alpha(2,6)$ sialyltransferases are also known and can be used in the methods of the invention. One example of a prokaryotic $\alpha(2,6)$ sialyltransferase is from *Photobacterium*. See, *e.g.*, Yamamoto *et al.*, *J. Biochem.* 120:104-110 (1996).

[0046] A "eukaryotic *N*-acetylglucosaminyltransferase" as used herein, refers to an *N*-acetylglucosaminyltransferase derived from a eukaryotic organism. The enzyme catalyzes the transfer of *N*-acetylglucosamine (GlcNAc) from a UDP-GlcNAc donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, *N*-acetylglucosaminyltransferase has a transmembrane domain, a stem region, and a catalytic domain.

[0047] A " β -1,2-*N*-eukaryotic *N*-acetylglucosaminyltransferase I (GnTI or GNTI)" as used herein, refers to a β -1,2-*N*-acetylglucosaminyltransferase I derived from a eukaryotic organism. Like other eukaryotic glycosyltransferases, GnTI has a transmembrane domain, a stem region, and a catalytic domain. Eukaryotic GnTI proteins include, *e.g.*, human, accession number NP_002397; Chinese hamster, accession number AAK61868; rabbit, accession number AAA31493; rat, accession number NP_110488; golden hamster, accession number AAD04130; mouse, accession number P27808; zebrafish, accession number AAH58297; *Xenopus*, accession number CAC51119; *Drosophila*, accession number NP_525117; *Anopheles*, accession number XP_315359; *C. elegans*, accession number NP_497719; *Physcomitrella patens*, accession number CAD22107; *Solanum tuberosum*, accession number CAC80697; *Nicotiana tabacum*, accession number CAC80702; *Oryza sativa*, accession number CAD30022; *Nicotiana benthamiana*, accession number CAC82507; and *Arabidopsis thaliana*, accession number NP_195537, each of which are herein incorporated by reference. Other eukaryotic *N*-acetylglucosaminyltransferase proteins that

can be used in the present invention are include, *e.g.*, BGnT-I, GnT-II, GnT-III, GnT-IV (*e.g.*, GnT-IVa and GnT-IVb), GnT-V, GnT-VI, and GnT-IVH.

[0048] Other eukaryotic *N*-acetylglucosaminyltransferase proteins can be produced using the methods of the present invention and include, *e.g.*, maniac fringe protein, and MGNT1 proteins.

[0049] A "nucleotide sugar : polypeptide glycosyltransferase protein" as used herein, refers to a glycosyltransferase protein that catalyzes the transfer of a sugar moiety from a donor substrate to an amino acid acceptor. The amino acid acceptor is an amino acid that includes a hydroxyl group, *e.g.*, threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids. For eukaryotic nucleotide sugar : polypeptide glycosyltransferase proteins, the amino acid acceptor is typically a serine or threonine residues. For prokaryotic nucleotide sugar : polypeptide glycosyltransferase proteins, more variability is seen in the amino acid acceptor; the most commonly used amino acid residues are threonine, serine, and tyrosine. Eukaryotic nucleotide sugar : polypeptide glycosyltransferase proteins include the GalNAcT proteins described below. Some examples pf nucleotide sugar : polypeptide glycosyltransferase proteins are disclosed at, *e.g.*, CAZy family 27 (Figure 18).

[0050] A "eukaryotic *N*-acetylgalactosaminyltransferase (GalNAcT)" as used herein, refers to an *N*-acetylgalactosaminyltransferase isolated from a eukaryotic organism. The enzyme catalyzes the transfer of *N*-acetylgalactosamine (GalNAc) from a UDP-GalNAc donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, GalNAcT enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of GalNAcT enzymes have been isolated and characterized, *e.g.*, GalNAcT1, accession number X85018; GalNAcT2, accession number X85019 (both described in White *et al.*, *J. Biol. Chem.* 270:24156-24165 (1995)); and GalNAcT3, accession number X92689 (described in Bennett *et al.*, *J. Biol. Chem.* 271:17006-17012 (1996), each of which are herein incorporated by reference). At present eukaryotic 20 GalNAcT proteins, *i.e.*, GalNAcT1-20) are known and can be used in the methods of the invention.

[0051] A "eukaryotic galactosyltransferase as used herein, refers to a galactosyltransferase derived from a eukaryotic organism. The enzyme catalyzes the transfer of galactose from a UDP-Gal donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, galactosyltransferases have a transmembrane domain, a stem region, and a catalytic domain.

[0052] A "eukaryotic β -1,4-galactosyltransferase (GalT1) as used herein, refers to a β -1,4-galactosyltransferase derived from a eukaryotic organism. The enzyme catalyzes the transfer of galactose from a UDP-Gal donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, GalT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of GalT1 enzymes have been isolated and characterized, *e.g.*, the full length bovine sequence, D'Agostaro *et al.*, *Eur. J. Biochem.* 183:211-217 (1989) and accession number CAA32695, each of which are herein incorporated by reference.

[0053] A "eukaryotic core I galactosyltransferase (Core 1 GalT1 or Core-1-Gal-T1)" as used herein refers to a protein with Core 1 β 1,3-Galactosyltransferase activity. Like other eukaryotic glycosyltransferases, Core 1 GalT1 enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of Core 1 GalT1 enzymes have been isolated and characterized, *e.g.*, the *Drosophila* and human sequences. The human protein is characterized in Ju *et al.*, *J. Biol. Chem.* 277 (1), 178-186 (2002), which is herein incorporated by reference for all purposes.

[0054] A " β -1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (Core 2 GlcNAcT or Core-2-GlcNAcT)" as used herein refers to a protein with Core 2 β 1,6-GlcNActransferase activity. Like other eukaryotic glycosyltransferases, Core 2 GlcNAcT enzymes have a transmembrane domain, a stem region, and a catalytic domain. A number of Core 2 GlcNAcT enzymes have been isolated and characterized and are disclosed at, *e.g.*, CAZy family 14. The CAZy family 14 includes examples of prokaryotic proteins with Core 2 β 1,6-GlcNActransferase activity. Such prokaryotic proteins can also be used in the invention.

[0055] A "eukaryotic fucosyltransferase" as used herein, refers to a fucosyltransferase derived from a eukaryotic organism. The enzyme catalyzes the transfer of fucose from a UDP-fucose donor to an acceptor molecule. Like other eukaryotic glycosyltransferases, fucosyltransferases have a transmembrane domain, a stem region, and a catalytic domain.

[0056] Other eukaryotic glycosyltransferase proteins that can be used in the present invention include, *e.g.*, dolichyl-phosphate mannosyltransferase polypeptide 1, or Dpm1, α -1,6-mannosyltransferase, α -1,3-mannosyltransferase, and β -1,4-mannosyltransferase proteins.

[0057] Many prokaryotic glycosyltransferases are also known, *e.g.*, galactosyltransferases, GalNAc transferases, GlcNAc transferases, mannosyltransferases, glucosyltransferases, xylosyltransferases, and fucosyltransferases, and can be used in the methods of the invention.

[0058] A "therapeutic protein" as used herein, refers a protein, peptide, glycoprotein or glycopeptide that is administered to a subject to treat disease or dysfunction or to improve health of the subject. In a preferred embodiment the subject is a human. In a further preferred embodiment, the therapeutic protein is a human protein. A soluble therapeutic protein refers to a therapeutic protein that is soluble in an aqueous solution. In some embodiments the soluble therapeutic protein is soluble in an intracellular compartment of a prokaryotic cell. Most of the therapeutic protein or a fraction of the expressed therapeutic protein can be soluble in the intracellular compartment of a prokaryotic cell. In another embodiment the soluble therapeutic protein is an active protein, *e.g.*, has enzymatic activity, binding activity, or ability to elicit an immune response in a mammal, *e.g.*, a human. In an additional embodiment, the therapeutic protein is glycosylated or otherwise modified by one or more glycosyltransferases produced in a microorganism that has an oxidizing intracellular environment. In another embodiment, the therapeutic protein is produced in a microorganism that has an oxidizing intracellular environment and is glycosylated by one or more heterologous, *e.g.*, eukaryotic, glycosyltransferases produced the same microorganism. In one embodiment, a therapeutic protein is one of the following: FGF-20, FGF-21, NT-3, or glucocerebrosidase. In another embodiment, a therapeutic protein is one of the following: human FGF-20, human FGF-21, human NT-3 or human glucocerebrosidase. Therapeutic proteins are not limited to those listed above, and can include, *e.g.*, proteins listed in Table 1.

[0059] An "O-glycosylated therapeutic protein" or an "O-glycosylated soluble therapeutic protein" as used herein refer to a therapeutic protein that has been modified to include at least one sugar residue conjugated to a hydroxyl group on an amino acid of the therapeutic protein.

[0060] An "unpaired cysteine residue" as used herein, refers to a cysteine residue, which in a correctly folded protein (*i.e.*, a protein with biological activity), does not form a disulfide bond with another cysteine residue.

[0061] A "redox couple" refers to mixtures of reduced and oxidized thiol reagents and include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)).

[0062] The term "oxidant" or "oxidizing agent" refers to a compound which oxidizes molecules in its environment, *i.e.*, which changes the molecules in its environment to become more oxidized and more oxidizing. An oxidant acts by accepting electrons, thereby becoming itself reduced after having oxidized a substrate. Thus, an oxidant is an agent which accepts
5 electrons.

[0063] The term "oxidizing conditions" or "oxidizing environment" refers to a condition or an environment in which a substrate is more likely to become oxidized than reduced. For example, the periplasm of a wild type *E. coli* cell constitutes an oxidizing environment, whereas the cytoplasm of a wild type *E. coli* cell is a reducing environment. In an oxidizing
10 environment, there is a greater likelihood that disulfide bonds will form.

[0064] An enzyme in an "oxidized state" refers to an enzyme that has fewer electrons than its reduced form.

[0065] The term "reductant" or "reducing agent" refers to a compound which reduces molecules in its environment, *i.e.*, which changes molecules in its environment to become
15 more reduced and more reducing. A reducing agent acts by donating electrons, thereby becoming itself oxidized after having reduced a substrate. Thus, a reducing agent is an agent which donates electrons. Examples of reducing agents include dithiothreitol (DTT), mercaptoethanol, cysteine, thioglycolate, cysteamine, glutathione, and sodium borohydride.

[0066] The term "reductase" refers to a thioredoxin reductase, glutathione or glutathione reductase (also referred to as "oxidoreductases") or any other enzyme that can reduce
20 members of the thioredoxin or glutaredoxin systems.

[0067] The term "reductase pathways" refers to the systems in cells which maintain the environment in reducing conditions, and includes the glutaredoxin system and the thioredoxin system.

[0068] The term "reducing conditions" or "reducing environment" refers to a condition or an environment in which a substrate is more likely to become reduced than oxidized. For example, the cytoplasm of a eukaryotic cell constitutes a reducing environment.
25

[0069] "Disulfide bond formation" or "disulfide bond oxidation", used interchangeably herein, refers to the process of forming a covalent bond between two cysteines present in one
30 or two polypeptides. Oxidation of disulfide bonds can be mediated by thiol-disulfide exchange between the active site cysteines of enzymes and cysteines in the target protein.

Disulfide bond formation can be catalyzed by enzymes which are referred to as catalysts of disulfide bond formation or can be catalyzed by chemical means, *e.g.*, an intracellular environment.

[0070] An enzyme in a "reduced state", has more electrons than its oxidized form.

5 [0071] "Disulfide bond reduction" refers to the process of cleaving a disulfide bond, thereby resulting in two thiol groups. Reduction of disulfide bonds is mediated by thiol-disulfide exchange between the active site cysteines of enzymes and cysteines in the target protein.

10 [0072] The term "disulfide bond isomerization" refers to an exchange of disulfide bonds between different cysteines, *i.e.*, the shuffling of disulfide bonds. Isomerization of disulfide bonds is mediated by thiol-disulfide exchange between the active site cysteines of enzymes and cysteines in the target protein and catalyzed by isomerases. In *E. coli*, isomerization is catalyzed by DsbC or DsbG a periplasmic disulfide bond oxidoreductase.

15 [0073] A "catalyst of disulfide bond formation" is an agent which stimulates disulfide bond formation. Such an agent must be in an oxidized state to be active.

[0074] A "catalyst of disulfide bond isomerization", also referred to as a "disulfide bond isomerase" is an agent which stimulates disulfide bond isomerization. Such an agent must be in a reduced form to be active.

20 [0075] The term "contacting" is used herein interchangeably with the following: combined with, added to, mixed with, passed over, incubated with, flowed over, etc.

[0076] "Chaperone proteins" are proteins that are known to promote proper folding of newly synthesized proteins. Chaperone proteins include, *e.g.*, trigger factor; members of the Hsp70 chaperone family, *e.g.* DnaK; members of the Hsp100 chaperone family, *e.g.* ClpB, and members of the Hsp60 chaperone family, *e.g.* GroEL. *See, e.g.*, Sorensen and
25 Mortensen, *BioMed Central*, www.microbialcellfactories.com/content/4/1/1. Chaperones are also known that allow protein folding at 4°C, *e.g.*, Cpn60 and Cpn 10 from *Oleispira antartica* RB8^T. *See, e.g., Id.* and Ferrer *et al.*, *Nat. Biotechnol.* 21:1266-1267 (2003).

[0077] "Protein disulfide isomerases" or "PDI proteins" can make or shuffle disulfide bonds. PDI proteins are described *e.g.*, in Georgiou *et al.* U.S. Patent No. 6,027,888, which
30 is herein incorporated by reference for all purposes. PDI proteins are derived from

eukaryotic and prokaryotic organisms. Eukaryotic PDI proteins include those of the Interpro family IPR005792 Protein disulphide isomerase. Exemplary eukaryotic PDI proteins include PDI proteins from *e.g.*, rat liver PDI, Ero1p and Pdi1p proteins from *Sacchromyces*.

5 Prokaryotic proteins include *e.g.*, DsbC from *E. coli*. See, *e.g.*, Frand *et al.*, *Trends in Cell Biol.* 10:203-210 (2000).

[0078] Other prokaryotic proteins that act to maintain the redox state of protein disulfide bonds include, *e.g.*, DsbB, DsbA, DsbC, DsbD, and DsbG from *E. coli*. These proteins are well known in the art and are described in, *e.g.*, Beckwith *et al.* U.S. Patent No. 6,872,563, which is herein incorporated by reference for all purposes.

10 [0079] The term "PEG" refers to poly(ethylene glycol). PEG is an exemplary polymer that has been conjugated to peptides. The use of PEG to derivatize peptide therapeutics has been demonstrated to reduce the immunogenicity of the peptides and prolong the clearance time from the circulation. For example, U.S. Pat. No. 4,179,337 (Davis *et al.*) concerns non-immunogenic peptides, such as enzymes and peptide hormones coupled to polyethylene
15 glycol (PEG) or polypropylene glycol. Between 10 and 100 moles of polymer are used per mole peptide and at least 15% of the physiological activity is maintained.

[0080] The term "specific activity" as used herein refers to the catalytic activity of an enzyme, *e.g.*, a recombinant glycosyltransferase of the present invention, and may be expressed in activity units. As used herein, one activity unit catalyzes the formation of 1
20 μmol of product per minute at a given temperature (*e.g.*, at 37°C) and pH value (*e.g.*, at pH 7.5). Thus, 10 units of an enzyme is a catalytic amount of that enzyme where 10 μmol of substrate are converted to 10 μmol of product in one minute at a temperature of, *e.g.*, 37 °C and a pH value of, *e.g.*, 7.5.

[0081] "N-linked" oligosaccharides are those oligosaccharides that are linked to a peptide
25 backbone through asparagine, by way of an asparagine-N-acetylglucosamine linkage. N-linked oligosaccharides are also called "N-glycans." Naturally occurring N-linked oligosaccharides have a common pentasaccharide core of $\text{Man}_3\text{GlcNAc}_2$. They differ in the presence of, and in the number of branches (also called antennae) of peripheral sugars such as N-acetylglucosamine, galactose, N-acetylgalactosamine, fucose and sialic acid. Optionally,
30 this structure may also contain a core fucose molecule and/or a xylose molecule. Using the soluble eukaryotic glycosyltransferases produced by the methods of the invention, oligosaccharides can be produced that mimic natural N-linked structures or that are designed

by the user. Soluble eukaryotic glycosyltransferases that generate N-linked oligosaccharides include, *e.g.*, GnT1, GalT1, and ST3Gal3 enzymes.

[0082] "O-linked" oligosaccharides are those oligosaccharides that are linked to a peptide backbone through threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids. Using the soluble eukaryotic glycosyltransferases produced by the methods of the invention, oligosaccharides can be produced that mimic natural O-linked structures or that are designed by the user. In one embodiment of the invention, soluble eukaryotic glycosyltransferases that generate O-linked oligosaccharides are expressed in one or more microorganisms that have an oxidizing, intracellular environment. Soluble eukaryotic glycosyltransferases that generate O-linked oligosaccharides include, *e.g.*, GalNAc-T2, Cor-1-Gal-T1, ST6GalNAc-1, and ST3Gal-1 enzymes.

[0083] A "substantially uniform glycoform" or a "substantially uniform glycosylation pattern," when referring to a glycoprotein species, refers to the percentage of acceptor substrates that are glycosylated by the glycosyltransferase of interest (*e.g.*, fucosyltransferase). It will be understood by one of skill in the art, that the starting material may contain glycosylated acceptor substrates. Thus, the calculated amount of glycosylation will include acceptor substrates that are glycosylated by the methods of the invention, as well as those acceptor substrates already glycosylated in the starting material.

[0084] The term "biological activity" refers to, *e.g.*, an enzymatic activity of a protein. For example, biological activity of a sialyltransferase refers to the activity of transferring a sialic acid moiety from a donor molecule to an acceptor molecule. Biological activity of a GalNAcT2 refers to the activity of transferring an *N*-acetylgalactosamine moiety from a donor molecule to an acceptor molecule. For GalNAcT2 proteins, an acceptor molecule can be a protein, a peptide, a glycoprotein, or a glycopeptide. Biological activity of a GnT1 protein refers to the activity of transferring a *N*-acetylglucosamine moiety from a donor molecule to an acceptor molecule. Biological activity of a galactosyltransferase refers to the activity of transferring a galactose moiety from a donor molecule to an acceptor molecule. Other biological activities include binding by *e.g.*, a hormone or receptor, induction of a second messenger system by *e.g.*, a cytokine, and ability to elicit an immune response.

[0085] "Commercial scale" refers to gram scale production of a glycosylated product in a single reaction. In preferred embodiments, commercial scale refers to production of at least about 0.2, 0.5, 1, 2, 5, 10, 15, 25, 50, 75, 80, 90 or 100, 125, 150, 175, 200, 500 or 1000

grams a glycosylated product in a single reaction. Commercial scale production of a soluble, active O-glycosylated therapeutic protein refers to gram scale production of a soluble, active O-glycosylated therapeutic protein.

[0086] The term “substantially” in the above definitions of “substantially uniform”

generally means at least about 60%, at least about 70%, at least about 80%, or more preferably at least about 90%, and still more preferably at least about 95% of the acceptor substrates for a particular glycosyltransferase are glycosylated.

[0087] The term “amino acid” refers to naturally occurring and unnatural amino acids, *e.g.*,

synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an

R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0088] “Protein”, “polypeptide”, or “peptide” refer to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a polypeptide. When the amino acids are α -amino acids, either the L-optical isomer or the D-optical isomer can be used. Additionally, unnatural amino acids, for example, β -alanine, phenylglycine and homoarginine are also included. Amino acids that are not gene-encoded may also be used in the present invention. Furthermore, amino acids that have been modified to include reactive groups may also be used in the invention. All of the amino acids used in the present invention may be either the D - or L -isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see*, Spatola, A. F., in CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983).

[0089] The term “recombinant” when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found
5 in the native form of the cell wherein the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques. A “recombinant protein” is one which has been produced by a recombinant cell.
10 In preferred embodiments, a recombinant eukaryotic glycosyltransferase is produced by a recombinant bacterial cell.

[0090] A “fusion protein” refers to a protein comprising amino acid sequences that are in addition to, in place of, less than, and/or different from the amino acid sequences encoding the original or native full-length protein or subsequences thereof. More than one additional
15 domain can be added to a glycosyltransferase as described herein, *e.g.*, an accessory domain and an epitope tag or purification tag, or multiple epitope tags or purification tags.

[0091] Components of fusion proteins include “accessory enzymes” and/or “purification tags.” An “accessory enzyme” as referred to herein, is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate for a glycosyltransferase. An
20 accessory enzyme can, for example, catalyze the formation of a nucleotide sugar that is used as a donor moiety by a glycosyltransferase. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar. Examples of accessory enzymes, and fusion of accessory enzymes are disclosed, *e.g.*, in PCT application
25 CA98/01180, filed on December 15, 1998.

[0092] The recombinant proteins of the invention can be constructed and expressed as a fusion protein with a molecular “purification tag” at one end, which facilitates purification of the protein. Such tags can also be used for immobilization of a protein of interest during the glycosylation reaction. Suitable tags include “epitope tags,” which are a protein sequence
30 that is specifically recognized by an antibody. Epitope tags are generally incorporated into fusion proteins to enable the use of a readily available antibody to unambiguously detect or isolate the fusion protein. A “FLAG tag” is a commonly used epitope tag, specifically

recognized by a monoclonal anti-FLAG antibody, consisting of the sequence

AspTyrLysAspAspAsp AspLys or a substantially identical variant thereof. Other epitope tags that can be used in the invention include, *e.g.*, myc tag, AU1, AU5, DDDDK (EC5), E tag, E2 tag, Glu-Glu, a 6 residue peptide, EYMPME, derived from the Polyoma middle T protein, HA, HSV, IRS, KT3, S tage, S1 tag, T7 tag, V5 tag, VSV-G, β -galactosidase, Gal4, green fluorescent protein (GFP), luciferase, protein C, protein A, cellulose binding protein, GST (glutathione S-transferase), a step-tag, Nus-S, PPI-ases, Pfg 27, calmodulin binding protein, dsb A and fragments thereof, and granzyme B. Epitope peptides and antibodies that bind specifically to epitope sequences are commercially available from, *e.g.*, Covance Research Products, Inc.; Bethyl Laboratories, Inc.; Abcam Ltd.; and Novus Biologicals, Inc.

[0093] Other suitable purification tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide or other poly-histidine peptides, which will bind to metal ions such as nickel or cobalt ions. Proteins comprising purification tags can be purified using a binding partner that binds the purification tag, *e.g.*, antibodies to the purification tag, nickel or cobalt ions or resins, and amylose, maltose, or a cyclodextrin. Purification tags also include starch binding domains, *E. coli* thioredoxin domains (vectors and antibodies commercially available from *e.g.*, Santa Cruz Biotechnology, Inc. and Alpha Diagnostic International, Inc.), and the carboxy-terminal half of the SUMO protein (vectors and antibodies commercially available from *e.g.*, Life Sensors Inc.). Starch binding domains, such as a maltose binding domain from *E. coli* and SBD (starch binding domain) from an amylase of *A. niger*, are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacyclodextrin (BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0094] Recombinant proteins can also include a self-cleaving protein tag, such as an "intein". Inteins facilitate removal of, *e.g.*, a purification or epitope tag. Inteins and kits for their use are commercially available, *e.g.*, from New England Biolabs.

[0095] The term "functional domain" with reference to glycosyltransferases, refers to a domain of the glycosyltransferase that confers or modulates an activity of the enzyme, *e.g.*, acceptor substrate specificity, catalytic activity, binding affinity, localization within the Golgi apparatus, anchoring to a cell membrane, or other biological or biochemical activity. Examples of functional domains of glycosyltransferases include, but are not limited to, the

catalytic domain, stem region, signal-anchor or transmembrane domain, and amino-terminal cytoplasmic tail.

[0096] The terms “expression level” or “level of expression” with reference to a protein refers to the amount of a protein produced by a cell. The amount of protein produced by a cell can be measured by the assays and activity units described herein or known to one skilled in the art. One skilled in the art would know how to measure and describe the amount of protein produced by a cell using a variety of assays and units, respectively. Thus, the quantitation and quantitative description of the level of expression of a protein, *e.g.*, a glycosyltransferase, is not limited to the assays used to measure the activity or the units used to describe the activity, respectively. The amount of protein produced by a cell can be determined by standard known assays, for example, the protein assay by Bradford (1976), the bicinchoninic acid protein assay kit from Pierce (Rockford, Illinois), or as described in U.S. Patent No. 5,641,668. Another method of determining protein expression is to analyze a lysate or other sample containing the protein using gel electrophoresis, *e.g.*, SDS-PAGE, followed by a visualization step. Visualization steps include protein dyes and stains, *e.g.*, Comassie or silver stain, or immunoassays, such as western blot analysis using an antibody that will specifically bind to the protein of interest. Antibodies can be directed against the glycosyltransferase or against a purification or epitope tag covalently bound to the protein.

[0097] The term “enzymatic activity” refers to an activity of an enzyme and may be measured by the assays and units described herein or known to one skilled in the art. Examples of an activity of a glycosyltransferase include, but are not limited to, those associated with the functional domains of the enzyme, *e.g.*, acceptor substrate specificity, catalytic activity, binding affinity, localization within the Golgi apparatus, anchoring to a cell membrane, or other biological or biochemical activity.

[0098] A “stem region” with reference to glycosyltransferases refers to a protein domain, or a subsequence thereof, which in the native glycosyltransferases is located adjacent to the signal anchor or transmembrane domain, between the membrane region and the shortest catalytic domain, and has been reported to function as a retention signal to maintain the glycosyltransferase in the Golgi apparatus and as a site of proteolytic cleavage. Stem regions generally start with the first hydrophilic amino acid following the hydrophobic transmembrane domain and end at the catalytic domain, or in some cases the first cysteine residue following the transmembrane domain. Exemplary stem regions include, but are not

limited to, the stem region of fucosyltransferase VI, amino acid residues 40-54; the stem region of mammalian GnT1, amino acid residues from about 36 to about 103 (see, *e.g.*, the human enzyme); the stem region of mammalian GalT1, amino acid residues from about 71 to about 129 (see *e.g.*, the bovine enzyme); the stem region of mammalian ST3GalIII, amino acid residues from about 29 to about 84 (see, *e.g.*, the rat enzyme); the stem region of invertebrate Core-1-Gal-T1, amino acid residues from about 36 to about 102 (see *e.g.*, the *Drosophila* enzyme); the stem region of mammalian Core-1-Gal-T1, amino acid residues from about 32 to about 90 (see *e.g.*, the human enzyme); the stem region of mammalian ST3Gal1, amino acid residues from about 28 to about 61 (see *e.g.*, the porcine enzyme) or for the human enzyme amino acid residues from about 18 to about 58; the stem region of mammalian ST6GalNAc-1, amino acid residues from about 30 to about 207 (see *e.g.*, the murine enzyme), amino acids 35-278 for the human enzyme or amino acids 37-253 for the chicken enzyme; the stem region of mammalian GalNAc-T2, amino acid residues from about 71 to about 129 (see *e.g.*, the rat enzyme).

[0099] A “catalytic domain” refers to a protein domain, or a subsequence thereof, that catalyzes an enzymatic reaction performed by the enzyme. For example, a catalytic domain of a sialyltransferase will include a subsequence of the sialyltransferase sufficient to transfer a sialic acid residue from a donor to an acceptor saccharide. A catalytic domain can include an entire enzyme, a subsequence thereof, or can include additional amino acid sequences that are not attached to the enzyme, or a subsequence thereof, as found in nature. An exemplary catalytic region is, but is not limited to, the catalytic domain of fucosyltransferase VII, amino acid residues 39-342; the catalytic domain of mammalian GnT1, amino acid residues from about 104 to about 445 (see, *e.g.*, the human enzyme); the catalytic domain of mammalian GalT1, amino acid residues from about 130 to about 402 (see *e.g.*, the bovine enzyme); and the catalytic domain of mammalian ST3Gal3, amino acid residues from about 85 to about 374 (see, *e.g.*, the rat enzyme). Catalytic domains and truncation mutants of GalNAc-T2 proteins are described in USSN 60/576,530 filed June 3, 2004; and US provisional patent application Attorney Docket Number 040853-01-5149-P1, filed August 3, 2004; both of which are herein incorporated by reference for all purposes. Catalytic domains can also be identified by alignment with known glycosyltransferases.

[0100] A “subsequence” refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (*e.g.*, protein) respectively.

[0101] A "glycosyltransferase truncation" or a "truncated glycosyltransferase" or grammatical variants, refer to a glycosyltransferase that has fewer amino acid residues than a naturally occurring glycosyltransferase, but that retains enzymatic activity. Truncated glycosyltransferases include, *e.g.*, truncated GnT1 enzymes, truncated GalT1 enzymes, truncated ST3GalIII enzymes, truncated GalNAc-T2 enzymes, truncated Core 1 GalT1 enzymes, amino acid residues from about 32 to about 90 (see *e.g.*, the human enzyme); truncated ST3Gal1 enzymes, truncated ST6GalNAc-1 enzymes, and truncated GalNAc-T2 enzymes. Any number of amino acid residues can be deleted so long as the enzyme retains activity. In some embodiments, domains or portions of domains can be deleted, *e.g.*, a signal-anchor domain can be deleted leaving a truncation comprising a stem region and a catalytic domain; a signal-anchor domain and a portion of a stem region can be deleted leaving a truncation comprising the remaining stem region and a catalytic domain; or a signal-anchor domain and a stem region can be deleted leaving a truncation comprising a catalytic domain. Glycosyltransferase truncations can also occur at the C-terminus of the protein. For example, some GalNAcT enzymes have a C-terminal lectin domain that can be deleted without diminishing enzymatic activity.

[0102] The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single-or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

[0103] A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of affecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to be transcribed (*e.g.*, a nucleic acid encoding a desired polypeptide), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette. In preferred embodiments, a

recombinant expression cassette encoding an amino acid sequence comprising a eukaryotic glycosyltransferase is expressed in a bacterial host cell.

5 [0104] A “heterologous sequence” or a “heterologous nucleic acid”, as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous glycoprotein gene in a eukaryotic host cell includes a glycoprotein-encoding gene that is endogenous to the particular host cell that has been modified. Modification of the heterologous sequence may occur, *e.g.*, by treating the DNA with a restriction enzyme to generate a DNA fragment that is capable of being operably linked to the promoter. Techniques such as site-directed mutagenesis are also useful
10 for modifying a heterologous sequence.

[0105] The term “isolated” refers to material that is substantially or essentially free from components which interfere with the activity of an enzyme. For a saccharide, protein, or nucleic acid of the invention, the term “isolated” refers to material that is substantially or essentially free from components which normally accompany the material as found in its
15 native state. Typically, an isolated saccharide, protein, or nucleic acid of the invention is at least about 80% pure, usually at least about 90%, and preferably at least about 95% pure as measured by band intensity on a silver stained gel or other method for determining purity. Purity or homogeneity can be indicated by a number of means well known in the art. For example, a protein or nucleic acid in a sample can be resolved by polyacrylamide gel
20 electrophoresis, and then the protein or nucleic acid can be visualized by staining. For certain purposes high resolution of the protein or nucleic acid may be desirable and HPLC or a similar means for purification, for example, may be utilized.

[0106] The term “operably linked” refers to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription
25 factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence.

[0107] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or protein sequences, refer to two or more sequences or subsequences that are the same
30 or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

[0108] The phrase “substantially identical,” in the context of two nucleic acids or proteins, refers to two or more sequences or subsequences that have at least greater than about 60% nucleic acid or amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

[0109] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0110] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat’l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally, Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)).

[0111] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschuel *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence

pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for
5 initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to
10 calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W , T , and X determine the sensitivity and speed of the alignment. The BLASTN
15 program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, $M=5$, $N=-4$, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

20 [0112] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid
25 sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0113] A further indication that two nucleic acid sequences or proteins are substantially
30 identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with the protein encoded by the second nucleic acid, as described below. Thus, a protein is typically substantially identical to a second protein, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences

are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

[0114] The phrase “hybridizing specifically to” refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA.

[0115] The term “stringent conditions” refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 15°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at T_m , 50% of the probes are occupied at equilibrium).

Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is typically at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42° C, or, 5x SSC, 1% SDS, incubating at 65° C, with wash in 0.2x SSC, and 0.1% SDS at 65° C. For PCR, a temperature of about 36° C is typical for low stringency amplification, although annealing temperatures may vary between about 32-48° C depending on primer length. For high stringency PCR amplification, a temperature of about 62° C is typical, although high stringency annealing temperatures can range from about 50° C to about 65° C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90-95° C for 30-120 sec, an annealing phase lasting 30-120 sec, and an extension phase of about 72° C for 1-2 min. Protocols and guidelines for low and high stringency amplification reactions are available,

e.g., in Innis, et al. (1990) *PCR Protocols: A Guide to Methods and Applications* Academic Press, N.Y.

[0116] The phrases “specifically binds to a protein” or “specifically immunoreactive with”, when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

[0117] “Conservatively modified variations” of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are “silent variations,” which are one species of “conservatively modified variations.” Every polynucleotide sequence described herein which encodes a protein also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and UGG which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each “silent variation” of a nucleic acid which encodes a protein is implicit in each described sequence.

[0118] Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids

(typically less than 5%, more typically less than 1%) in an encoded sequence are “conservatively modified variations” where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

5 [0119] One of skill will appreciate that many conservative variations of proteins, *e.g.*, glycosyltransferases, and nucleic acid which encode proteins yield essentially identical products. For example, due to the degeneracy of the genetic code, “silent substitutions” (*i.e.*, substitutions of a nucleic acid sequence which do not result in an alteration in an encoded protein) are an implied feature of every nucleic acid sequence which encodes an amino acid.
10 As described herein, sequences are preferably optimized for expression in a particular host cell used to produce the chimeric glycosyltransferases (*e.g.*, yeast, human, and the like). Similarly, “conservative amino acid substitutions,” in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties (*see*, the definitions section, *supra*), are also readily identified as being highly similar to a
15 particular amino acid sequence, or to a particular nucleic acid sequence which encodes an amino acid. Such conservatively substituted variations of any particular sequence are a feature of the present invention. *See also*, Creighton (1984) *Proteins*, W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are
20 also “conservatively modified variations”.

[0120] The practice of this invention can involve the construction of recombinant nucleic acids and the expression of genes in host cells, preferably bacterial host cells. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids such
25 as expression vectors are well known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques*, *Methods in Enzymology* volume 152 Academic Press, Inc., San Diego, CA (Berger); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between
30 Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1999 Supplement) (Ausubel). Suitable host cells for expression of the recombinant polypeptides are known to those of skill in the art, and include, for example, prokaryotic cells, such as *E. coli*, and eukaryotic cells including insect, mammalian and fungal cells (*e.g.*, *Aspergillus niger*)

[0121] Examples of protocols sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.* (1987) U.S. Patent No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874; Lomell *et al.* (1989) *J. Clin. Chem.* 35: 1826; Landegren *et al.* (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117. Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0122] The present invention provides for the first time methods of producing soluble, active O-glycosylated therapeutic proteins in prokaryotes. In some embodiments, the solubility of the therapeutic proteins is improved by expression in prokaryotes that have an intracellular oxidizing environment. The intracellular oxidizing environment enhances disulfide bond formation and permits correct folding of both the therapeutic proteins and heterologous glycosyltransferases. The activity and solubility of those proteins are generally more convenient to measure than protein refolding and disulfide bond formation. Therefore, protein activity and solubility are used as surrogates for determination of correct protein folding. The O-glycosylation is provided by co-expression of a heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase protein that transfers a sugar moiety to an amino acid acceptor on the therapeutic protein. The glycosylation of the therapeutic protein occurs in the cytoplasm of the prokaryote. In preferred embodiments, the amino acid acceptor is a serine or threonine residue. In this way the invention provides an efficient way of reproducing a system of eukaryotic post-translational modification in a prokaryotic cell and of making large amounts of soluble active O-glycosylated therapeutic protein in a prokaryotic cell.

[0123] Some therapeutic proteins are soluble in prokaryotic microorganisms that have a reducing environment, as are some glycosyltransferases, *e.g.*, prokaryotic

glycosyltransferases, including nucleotide sugar : polypeptide glycosyltransferase proteins.

Therefore, in some embodiments the methods of the invention can be performed using prokaryotic host cells that have a reducing environment, *e.g.*, expression of O-therapeutic proteins and heterologous glycosyltransferases that are soluble and active in a reducing

environment. One example of a prokaryotic microorganisms that has a reducing environment is an unmodified *E. coli* cell.

[0124] The present invention also provides for the first time prokaryotic cells that comprise a soluble, active O-glycosylated therapeutic protein and a heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase protein. The prokaryotic cells can have an intracellular oxidizing environment or an intracellular reducing environment. The heterologous soluble active nucleotide sugar : polypeptide glycosyltransferase protein, any other heterologous glycosyltransferases, and the soluble, active O-glycosylated therapeutic protein are expressed in the cytoplasm of the prokaryotic cell.

II. O-glycosylation of proteins

[0125] O-glycosylation is the conjugation of a sugar group to protein through an amino acid that has a hydroxyl group. O-glycosylation of proteins occurs in both eukaryotes and prokaryotes.

[0126] In eukaryotes O-glycosylation occurs in the endoplasmic reticulum at serine or threonine residues on a protein. Eukaryotic O-glycosylated proteins are found, *e.g.*, within the endoplasmic reticulum, the golgi apparatus, and are frequently secreted out of the cell. Sugars that are transferred to an amino acid O-glycosylation site on eukaryotic proteins include GalNAc, GlcNAc, mannose, xylose, glucose, and fucose. The most commonly added sugar is GalNAc, which forms a mucin type linkage with serine or threonine. GalNAc is transferred to proteins by a family of GalNAc transferases, which includes at least 24 members in human alone. *See, e.g.*, Ten Hagen *et al.*, *Glycobiol.* 13:1R-16R (2003). Additional sugars can be added to the first conjugated sugar by additional heterologous glycosyltransferases.

[0127] Prokaryotes, *i.e.*, archaea and bacteria, also have endogenous O-glycosylated proteins which typically are found on the outer layer of the cell, *e.g.*, as part of the S-layer, or on extra cellular proteins, such as flagellin and pillin. *See, e.g.*, Schaffer *et al.*, *Proteomics*, 1:248-261 (2001); Benz and Schmidt, *Mole. Microbiol.* 45:267-276 (2002); and Eichler,

Microbiol. 149:3347-3351 (2003). A variety of sugars are added to amino acids on prokaryotic proteins through O-linkages, *e.g.*, GalNAc, GlcNAc, mannose, xylose, glucose, and fucose. Additional sugars can be added to the first conjugated sugar by additional heterologous glycosyltransferases.

5 [0128] It can be useful for those of skill to determine the composition of an oligosaccharide added to a protein in an O-linkage and/or to determine the amino acid site of O-linkage on a particular protein. For example, one of skill can determine the composition and/or amino acid site of an O-linked oligosaccharide on a naturally occurring therapeutic protein, *e.g.*, a hormone or growth factor, before using the disclosed methods. Or one of skill can determine
10 the composition or amino acid site of a soluble, active O-glycosylated therapeutic protein after using the disclosed methods to verify the production of the desired product.

[0129] Methods to determine the composition and/or amino acid site of an O-linked oligosaccharide on a protein are known to those of skill in the art. See, *e.g.*, Schaffer *et al.*, *Proteomics*, 1:248-261 (2001). For example, glycoconjugates can be detected on proteins
15 that have been separated by SDS-PAGE, using *e.g.*, the Schiff staining method. Separated proteins can also be transferred to nylon or PVDF membrane and then assayed for the presence of glycoconjugates after labeling with digoxigenin. Glycosylation of proteins modifies their behavior on SDS-PAGE. Thus, a protein suspected of being glycosylated can be compared on SDS-PAGE to a non-glycosylated control. Lectins can also be used analyze
20 a protein believed to be glycosylated and to identify conjugated sugar moieties. The identity of an O-linked oligosaccharide moiety can be determined using, *e.g.*, gas chromatography, mass spectrometry techniques (MALDI-MS) or liquid chromatography-electrospray mass spectrometry. See, *e.g.*, Benz and Schmidt, *Mole. Microbiol.* 45:267-276 (2002).

[0130] Once a site for glycosylation and composition of the glycoconjugate have been
25 determined for a particular therapeutic protein, the appropriate heterologous glycosyltransferase(s) and prokaryotic cells can be selected for use in producing a soluble, active O-linked therapeutic protein. In preferred embodiments, the therapeutic protein and heterologous glycosyltransferase(s) are expressed in the cytoplasm of the prokaryotic cell and the O-linked glycosylation of the therapeutic protein occurs in the cytoplasm of the
30 prokaryotic cell. In a further preferred embodiment, the therapeutic protein and heterologous glycosyltransferase(s) are expressed in the cytoplasm of the prokaryotic cell and the therapeutic protein is glycosylated in the cytoplasm and then transported to the periplasm or

extracellular where disulfide bonds are formed in the therapeutic protein. In another preferred embodiment, transfer of the therapeutic protein and heterologous glycosyltransferase(s) in the periplasm or outside of the living prokaryotic cell is excluded from the invention.

5 III. Soluble, active O-glycosylated therapeutic proteins

[0131] Therapeutic proteins are typically recombinant eukaryotic protein that are administered to a subject, *e.g.*, a higher mammal and preferably a human, to treat a disease or to alleviate symptoms of a disease. Other uses for therapeutic proteins include, *e.g.*, use as a vaccine, that is eliciting an immune response in a subject. Therapeutic proteins are not typically produced commercially in prokaryotes such as bacteria. Reasons include, *e.g.*, inefficient production because of expression in inclusion bodies (insoluble protein) or production of proteins that are not glycosylated or that are not correctly glycosylated. The methods of the invention overcome these problems, first by enhancing the solubility of the therapeutic protein expressed in the prokaryote and second by allowing glycosylation of the therapeutic protein with in the prokaryotic cell. Therapeutic proteins include the proteins listed in Table 1, any of which can be O-glycosylated using the methods of disclosed herein.

Table 1. Preferred therapeutic proteins

<u>Hormones and Growth Factors</u>	<u>Receptors and Chimeric Receptors</u>
Granulocyte colony stimulating factor (G-CSF)	CD4
Granulocyte-macrophage colony stimulating factor (GM-CSF)	Tumor Necrosis Factor receptor (TNF-R)
TPO	TNF-R:IgG Fc fusion
Erythropoietin (EPO)	Alpha-CD20
EPO variants	PSGL-1
Follicle Stimulating Hormone (FSH)	Complement
Human Growth Hormone (HGH)	GlyCAM or its chimera
Insulin	N-CAM or its chimera
alpha-TNF	<u>Monoclonal Antibodies (Immunoglobulins)</u>
Leptin	MAB-anti-RSV
Human chorionic gonadotropin	MAB-anti-IL-2 receptor
Bone morphogenetic protein 2 (BMP-2)	MAB-anti-CEA
Bone morphogenetic protein 7 (BMP-7)	MAB-Glycoprotein IIb/IIIa (Reopro™)
Fibroblast growth factor-20 (FGF-7)	MAB-anti-EGF
Fibroblast growth factor-20 (FGF-20)	MAB-Her-2 (Herceptin™)
Fibroblast growth factor-20 (FGF-21)	MAB-CD20 (Rituxan™)
Neurotrophin-3	MAB-alpha-CD3
	TNF receptor-IgG Fc fusion (Enbrel™)
	MAB-TNFα (Remicade™)
	MAB-CD4
	MAB-PSGL-1

<u>Enzymes and Inhibitors</u>	<u>Mab-anti F protein of Respiratory</u>
Tissue-type plasminogen activator (TPA)	Syncytial Virus
TPA variants	Anti-thrombin-III
Urokinase	Others
Factor VII clotting factor	Hepatitis B surface antigen (HbsAg)
Factor VIII	Chimeric diphtheria toxin-IL-2
Factor IX clotting factor	<u>Transport Proteins</u>
Factor X	Transferrin
Factor XIII	
hrDNase	
Glucocerebrosidase (Cerezyme™)	
Hirudin	
α 1 antitrypsin (α 1 protease inhibitor)	
Antithrombin III	
Acid α -glucosidase (acid maltase)	
α galactosidase A	
α -L-iduronidase	
Urokinase	
<u>Cytokines and Chimeric Cytokines</u>	
Interleukin-1 (IL-1), 1B, 2, 3, 4,	
Interleukin-21	
Interleukin-22	
Interferon-alpha (IFN-alpha)	
Interferon -alpha-2b	
Interferon -beta	
Interferon -gamma	
Interferon -omega	

[0132] The amino acid sequences of many potential soluble, active O-glycosylated therapeutic proteins are known to those of skill. The amino acid site and composition of a glycoconjugate can be determined by *e.g.*, repeating all or part of the O-glycosylation pattern of a naturally occurring soluble, active O-glycosylated therapeutic protein. The amino acid site and composition of a glycoconjugate on a naturally occurring soluble, active O-glycosylated therapeutic protein can be determined empirically, as described above. Or those of skill can make use of a previously determined amino acid site and composition of a glycoconjugate on a naturally occurring soluble, active O-glycosylated therapeutic protein. Many are known and have been collected for use by those of skill at *e.g.*, O-GLYCBASE, a database of O-glycosylated proteins. *See, e.g.*, www.cbs.dtu.dk/databases/OGLYCBASE/.

[0133] Other preferred therapeutic proteins that can be produced using the method of the invention are disclosed in Application No. PCT/US02/32263, filed October 9, 2002; Provisional Patent Application No. 60/448,381, filed February 19, 2003; Provisional Patent

Application No. 60/438,582, filed January 6, 2003; Provisional Patent Application No. 60/407,527, filed August 28, 2002; Provisional Patent Application No. 60/404,249, filed August 16, 2002; Provisional Patent Application No. 60/396,594, filed July 17, 2002; Provisional Patent Application No. 60/391,777, filed June 25, 2002; Provisional Patent Application No. 60/387,292, filed June 7, 2002; Provisional Patent Application No. 60/334,301, filed November 28, 2001; Provisional Patent Application No. 60/334,233, filed November 28, 2001; Provisional Patent Application No. 60/344,692, filed October 19, 2001; and Provisional Patent Application No. 60/328,523, filed October 10, 2001; and in the following US Patent Application Publications 20040142856, 20040137557, 20040132640, 20040126838, 20040115168, 20040082026, 20040077836, 20040063911, 20040043446. The preferred therapeutic proteins in the above references are also referred to as preferred peptides for remodeling.

[0134] The amino acid sequence of the therapeutic proteins includes an O-linked glycosylation site. The O-linked glycosylation site can be a naturally occurring site, or can result from manipulation to introduce an O-linked glycosylation site into the amino acid sequence. Exemplary proteins with O-linked glycosylation sites include, *e.g.*, granulocyte colony stimulating factor (G-CSF), *e.g.*, 175 and 178 amino acid wild types (with or without N-terminal methionine residues), interferon (*e.g.*, interferon alpha, *e.g.*, interferon alpha 2b, or interferon alpha 2a), granulocyte macrophage colony stimulating factor (GM-CSF), human growth hormone, interleukin (*e.g.*, interleukin 2), and fibroblast growth factor (FGF). Examples of wild-type and mutant proteins and peptides are found in, *e.g.*, PCT/US2004/014254, filed May 7, 2004; U.S. Provisional Patent Application No. 60/469,114, filed May 9, 2003; U.S. Provisional Patent Application No. 60/494,751, filed August 13, 2003; U.S. Provisional Patent Application No. 60/495,076, filed August 14, 2003; U.S. Provisional Patent Application No. 60/535,290, filed January 8, 2003; PCT/US05/000799, filed January 10, 2005; USSN 11/033365 filed January 10, 2005; U.S. Provisional Patent Application No. 60/535284, filed January 8, 2004; U.S. Provisional Patent Application No. 60/544411, filed February 12, 2004; U.S. Provisional Patent Application No. 60/546631, filed February 20, 2004; U.S. Provisional Patent Application No. 60/555813, filed March 23, 2004; U.S. Provisional Patent Application No. 60/570891, filed May 12, 2004; PCT/US05/39226, filed October 31, 2005; and U.S. Provisional Patent Application No. 60/623342, filed October 29, 2004; each of which is herein incorporated by reference for all purposes.

[0135] Examples of synthetic O-glycosylation sites that can be inserted into an amino acid sequence are disclosed in US Provisional Patent Application 60/832,461, filed July 21, 2006; which is herein incorporated by reference for all purposes.

[0136] The invention also encompasses therapeutic proteins that have been modified to increase resistance to proteases. In one embodiment, the protease resistant therapeutic protein is a human growth hormone protein. Exemplary protease resistant therapeutic proteins are found in *e.g.*, U.S. Provisional Patent Application No. 60/669736, filed April 8, 2005; U.S. Provisional Patent Application No. 60/710401, filed August 22, 2005; and U.S. Provisional Patent Application No. 60/720030, filed September 23, 2005; each of which is herein incorporated by reference for all purposes.

[0137] After expression of the soluble, active O-glycosylated therapeutic protein using the methods of the invention, the soluble, active O-glycosylated therapeutic protein will preferably be an active protein. Those of skill will recognize how to determine the activity of an O-glycosylated therapeutic protein. Enzymatic assays of O-glycosylated enzymes or inhibitors, *e.g.*, clotting factors, can be performed to determine activity. Some O-glycosylated therapeutic protein can be assayed for biological activity. For example, O-glycosylated hormones or growth factors can be assayed for binding to an appropriate receptor or for an appropriate response in a cell based or animal model. O-glycosylated receptor proteins can be assayed for binding to an appropriate ligand or for an appropriate response in a cell based or animal model. O-glycosylated antibodies or immunoglobulin proteins can be assayed for binding to an appropriate antigen. O-glycosylated cytokines can be assayed for activity using an appropriate cell based model or animal model. O-glycosylated therapeutic proteins that are vaccines, can be assayed for an ability to elicit an immune response in a model animal or in a human.

[0138] In one embodiment, a soluble, active O-glycosylated therapeutic protein made by the methods described herein has activity levels, *e.g.*, U/cell or U/mg protein, up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than activity levels of the same therapeutic protein expressed in a microorganism with a reducing environment. In another embodiment, a soluble, active O-glycosylated therapeutic protein made by the methods described herein has activity levels, *e.g.*, U/cell or U/mg protein, up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than activity levels of the

same therapeutic protein expressed in a microorganism without a heterologous glycosyltransferase.

[0139] In a preferred embodiment, the soluble, active O-glycosylated therapeutic protein has improved therapeutic properties as compared to, *e.g.*, the therapeutic protein expressed in a prokaryotic or eukaryotic cell without expression of a heterologous glycosyltransferase. Improved therapeutic properties include, *e.g.*, increased bioavailability of the therapeutic protein in a subject, increased half life of the therapeutic protein in a subject, enhanced pharmacokinetics, enhanced pharmacodynamics, improved biodistribution, providing a polyvalent species, improved water solubility, enhanced or diminished lipophilicity, and tissue targeting. Tests for improved therapeutic properties can be performed in *e.g.*, humans and model animal systems, *e.g.*, rodents, cats, dogs, and non-human primates.

[0140] In one embodiment, a soluble, active O-glycosylated therapeutic protein made by the methods described herein has improved therapeutic properties, up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than those of the same therapeutic protein expressed in a microorganism with a reducing environment. In another embodiment, a soluble, active O-glycosylated therapeutic protein made by the methods described herein has improved therapeutic properties, up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than those of the same therapeutic protein expressed in a microorganism without a heterologous soluble, active glycosyltransferase.

IV. Soluble, active glycosyltransferases

[0141] Soluble, active glycosyltransferases are used to conjugate an O-linked sugar or oligosaccharide to a soluble active therapeutic protein of the invention. The conjugation occurs within the cytoplasm of a prokaryote that has an intracellular oxidizing environment. Soluble active glycosyltransferases include, *e.g.*, nucleotide sugar : polypeptide glycosyltransferase proteins and other glycosyltransferases.

[0142] At least one heterologous glycosyltransferase is used in the methods of the invention. The total number of glycosyltransferases will depend on the needs of the user. The glycosyltransferase can be a eukaryotic glycosyltransferase or a prokaryotic glycosyltransferase. Active eukaryotic glycosyltransferases can be produced in prokaryotes that have an oxidizing environment. *See, e.g.*, U.S. Provisional Application No. 60/665,396, filed March 24, 2005; U.S. Provisional Application No. 60/668,899, filed April 5, 2005; U.S.

Provisional Application No. 60/732,409, filed October 31, 2005; and International Application PCT/US06/11065, filed March 24, 2006; each of which are herein incorporated by reference for all purposes. In prokaryotes that have a reducing environment, *e.g.*, unmodified *E. coli* species, eukaryotic glycosyltransferases and other heterologous proteins frequently are insoluble and expressed in inclusion bodies. The ability to co-express heterologous soluble, active glycosyltransferases with the therapeutic protein of choice allows the *in vivo* O-linked glycosylation of the therapeutic protein to occur in the cytoplasm of the prokaryotic host cell.

[0143] The first glycosylation step is carried out by a nucleotide sugar : polypeptide glycosyltransferase protein, *i.e.*, a glycosyltransferase protein that transfers a sugar moiety from a donor substrate to an amino acid acceptor on the therapeutic protein. In eukaryotes, commonly used nucleotide sugar : polypeptide glycosyltransferase proteins include GalNAc transferase proteins. These proteins transfer a GalNAc moiety from UDP-GalNAc to a serine or threonine residue on a protein. In one embodiment, a single nucleotide sugar : polypeptide glycosyltransferase protein is used in the invention and a soluble, active O-glycosylated therapeutic protein comprising an O-linked amino acid is produced.

[0144] N-acetylgalactosaminyltransferases are of use in practicing the present invention, particularly for binding a GalNAc moiety to an amino acid of the O-linked glycosylation site of the peptide. Suitable N-acetylgalactosaminyltransferases include, but are not limited to, $\alpha(1,3)$ N-acetylgalactosaminyltransferase, $\beta(1,4)$ N-acetylgalactosaminyltransferases (Nagata *et al.*, *J. Biol. Chem.* **267**: 12082-12089 (1992) and Smith *et al.*, *J. Biol. Chem.* **269**: 15162 (1994)) and polypeptide N-acetylgalactosaminyltransferase (Homa *et al.*, *J. Biol. Chem.* **268**: 12609 (1993)).

[0145] Production of proteins such as the enzyme GalNAc T_{I-XX} from cloned genes by genetic engineering is well known. See, *eg.*, U.S. Pat. No. 4,761,371. One method involves collection of sufficient samples, then the amino acid sequence of the enzyme is determined by N-terminal sequencing. This information is then used to isolate a cDNA clone encoding a full-length (membrane bound) transferase which upon expression in the insect cell line Sf9 resulted in the synthesis of a fully active enzyme. The acceptor specificity of the enzyme is then determined using a semiquantitative analysis of the amino acids surrounding known glycosylation sites in 16 different proteins followed by *in vitro* glycosylation studies of synthetic peptides. This work has demonstrated that certain amino acid residues are

overrepresented in glycosylated peptide segments and that residues in specific positions surrounding glycosylated serine and threonine residues may have a more marked influence on acceptor efficiency than other amino acid moieties.

[0146] The second glycosylation step is carried out by a glycosyltransferase that can transfer a sugar moiety to the amino acid conjugated sugar on the soluble, active O-glycosylated therapeutic protein. In one embodiment the first glycosylation step is performed by a GalNAc transferase protein and a second glycosylation step is performed by a Core 1 Galactose transferase 1 protein (Core 1 Gal T1), which transfers a galactose moiety to the amino acid-conjugated GalNAc moiety. Alternatively, the second glycosyltransferase can be a sialyltransferase, *e.g.*, an ST6GalNAc1 protein.

[0147] The third glycosylation step is carried out by a glycosyltransferase that can transfer a sugar moiety to the terminal conjugated sugar on the soluble, active O-glycosylated therapeutic protein. For example, if the second glycosylation step was performed by a Core 1 Gal T1 protein, the third glycosylation step can be carried out by, *e.g.*, an ST3Gal1 protein to add a sialic acid moiety to the soluble, active O-glycosylated therapeutic protein. Additional or different steps can be performed as desired by the user.

[0148] In some embodiments, eukaryotic glycosyltransferases are used in all or some of the glycosylation steps. Any eukaryotic glycosyltransferase can be used in the methods of the present invention. The eukaryotic glycosyltransferases can be the naturally occurring, unmodified proteins or can be glycosyltransferases that have been modified to enhance catalytic activity, or stability, or other characteristics of the proteins. Modification of eukaryotic glycosyltransferases include *e.g.*, truncation of the protein to remove *e.g.*, the stem region, the signal-anchor domain, or a portion of the stem region or the signal-anchor domain, or removal of both the stem region and the signal-anchor domain; or removal of an unpaired cysteine residue by substitution to another amino acid residue. A glycosyltransferase can also be truncated at the C-terminus to remove a non-catalytic domain or domains. For example, a C-terminal lectin domain can be removed from GalNAcT enzymes without diminishing enzymatic activity. Modified glycosyltransferases are described *e.g.*, in USSN 60/542,210, filed February 4, 2004; USSN 60/599,406, filed August 6, 2004; USSN 60/627,406, filed November 12, 2004; USSN 60/576,433, filed June 3, 2004; USSN 60/650,011, filed February 4, 2005; PCT/US05/19583, filed June 3, 2005; USSN 60/576,530, filed June 3, 2004; USSN 60/598,584, filed August 3, 2004; PCT/US05/19442,

filed June 3, 2005; PCT/US05/03856, filed February 4, 2005; WO 2004/063344; U.S. Provisional Application No. 60/665,396, filed March 24, 2005; U.S. Provisional Application No. 60/668,899, filed April 5, 2005; U.S. Provisional Application No. 60/732,409, filed October 31, 2005; and International Application PCT/US06/11065, filed March 24, 2006; each of which is herein incorporated by reference for all purposes

[0149] Preferred embodiments of the invention include use of, *e.g.*, a eukaryotic *N*-acetylglucosaminyltransferase (GnTI or GNTI, GnTII or GNTII, GnTIII or GNTIII, GnTIV or GNTIV, GnTV or GNTV, GnTIV or GNTIV or a Core 2 GalNAcT); a eukaryotic *N*-acetylgalactosaminyltransferase (GalNAcT, *e.g.*, GalNAcT1, GalNAcT2, or GalNAcT3); any galactosyltransferase, *e.g.*, a eukaryotic β -1,4-galactosyltransferase (GalT1) or a eukaryotic core I galactosyltransferase (Core-1-Gal-T1); any eukaryotic sialyltransferase, *e.g.*, a eukaryotic α (2,3)sialyltransferase (ST3Gal3), or a eukaryotic α -N-acetylgalactosaminyl α -2,6-sialyltransferase I (ST6GalNAc-1), or a eukaryotic gal β 1,3GalNAc α 2,3-sialyltransferase (ST3Gal-1); and any eukaryotic fucosyltransferase. Many examples of proteins having the above listed activities are known, see, *e.g.*, afmb.cnrs-mrs.fr/CAZY/, in particular Glycosyltransferase Families 2, 4, 6, 7, 10, 1, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 34, 37, 38, 41, 42, 49, 52, 54, 65, or 68. Other preferred embodiments include use of prokaryotic glycosyltransferases.

[0150] As indicated above, the glycosyltransferases can be modified before production using the disclosed methods. Modifications include, *e.g.*, truncation of the glycosyltransferase to remove all or a portion of a non-catalytic domain, such as a cytoplasmic domain, a signal-anchor domain, a stem region, and or a lectin or ricin domain. Exemplary truncated glycosyltransferase that can be produced in the present invention include, *e.g.*, ST3Gal III (Δ 27, Δ 28, Δ 73, Δ 85, Δ 86), human GnTI (Δ 103), bovine GalT1 (Δ 40, Δ 129, Δ 70), human GalNAcT2 (Δ 51, Δ 40, Δ 73, Δ 94, Δ 51 Δ 445, Δ 53, Δ 53 Δ 445), ST3Gal1 (Δ 45), *Drosophila* Core-1-Gal-T1 (Δ 31, Δ 50), and human ST6GalNAc1 mutants as described in U.S. Provisional Application No. 60/665,396, filed March 24, 2005; U.S. Provisional Application No. 60/668,899, filed April 5, 2005; U.S. Provisional Application No. 60/732,409, filed October 31, 2005; and International Application PCT/US06/11065, filed March 24, 2006; each of which are herein incorporated by reference for all purposes.

[0151] In a preferred embodiment, the first glycosylation step is carried out by a eukaryotic GalNAcT protein. Such proteins are exemplified as CAZy family 27, *see, e.g.*, Figure 18.

GalNAcT proteins have been aligned and the alignments used to identify domains within the proteins, *e.g.*, transmembrane, stem, catalytic and lectin or ricin domains. *See, e.g.*, Ten Hagen *et al. Glycobiology* 13:1R-16R (2003), which is herein incorporated by reference for all purposes. Such alignments can also be used by those of skill to identify truncations, substitutions and other modifications that can be made while retaining activity of the proteins. *See, e.g.*, Ten Hagen *et al.*, above.

[0152] CAZy family 27 also includes bacterial proteins that can be used in the invention, as can eukaryotic or prokaryotic polypeptide : nucleotide sugar transferases that transfer different sugars, *e.g.*, GlcNAc, glucose, fucose, and xylose to an amino acid.

[0153] In further embodiments, prokaryotic glycosyltransferases are used to carry out one or more glycosylation steps of the invention. Bacterial glycosyltransferases that can be used in the invention include, *e.g.*, sialyltransferases, galactosyltransferases, fucosyltransferases, and GalNAc transferases. Examples of bacterial glycosyltransferases are disclosed at, *e.g.*, US Patent No. 6,503,744, issued January 7, 2003; US Patent No. 6,699,705, issued March 2, 2004; US Patent No. 6,096,529, issued August 1, 2000; US Patent No. 6,689,604, issued February 10, 2004; International Application No. PCT/US2005/001614 filed January 21, 2005; International Application No. PCT/CA2005/001432, filed September 16, 2005; USSN 60/670608, filed April 11, 2005; and USSN 60/764171, filed January 31, 2006.

V. O-linked glycan structures

[0154] For a review of O-linked glycan structures, see Schachter and Brockhausen, *The Biosynthesis of Branched O-Linked Glycans*, 1989, Society for Experimental Biology, pp. 1-26 (Great Britain).

[0155] The following are exemplary O-linked glycan structures:

[0156] Ser/Thr- α -1-GalNAc[$\rightarrow$ β -1,6 $\rightarrow$ GlcNAc $\rightarrow$ β -1,4 $\rightarrow$ Gal $\rightarrow$ α 2,3-Sia] β -1,3-Gal

[0157] Ser/Thr- α -1-GalNAc[$\rightarrow$ β -1,6 $\rightarrow$ GlcNAc $\rightarrow$ β -1,4 $\rightarrow$ Gal $\rightarrow$ α 2,3-Sia] β -1,3-Gal- α 2,3-Sia

[0158] Ser/Thr- α -1-GalNAc $\rightarrow$ β -1,3-Gal- α 2,3-Sia

[0159] Ser/Thr- α -1-GalNAc $\rightarrow$ β -1,3-Gal

[0160] Ser/Thr- α -1-GalNAc- α 2,6-Sia

[0161] Ser/Thr- α -1-GalNAc

[0162] The O-glycan structure can be made using any appropriate combination(s) of eukaryotic and prokaryotic glycosyltransferases. Examples of naturally-occurring O-glycan structures are provided in Figure 16. However, those of skill will recognize that alternative O-glycan structures can be made using appropriate glycosyltransferases and activated sugar substrates.

[0163] The identity of the sugar transferred to an amino acid residue is not limited to GalNAc. Other sugars can be transferred to oxygen-containing amino acid residues using appropriate glycosyltransferases, as shown in Figure 17. Thus, GalNAc is transferred to an amino acid residue by GalNAc transferases; GlcNAc is transferred to an amino acid residue by GlcNAc transferases; fucose is transferred to an amino acid residue by fucosyltransferases; glucose is transferred to an amino acid residue by glucosyltransferases; and mannose is transferred to an amino acid residue by mannosyltransferases.

[0164] While Figures 16 and 17 demonstrate O-glycosylation on serine or threonine residues, other oxygen-containing amino acid residues can accept sugar molecules, *e.g.*, hydroxyproline. In addition, glucose can be enzymatically conjugated to tryptophan residues.

[0165] An example of an O-glycosylation protocol, beginning with transfer of GalNAc to a serine or threonine residue follows. The GalNAc residue is conjugated to the serine or threonine residue by a heterologous GalNAcT protein. Exemplary GalNAcT proteins include eukaryotic GalNAcT1-20 and bacterial enzymes and are listed in Figure 18.

[0166] The galactose residue is conjugated to the GalNAc sugar by a heterologous galactosyltransferase protein. Exemplary galactosyltransferase protein include eukaryotic core 1 Gal T1 proteins and bacterial proteins that conjugate galactose to GalNAc in a β -1,3 linkage. Sialic acid residues can be added to the galactose sugar by a sialyltransferase protein with α 2,3 activity. Exemplary sialyltransferase proteins include ST3 Gal1, ST3 Gal2 or *Campylobacter* CstI protein.

[0167] The GlcNAc residue is conjugated to GalNAc sugar by a heterologous GlcNAcT protein. Exemplary GlcNAcT proteins are disclosed at, *e.g.*, CAZy family 14. The CAZy family 14 includes examples of eukaryotic and prokaryotic proteins with Core 2 β 1,6-GlcNAc transferase activity that can be used in the invention. The galactose residue is conjugated to the GlcNAc sugar by a heterologous galactosyltransferase protein specific for a β (1-4) linkage, *e.g.*, a eukaryotic Gal T1 or Gal T7 or a *Neisseria* lgtB protein. Other

bacterial proteins are members of CAZy family 82. Sialic acid residues can be added to the galactose sugar by a sialyltransferase protein with $\alpha 2,3$ activity. Exemplary sialyltransferase proteins include ST3 Gal2, ST3 Gal3, ST3 Gal4 or *Campylobacter* CstI protein.

[0168] Additional modifications can occur on the terminal sialic acid residues, *e.g.*,
5 addition of sialic acid in an $\alpha 2,8$ linkage. Exemplary proteins include *e.g.*, ST8Sia I, ST8Sia II, ST8Sia III, ST8Sia IV or *Campylobacter* CstII protein. Fucose residues can also be added using, *e.g.*, eukaryotic fucosyltransferases V, VI, VIII, or *Helicobacter* fucosyltransferases.

VI. Intracellular, oxidizing environments in prokaryotes

[0169] In preferred embodiments, soluble, active O-glycosylated therapeutic proteins and
10 heterologous soluble, active glycosyltransferases are co-expressed in prokaryotic organisms that have intracellular oxidizing environments.

A. Prokaryotic microorganisms that have oxidizing intracellular environments

[0170] The method of the invention are carried out using prokaryotic microorganisms that have oxidizing intracellular environments. Such microorganisms include prokaryotic
15 microorganisms that have endogenous, intracellular oxidizing environments and prokaryotic microorganisms that are genetically manipulated to have an intracellular oxidizing environment.

[0171] Some prokaryotic organisms have endogenous, intracellular oxidizing environments and, thus, promote formation of protein disulfide bonds inside the cell. Oxidizing
20 intracellular compartments in prokaryotic organisms specifically exclude a bacterial periplasmic space. Prokaryotic organisms that have endogenous, intracellular oxidizing environments can be used in to produce soluble, active eukaryotic glycosyltransferases in an intracellular compartment. Prokaryotic organisms with endogenous, intracellular oxidizing environments include members of *e.g.*, *Pseudomonas* species, including *testosteroni*, *putida*,
25 *aeruginosa*, *syringae*, and *fluorescens*; some gram positive bacteria; and some gram negative bacteria. Additional *Pseudomonas* species and strains are described in, *e.g.*, U.S. Patent Application Publication No. US 2005/0186666, published August 25, 2005, which is herein incorporated by reference for all purposes. Gram positive bacteria include, *e.g.*, *Bacillus*, *Listeria*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, and *Clostridium* species.

[0172] Prokaryotic organisms with modification of a redox pathway can also be used in the methods of the invention to produce soluble, active eukaryotic glycosyltransferases or soluble, active therapeutic proteins. Modifications can be performed on prokaryotic organisms that have a reducing environment, *e.g.*, *E. coli* or other gram negative bacteria or some gram positive bacteria. The prokaryotic microorganisms are modified to promote an oxidizing intracellular environment, thereby enhancing intracellular disulfide bond formation and protein refolding of *e.g.*, therapeutic proteins and eukaryotic glycosyltransferases.

[0173] Many prokaryotic organisms use two pathways to reduce disulfide bonds that form in some cytoplasmic proteins, including recombinantly expressed proteins. The components of these pathways can be manipulated to promote formation of an intracellular oxidizing environment. The first pathway is the thioredoxin system, which generally includes a thioredoxin reductase and thioredoxin. Thioredoxin reductase maintains thioredoxin in a reduced state. The second pathway is the glutaredoxin system, which generally includes a glutathione oxidoreductase, glutathione, and glutaredoxins. Inactivating mutations of some components of these redox pathways can ultimately increase the formation of disulfide bonds in expressed proteins, and in the case of heterologous proteins expressed in the prokaryotic organism, can increase the solubility and activity of the expressed heterologous proteins. For example, in *E. coli* elimination of thioredoxin reductase activity results in an accumulation of oxidized thioredoxin that act as an oxidase in the intracellular compartment.

[0174] Some preferred examples are prokaryotic microorganisms that have reduced or absent reductase activity. For example, the activity of a thioredoxin reductase and/or a glutathione oxidoreductase can be reduced or eliminated to modify the intracellular environment, thereby producing an oxidizing intracellular environment that favors formation of disulfide bonds.

[0175] For example, *E. coli* strains that have mutations in both the thioredoxin reductase gene (*trxB*) and the glutathione oxidoreductase gene (*gor*) are able to express proteins with higher levels of disulfide bond formation. *See, e.g.*, Prinz *et al.*, *J. Biol Chem.* 272:15661-15667 (1997). These *trxB gor* double mutants grow very slowly on most growth media, although growth can be enhanced by addition of a reductant, such as DTT. However, the double mutant strains frequently give rise to suppressor mutant strains that retain the *trxB gor* mutations and that grow faster in medium lacking DTT. One example of a *trxB gor* suppressor mutation in *E. coli* is a mutation of the gene *ahpC*, which encodes a catalytic

subunit of the alkyl hydroperoxidase, AhpCF. This suppressor mutation adds a triplet to the DNA that encodes the catalytic site of the AhpCF enzyme. Fast growing double mutant *E. coli* strains, *e.g.*, *trxB*, *gor*, *supp* and *trxB*, *gshA*, *supp* strains are disclosed in *e.g.*, US Patent No. 6,872,563, which is herein incorporated by reference for all purposes. Such manipulated

5 *E. coli* strains, *e.g.*, *trxB*, *gor*, *supp* strains, are commercially available, *e.g.*, under the trade names ORIGAMI TM, ORIGAMI 2 TM, and ROSETTA-GAMI TM, from *e.g.*, EMD Biosciences, Inc. Other *E. coli* mutations can result in an oxidizing intracellular environment, *e.g.*, *trxB*, *gshA* and *trxB*, *gshA* *supp* strains.

[0176] Other manipulations of components of a redox pathway in a microorganism can be used to enhance formation of disulfide bonds in a protein, *e.g.*, a therapeutic protein and a heterologous glycosyltransferase. For example, proteins with oxidizing activity, *e.g.*, *E. coli* thioredoxin proteins in *trxB*, *gor* mutant strains, can be overexpressed in the prokaryotic microorganism. Another example is expression or overexpression of thioredoxin mutants that have enhanced oxidizing activity. Examples of such mutants are described in, *e.g.*,

10 Bessette, *et al.* *PNAS* 96:13703-13708 (1999). Targeted cytoplasmic expression of certain oxidizing enzymes can also be used to enhance formation of intracellular disulfide bonds. For example oxidizing proteins that are typically expressed in the periplasmic space, *e.g.*, DsbC, can be expressed in a bacterial cytoplasm by *e.g.*, deleting a periplasmic targeting sequence or including a cytoplasmic retention sequence. Other oxidizing periplasmic

15 proteins can be expressed in the bacterial cytoplasm to enhance oxidation of cytoplasmic proteins, *e.g.*, by deleting a periplasmic targeting sequence or including a cytoplasmic retention sequence.

[0177] Thioredoxin reductase nucleic acids, glutathione oxidoreductase nucleic acids, thioredoxin nucleic acids, glutathione nucleic acids, and nucleic acids encoding other proteins involved in maintenance of an intracellular redox environment can be identified in other

25 bacteria, *e.g.*, *Azotobacter sp.* (*e.g.*, *A. vinelandii*), *Pseudomonas sp.*, *Rhizobium sp.*, *Erwinia sp.*, *Escherichia sp.* (*e.g.*, *E. coli*), *Bacillus*, *Pseudomonas*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, *Paracoccus* and *Klebsiella sp.*, among many others. Such genes can be identified by sequence analysis and comparison to known thioredoxin reductase

30 genes, glutathione oxidoreductase genes, and genes encoding other proteins involved in maintenance of an intracellular redox environment or to the amino acid sequence of the encoded products. The encoded proteins can be further identified functionally by enzymatic assays or by genetic complementation assays of *E. coli* mutants of an appropriate gene

function. The endogenous thioredoxin reductase and glutathione oxidoreductase genes can be *e.g.*, mutated to inactivate the gene product using standard molecular biology techniques and those mutated strains can also be used to express proteins with increased levels of disulfide bond formation, as compared to unmutated strains.

5 *B. Identification of intracellular, oxidizing environments*

[0178] Protein refolding and protein activity frequently depend on the correct formation of disulfide bonds. Disulfide bonds are reversible thiol-disulfide (SH-SS) exchange reactions that are greatly influenced by the redox state of the environment surrounding the protein. In many cells, including *E. coli* and other prokaryotic organisms, glutathione, a tripeptide
10 containing cysteine, is an important thiol-disulfide redox buffer. The redox state of prokaryotic microorganisms is also affected by other proteins, such as thioredoxins. Reductase proteins, in turn, regulate the redox state of glutathione, glutaredoxins and thioredoxins. In *E. coli* glutathiones, encoded by gshA and gshB, regulates the redox state of glutaredoxins. Reductase proteins include, *e.g.*, thioredoxin reductase and glutathione
15 oxidoreductase. *E. coli* has thioredoxins encoded by trxA and trxC genes, glutaredoxin 1, glutaredoxin 2, and glutaredoxin 3, encoded by grxA, grxB, and grxC genes. Many of the proteins that regulate the oxidation state of a cell, *e.g.*, thioredoxin, glutathione, thioredoxin reductase and glutathione oxidoreductase, comprise an active site CX₁X₂C motif. The proteins also comprise a protein structural motif known as the thioredoxin fold.

20 [0179] One method to identify prokaryotes that have an oxidizing intracellular environment is to measure the ratio of reduced glutathione (GSH) to oxidized glutathione (GSSG). Optimum ratios of GSH/GSSG for protein folding have been determined. In vitro, maximum yields of properly folded protein occur at GSH/GSSG ratios of less than 50, preferably less than 40, more preferably less than 30, still more preferably less than 20, and most preferably
25 less than 10. In mammalian cells, cytoplasmic GSH/GSSG ratios ranged from 30/1 to 100/1, while secretory pathway (where most protein refolding occurs) GSH/GSSG ratios ranged from 1/1 to 3/1. Hwang *et al.*, *Science* 257:1496-1502 (1992). *E. coli* express very few intracellular proteins with disulfide bonds. *E. coli* proteins that have disulfide bonds are secreted into the periplasmic space, which has an oxidizing environment. Typical wild type
30 intracellular *E. coli* GSH/GSSG ratios ranged from 50/1 to 200/1. Hwang *et al. supra*.

[0180] The methods of the invention can be used to produce soluble eukaryotic glycosyltransferases in prokaryotic organisms that have an oxidizing intracellular

environment. Microorganisms with an oxidizing intracellular environment typically have GSH/GSSG ratios of less than 50, preferably less than 40, more preferably less than 30, still more preferably less than 20, and most preferably less than 10. Thus, in some embodiments, the microorganisms of the invention will have GSH/GSSG ratios that range, *e.g.*, from 0 to 50, or from 0.1 to 25, or from 0.5 to 10.

[0181] Prokaryotic organisms with intracellular environments can be identified by *e.g.*, determining the intracellular GSH/GSSG ratio of the prokaryotic organisms. Assays for total glutathione concentration are commercially available from, *e.g.*, Sigma. Assays for determination of a GSH/GSSG ratio are described, *e.g.*, in Hwang *et al.*, *Science* 257:1496-1502 (1992). Methods to quantify intracellular content of GSH and GSSG by derivitization with *N*-(1-pyrenyl)maleimide (NPM) followed by quantification using HPLC are described in Ostergaard, *et al.*, *J. Cell Biol.* 166:337-345 (2004).

[0182] A number of additional assays are available to those of skill to determine whether a prokaryotic organism has an intracellular, oxidizing environment. Those assays include measurement of glutathione reductase activity and glutathione pool redox state (Tuggle and Fuchs, *J. Bacter.* 162:448-450 (1985)), sensitivity to thiol-specific oxidants in growth medium (Prinz *et al.*, *J. Biol. Chem.* 272:15661-15667 (1997)), transcriptional activation of the OxyR gene in *E. coli* after exposure to hydrogen peroxide or diamide (Bessette *et al.*, *PNAS* 96:13703-13708 (1999)), measurement of the redox state of a reporter gene, such as a redox sensitive green fluorescent protein, (rxYFP) (Ostergaard *et al.*, *J. Cell Biol.* 166:337-345 (2004)), detection of glutathione using glutathione sensitive dyes such as monochlorobimane, CellTracker Green CMFDA, *o*-phthaldialdehyde, and naphthalene-2,3-dicaboxaldehyde from *e.g.*, Molecular Probes, and oxidation of cysteine residue in proteins after exposure of cells to a sulfhydryl-alkylating reagent, such as 4-acetamido-4'-maleimidystibene-2,2-disulfonic acid (Jurado *et al.*, *J. Mol. Biol.* 320:1-10 (2002)).

VII. Enhancement of protein solubilization

[0183] Reduction of disulfide bonds in heterologously expressed proteins, such as the therapeutic proteins and eukaryotic glycosyltransferase polypeptides used in the methods of the invention, frequently results in protein misfolding and precipitation out of solution. In bacterial cells such as *e.g.*, *E. coli*, misfolded proteins are expressed as insoluble inclusion bodies. Enzymes with correct disulfide bond formation and thus, proper folding are usually soluble. Solubilization of a protein is generally indicated by the presence of the protein in an

aqueous fraction after centrifugation at an appropriate speed for an appropriate period. In addition, expression of properly folded proteins results in increased levels of protein activity. Thus, assays of enzyme activity can also be used to determine whether proper protein folding has occurred.

5 [0184] Solubilization of *e.g.*, an O-glycosylated therapeutic protein and glycosyltransferase expressed in a microorganism with an oxidizing environment can be compared to solubilization of a therapeutic polypeptide and glycosyltransferase expressed in a microorganism with a reducing environment, *e.g.*, an *E. coli* strain with a reducing environment. In some embodiments, an O-glycosylated therapeutic protein expressed in a
10 microorganism with an oxidizing environment is expressed in a soluble fraction at levels that are up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than soluble levels of the same therapeutic protein expressed in a microorganism with a reducing environment. In other embodiments, an O-glycosylated therapeutic protein expressed in a microorganism with an oxidizing environment has activity levels, *e.g.*, U/cell
15 or U/mg protein, up to 1.1, 1.2, 1.5, 2, 3, 5, 10, 15, 20, 50, 100, 500, 1000, or up to 10,000 times greater than activity levels of the same therapeutic protein expressed in a microorganism with a reducing environment.

A. Characterization of protein solubility

[0185] In preferred embodiments, the therapeutic proteins and eukaryotic
20 glycosyltransferases are expressed as soluble proteins intracellularly within a prokaryotic microorganism. Solubility of the therapeutic proteins and eukaryotic glycosyltransferase polypeptides can be determined as disclosed above, by determining protein levels in an aqueous fraction after centrifugation at an appropriate speed for an appropriate period. Protein levels can be determined using methods known to those of skill in the art, *e.g.*,
25 immunoassays or direct comparison of proteins separated by, *e.g.*, SDS-PAGE. Immunoassays can be performed using antibodies specific for the therapeutic protein or eukaryotic glycosyltransferase polypeptide of interest or using antibodies specific for an epitope or purification tag that is covalently linked to the therapeutic protein or the eukaryotic glycosyltransferase polypeptide.

30 [0186] Solubility can also be determined by assaying an activity of the O-glycosylated therapeutic protein in a soluble fraction from a prokaryotic microorganism. In a preferred

embodiment, O-glycosylated therapeutic protein activity is measurable in a soluble intracellular fraction from a prokaryotic microorganism.

B. Enhancement of protein solubility

[0187] Further enhancement of solubility of therapeutic proteins and heterologous glycosyltransferase polypeptides can occur, *e.g.*, by reducing the rate of protein expression or by expressing the protein in combination with, *e.g.* a chaperone protein.

[0188] Enhancing the rate of formation of appropriate disulfide bonds can lead to higher expression of active soluble glycosyltransferases and O-glycosylated therapeutic proteins. Another method to enhance expression of active soluble glycosyltransferases and O-glycosylated therapeutic proteins is to reduce the rate of expression thereby allowing the nascent polypeptide more time to achieve a stable, soluble conformation. The combination of the two methods, as described herein, is a preferred embodiment of the invention. Maximal expression of a heterologous protein generally occurs under optimal growth condition for the host cells. One method to slow the expression of proteins is to slow the growth rate of the cells. In a preferred embodiment, host cells are grown at a temperature below their optimal growth temperature. For example, the optimal growth temperature of *E. coli* is 37°C. Therefore, a temperature less than optimal growth temperature for *E. coli* is less than 37°C, *e.g.*, between 4°C and 36°C, between 8°C and 33°C, between 12°C and 30°C, or between 18°C and 26°C, or at about 20°C, or at about 24°C.

[0189] The temperature used to slow protein production will depend on the optimal growth temperature of the host cells. As an example, *E. coli* and many other bacteria have an optimal growth temperature of 37°C. Thus, a temperature lower than an optimal growth temperature for *E. coli* or for other bacteria that grow optimally at 37°C could be between 4-35°C, between 12-30°C, or between 15-20°C. In a preferred embodiment the temperature lower than an optimal growth temperature for *E. coli* or for other bacteria that grow optimally at 37°C is between 18 and 23°C. For cells that grow optimally at 30°C, as do many yeasts, a temperature lower than an optimal growth temperature could be between 10 and 25°C, between 12 and 21°C, or between 15 and 20°C.

[0190] Another method to reduce the rate of expression of a heterologous protein is to vary the concentration of a molecule that regulates expression from an inducible promoter. For example, some *lacY* mutations allow protein expression to be controlled by varying the

amount of IPTG, the inducer molecule, in the medium. In preferred embodiments the concentration of IPTG in the medium is less than optimal for, e.g., expression of a protein that does not form inclusion bodies when over expressed in a prokaryotic microorganism.

[0191] In some embodiments, an O-glycosylated therapeutic protein is expressed in a microorganism that has an oxidizing environment and that further comprises a heterologous chaperone protein. Chaperone proteins include, e.g., trigger factor; members of the Hsp70 chaperone family, e.g. DnaK; members of the Hsp100 chaperone family, e.g. ClpB, and members of the Hsp60 chaperone family, e.g. GroEL. See, e.g., Sorensen and Mortensen, *BioMed Central*, www.microbialcellfactories.com/content/4/1/1. Chaperones are also known that allow protein folding at 4°C, e.g., Cpn60 and Cpn 10 from *Oleispira antartica* RB8^T. See, e.g., *Id.* and Ferrer *et al.*, *Nat. Biotechnol.* 21:1266-1267 (2003). Exemplary chaperonin proteins include, those listed in U.S. Provisional Application No. 60/665,396, filed March 24, 2005; U.S. Provisional Application No. 60/668,899, filed April 5, 2005; U.S. Provisional Application No. 60/732,409, filed October 31, 2005; and International Application PCT/US06/11065, filed March 24, 2006; each of which are herein incorporated by reference for all purposes.

[0192] Chaperone proteins can be specific for either the therapeutic protein or for a glycosyltransferase used in the methods of the invention. For example, a chaperone protein for a mammalian Core 1 Gal T1 protein has been described and can be used herein to enhance production and in vivo activity of a mammalian Core 1 Gal T1. See, e.g., Ju and Cummings, *Proc. Nat'l. Acad. Sci. USA* 99:16613-16618 (2002); Accession number AA578739 and Accession number NP_067525; each of which are herein incorporated by reference for all purposes.

[0193] In other embodiments, a therapeutic protein and a eukaryotic glycosyltransferase polypeptide are expressed in a microorganism that has an oxidizing environment that further comprises a heterologous protein disulfide isomerase (PDI). PDI proteins can make or shuffle disulfide bonds. PDI proteins are described e.g., in Georgiou *et al.* U.S. Patent No. 6,027,888, which is herein incorporated by reference for all purposes. PDI proteins include e.g., rat liver PDI, Ero1p and Pdi1p proteins from *Saccharomyces*. Prokaryotic proteins include e.g., DsbC from *E. coli*. See, e.g., Frand *et al.*, *Trends in Cell Biol.* 10:203-210 (2000). In some embodiments, DsbC proteins are expressed in a bacterial cytoplasm by e.g., deleting a periplasmic targeting sequence or including a cytoplasmic retention sequence.

Thus, DsbC proteins include truncations or other variants that exhibit disulfide isomerase activity, either intracellularly or in the periplasmic space.

[0194] Other prokaryotic proteins that act to maintain the redox state of protein disulfide bonds include, *e.g.*, DsbB, DsbA, DsbC, DsbD, and DsbG from *E. coli*. These proteins are well known in the art and are described in, *e.g.*, Beckwith *et al.* U.S. Patent No. 6,872,563, which is herein incorporated by reference for all purposes. In some embodiments, DsbB, DsbA, DsbC, DsbD, and DsbG are expressed in the bacterial cytoplasm to enhance oxidation of cytoplasmic proteins, *e.g.*, by deleting a periplasmic targeting sequence or including a cytoplasmic retention sequence.

[0195] In a further embodiment, a therapeutic protein and a eukaryotic glycosyltransferase polypeptide are expressed in a prokaryotic microorganism that has an oxidizing environment and that also comprises a heterologous chaperone protein and/or a heterologous PDI protein and or a protein such as DsbB, DsbA, DsbC, DsbD, and DsbG from *E. coli*.

VIII. Accessory enzymes and enhanced intracellular synthesis of donor substrates

[0196] Suitable donor substrates used by the heterologous soluble, active glycosyltransferases and methods of the invention include, but are not limited to, UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, GDP-Man, GDP-Fuc, UDP-GlcUA, UDP-GlcNH₂, UDP-GalNH₂, and CMP-sialic acid. Guo *et al.*, *Applied Biochem. and Biotech.* **68**: 1-20 (1997). As the reactions catalyzed by the heterologous soluble, active glycosyltransferases typically take place within the cell, the present invention also provides methods to enhance the production of donor substrates, *e.g.*, nucleotide sugars, used by the heterologous soluble, active glycosyltransferases.

[0197] Intracellular synthesis of donor substrates can be enhanced by expression of an "accessory enzyme." An accessory enzyme is an enzyme that is involved in catalyzing a reaction that, for example, forms a substrate or other reactant for a glycosyltransferase reaction. An accessory enzyme can, *e.g.*, catalyze the formation of a nucleotide sugar that is used as a sugar donor moiety by a glycosyltransferase. An accessory enzyme can also be one that is used in the generation of a nucleotide triphosphate that is required for formation of a nucleotide sugar, or in the generation of the sugar which is incorporated into the nucleotide sugar. The heterologous glycosyltransferases can be used in combination with an appropriate recombinant accessory enzymes, which may or may not be fused to the glycosyltransferase.

Examples of heterologous accessory enzyme and glycosyltransferase pairs include, *e.g.*, a galactosyltransferase and an epimerase, *e.g.*, a UDP-glucose 4' epimerase protein, a UDP-GlcNAc 4' epimerase protein or a dual function UDP-glucose 4' epimerase protein/ UDP-GlcNAc 4' epimerase protein; or a sialyltransferase and a sialic acid synthetase. Other
5 examples are disclosed in PCT Patent Application PCT/CA98/01180, published as WO99/31224 on June 24, 1999; which is herein incorporated by reference for all purposes.

[0198] Sugar nucleotides serve as donors for the *in vivo* glycosylation of a therapeutic protein. Production of sugar nucleotides can be enhanced to increase intracellular sugar nucleotide pools in a variety of ways. One method to increase production of any sugar
10 nucleotide is to increase expression of the biosynthetic enzymes that make the sugar nucleotide. Expression of single enzymes, multiple enzymes or an entire biosynthetic pathway, *e.g.*, an operon, can be increased. For example the appropriate nucleic acids can be cloned in endogenous operons with endogenous regulatory sequences or can be cloned into synthetic expression cassettes, including polycistronic expression cassettes. Enzymatic
15 syntheses of sugar nucleotides are reviewed in Bulter and Elling, *Glycoconj J* 16:147-159 (1999); some non-limiting *in vivo* examples follow.

[0199] For example, for synthesis of UDP-GlcNAc in *E. coli*, the gene products GlmS, GlmM, and GlmU are required to convert Fructose-6P to UDP-GlcNAc. *glmU* and *glmS* are adjacent in the genome, and could be cloned out as a single fragment including regulatory
20 structures, and included in an expression construct or reintroduced into the bacterial genome. The expression construct could be reengineered to also include the *glmM* gene, thus, adding another pathway gene to a natural operon.

[0200] Also to increase UDP-GlcNAc production, expression of the acetyl-CoA synthase (*acs*) gene can be increased. *See, e.g.*, Lin *et al.*, *Appl Micro Biotech* 71:870-874 (2006).
25 Increased expression of *acs* boosts the pool of CoA available the acetylation of GlcN-1P by GlmU.

[0201] In addition to endogenous genes from the host microorganism, orthologs can be expressed to increase UDP-GlcNAc production. For a list of genes and pathways, *see, e.g.*, Milewski *et al.* *Yeast* 23:1-14(2006). In particular yeast gene products use different
30 intermediates than the bacterial pathway. For example, yeast acetylate GlcN-6P, then a phospho-GlcNAc mutase converts it to GlcNAc-1P; whereas bacteria first convert Glc6P to Glc1P, and then acetylate it to form GlcNAc-1P.

[0202] For synthesis of UDP-Gal, most of the *E. coli* biosynthetic genes occur in a single operon that can be cloned into an expression construct or a host cell genome, in whole or in part. Bacteria typically generate UDP-Gal using two pathways: conversion of UDP-Glc to UDP-Gal by GalE, or from galactose, *i.e.*, Gal $\xrightarrow{\text{galK}}$ Gal-1P $\xrightarrow{\text{galT}}$ UDP-Glc $\rightarrow$

5 UDP-Gal + Glc1P. For the second pathway, most of the *E. coli* biosynthetic genes occur in a single operon that can be cloned into an expression construct or a host cell genome, in whole or in part.

[0203] An alternative method to synthesize UDP-Gal is by addition and expression of a sucrose synthase gene and feeding of sucrose. *See, e.g.*, Elling *et al.* *Glycobiology* 3:349-355(1993) and Wang *et al.*, *Arch Biochem Biophys* 371:70-82 (1999). Most *E. coli* K12 strains do not metabolize sucrose. Thus, the reaction is sucrose + UDP $\xrightarrow{\text{Sucrose synthase}}$ UDP-Glc + fructose. An advantage of this approach is that the fructose can be used in the UDP-GlcNAc pathway. Note that while some sucrose enters the cell through other sugar import pathways, addition of natural sucrose transporters to the strain improves import. *See, e.g.*

15 Jahreis *et al.* *J. Bacteriol.* 184, 5307-5316 (2002). The sucrose transport protein can be manipulated to remove hydrolase activity. If sucrose is phosphorylated on import into the cell, a sucrose-phosphate synthase could also be used; a phosphor-mutase enzyme may be needed to convert sucrose-6P $\leftrightarrow$ sucrose-1P). *See, e.g.* Lunn, *Gene* 303:187-196 (2003); and Huynh *et al.* *Acta Crystallogr B* 61:116-117 (2005).

20 [0204] A number of methods can be used to increase production of UDP-GalNAc. For example, UDP-GlcNAc can be converted to UDP-GalNAc using a dual specificity Glc/GlcNAc 4' epimerase, *e.g.*, GNE or GalE. Some bacterial galactose epimerase proteins are not able to convert UDP-Glc. UDP-GlcNAc biosynthetic genes including *glmS*, *glmM*, and *glmU* can also be overexpressed to increase production of UDP-GalNAc.

25 [0205] Mammalian proteins can synthesize UDP-GalNAc directly from GalNAc, UTP, and ATP. *See, e.g.*, Bourgeaux *et al.* *Bioorg Med Chem Lett* 15:5459-5462 (2005), using mammalian GalNAc kinase and UDP-GalNAc pyrophosphorylase (AGX1). A bacterial GlcNAc pyrophosphorylase (*GlmU*) can also accomplish this last step. *See, e.g.* Szumilo *et al.* *JBC* 271:13147-13154(1996).

30 [0206] Some of the sugar kinases make the 6P sugar, not a 1P sugar. In these cases, a phosphor-sugar mutase would be required to transfer the phosphate from the 6 to the 1 position. An example is the endogenous bacterial *GlmM*, which converts GlcN-6P to GlcN-

1P in the UDP-GlcNAc biosynthetic pathway. Addition of a sugar-specific mutase activity that is otherwise lacking in the bacterium may enable certain UDP-sugar synthetic pathways to act on now-available precursors. For example, yeast AGM1, converts GlcNAc-6P to 1P, which could then be converted to UDP-GlcNAc by endogenous bacterial glmU.

5 [0207] For other nucleotide sugars, synthetic genes from other species could be added, for example, from higher eukaryotes or pathogenic bacteria, or specific pathways altered to enhance specific sugar nucleotide production. For example, overexpression of bacterial rcsA and deletion of bacterial wcaJ significantly enhances the production of GDP-fucose, *See e.g., Dumon et al. Glycoconj J* 18, 465-474 (2001). CMP-sialic acid (CMP-NeuAc) can be
10 produced by deleting bacterial nanA, and introducing a CMP-NeuAc synthase (CNS) from, *e.g., Neisseria meningitidis, see e.g. Priem et al. Glycobiology* 12, 235-240 (2002), and Gilbert *et al. Biotechnol. Lett.* 19, 417-420.

[0208] Synthesis of donor substrates can also be enhanced by growth of host prokaryotic microorganisms in medium that includes a donor substrate or a precursor to a donor substrate,
15 *e.g., galactose or GalNAc. See, e.g., Priem et al., Glycobiology* 12:235-240 (2002). The microorganism can be altered to facilitate uptake of precursor molecule from the medium, *e.g., by expressing a sugar transport protein. For example, if lactose is used, E. coli cells that express the LacY permease can be used.*

[0209] Synthesis of donor substrates can be enhanced by inhibiting or knocking out
20 pathways that breakdown sugar nucleotide precursors. For example, in *E. coli*, break down of sialic acid can be minimized by using a host strain that lacks aldolase activity, *see, e.g., Priem et al. Glycobiology* 12, 235-240 (2002). Also in *E. coli*, when lactose is, for example, part of an oligosaccharide on the glycosylated therapeutic protein, lactose breakdown can be minimized by using host cells that are *LacZ*⁻. Other examples of enzymes that breakdown
25 sugar precursors and are thus, targets for inhibition or deletion include RfbA and RffH which interfere with UDP-Glc production by siphoning Glc-1P to TDP-Glc. Glc-1P pools decrease after dephosphorylation by agp, thus agp could be targeted. Similarly, manA could be targeted as it interferes with production of UDP-GlcNAc by isomerizing Fruc-6P to man-6P.

[0210] Synthesis of donor substrates can also be enhanced by inhibiting or knocking out
30 pathways that breakdown the sugar nucleotide itself. For example, targets include ugd which oxidizes UDP-Glc; glf which converts UDP-Gal to UDP-galactofuranose; and RffE which converts UDP-GlcNAc to UDP-ManNAc.

[0211] Additional methods to enhance production of a donor substrate are disclosed in International Application No. PCT/US05/033532, filed September 19, 2005; which is herein incorporated by reference for all purposes.

IX. Conjugation of modified sugar residues

5 [0212] The modified sugars are conjugated to a glycosylated or non-glycosylated peptide or protein using an appropriate enzyme to mediate the conjugation. Preferably, the concentrations of the modified donor sugar(s), enzyme(s) and acceptor peptide(s) or protein(s) are selected such that glycosylation proceeds until the acceptor is consumed. The considerations discussed below, while set forth in the context of a sialyltransferase, are
10 generally applicable to other glycosyltransferase reactions.

[0213] The present invention also provides for the industrial-scale production of modified peptides. As used herein, an industrial scale generally produces *e.g.*, at least one microgram, one milligram, or one gram of finished, purified conjugate.

[0214] In the discussion that follows, the invention is exemplified by the conjugation of
15 modified sialic acid moieties to a glycosylated peptide. The exemplary modified sialic acid is labeled with PEG. The focus of the following discussion on the use of PEG-modified sialic acid and glycosylated peptides is for clarity of illustration and is not intended to imply that the invention is limited to the conjugation of these two partners. One of skill understands that the discussion is generally applicable to the additions of modified glycosyl moieties other
20 than sialic acid. Moreover, the discussion is equally applicable to the modification of a glycosyl unit with agents other than PEG including other water-soluble polymers, therapeutic moieties, and biomolecules.

[0215] An enzymatic approach can be used for the selective introduction of PEGylated or PPGylated carbohydrates onto a peptide or glycopeptide. The method utilizes modified
25 sugars containing PEG, PPG, or a masked reactive functional group, and is combined with the appropriate glycosyltransferase. By selecting the glycosyltransferase that will make the desired carbohydrate linkage and utilizing the modified sugar as the donor substrate, the PEG or PPG can be introduced directly onto the peptide backbone, onto existing sugar residues of a glycopeptide or onto sugar residues that have been added to a peptide.

30 [0216] An acceptor for the sialyltransferase is present on the peptide to be modified by the methods of the present invention either as a naturally occurring structure or one placed there

recombinantly, enzymatically or chemically. Suitable acceptors, include, for example, galactosyl acceptors such as Gal β 1,4GlcNAc, Gal β 1,4GalNAc, Gal β 1,3GalNAc, lacto-N-tetraose, Gal β 1,3GlcNAc, Gal β 1,3Ara, Gal β 1,6GlcNAc, Gal β 1,4Glc (lactose), GalNAc and sialic acid containing structures, and other acceptors known to those of skill in the art (*see*,
5 *e.g.*, Paulson *et al.*, *J. Biol. Chem.* 253: 5617-5624 (1978)).

[0217] In one embodiment, an acceptor for the sialyltransferase is present on the glycopeptide to be modified upon *in vivo* synthesis of the glycopeptide. Such glycopeptides can be sialylated using the claimed methods without prior modification of the glycosylation pattern of the glycopeptide. Alternatively, the methods of the invention can be used to
10 sialylate a peptide that does not include a suitable acceptor; one first modifies the peptide to include an acceptor by methods known to those of skill in the art. In an exemplary embodiment, a GalNAc residue is added by the action of a GalNAc transferase.

[0218] In an exemplary embodiment, the galactosyl acceptor is assembled by attaching a galactose residue to an appropriate acceptor linked to the peptide, *e.g.*, a GalNAc. The
15 method includes incubating the peptide to be modified with a reaction mixture that contains a suitable amount of a galactosyltransferase (*e.g.*, gal β 1,3 or gal β 1,4), and a suitable galactosyl donor (*e.g.*, UDP-galactose). The reaction is allowed to proceed substantially to completion or, alternatively, the reaction is terminated when a preselected amount of the galactose residue is added. Other methods of assembling a selected saccharide acceptor will be
20 apparent to those of skill in the art.

[0219] Methods for conjugation of modified sugars to peptides or proteins are found *e.g.*, in USSN 60/328,523 filed October 10, 2001; USSN 60/387,292, filed June 7, 2002; USSN 60/391,777 filed June 25, 2002; USSN 60/404,249 filed August 16, 2002; and PCT/US02/32263; each of which are herein incorporated by reference for all purposes.

25 X. Expression of proteins in prokaryotic host cells

[0220] Soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides of the invention can be expressed in a variety of prokaryotic microorganisms with oxidizing intracellular environments, including *E. coli*, and other bacterial hosts, as described above.

30 [0221] Once expressed in a prokaryotic organism that has an oxidizing intracellular environment, the soluble, active O-glycosylated therapeutic proteins and glycosyltransferase

polypeptides can be isolated using standard protein purification techniques and used therapeutically or further modified *in vitro*.

[0222] Typically, the polynucleotide that encodes the heterologous polypeptide is placed under the control of a promoter that is functional in the desired prokaryotic organisms that has an oxidizing environment. An extremely wide variety of promoters are well known, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include one or more of these control sequences are termed "expression cassettes." Accordingly, the invention provides expression cassettes into which the nucleic acids that encode fusion proteins are incorporated for high level expression in a desired microorganism that has an oxidizing environment.

[0223] Examples of expression vectors include, e.g., the pCWin1 vector and pCWin2 vector, both disclosed in WO 2005/067601, which is herein incorporated by reference for all purposes.

[0224] Expression control sequences that are suitable for use in a particular host cell are often obtained by cloning a gene that is expressed in that cell. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (*lac*) promoter systems (Change *et al.*, *Nature* (1977) 198: 1056), the tryptophan (*trp*) promoter system (Goeddel *et al.*, *Nucleic Acids Res.* (1980) 8: 4057), the *tac* promoter (DeBoer, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); and the lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake *et al.*, *Nature* (1981) 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes can be used.

[0225] For expression of soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides in prokaryotic cells other than *E. coli*, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, the hybrid *trp-lac* promoter functions in *Bacillus* in addition to *E. coli*.

Promoters are known for other bacterial species, *e.g. Pseudomonas*. See, *e.g.*, U.S. Patent Application Publication No. US 2005/0186666, published August 25, 2005, which is herein incorporated by reference for all purposes.

[0226] A ribosome binding site (RBS) is conveniently included in the expression cassettes of the invention. An RBS in *E. coli*, for example, consists of a nucleotide sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine and Dalgarno, *Nature* (1975) 254: 34; Steitz, *In Biological regulation and development: Gene expression* (ed. R.F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

[0227] Either constitutive or regulated promoters can be used in the present invention.

Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the fusion proteins is induced. High level expression of heterologous proteins slows cell growth in some situations and may not be desired in all situations, see below. An inducible promoter is a promoter that directs expression of a gene where the level of expression is alterable by environmental or developmental factors such as, for example, temperature, pH, anaerobic or aerobic conditions, light, transcription factors and chemicals. Such promoters are referred to herein as "inducible" promoters, which allow one to control the timing of expression of the glycosyltransferase or enzyme involved in nucleotide sugar synthesis. For *E. coli* and other bacterial host cells, inducible promoters are known to those of skill in the art. These include, for example, the *lac* promoter, the bacteriophage lambda P_L promoter, the hybrid *trp-lac* promoter (Amann *et al.* (1983) *Gene* 25: 167; de Boer *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA* 80: 21), and the bacteriophage T7 promoter (Studier *et al.* (1986) *J. Mol. Biol.*; Tabor *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA* 82: 1074-8). These promoters and their use are discussed in Sambrook *et al.*, *supra*. A particularly preferred inducible promoter for expression in prokaryotes is a dual promoter that includes a *tac* promoter component linked to a promoter component obtained from a gene or genes that encode enzymes involved in galactose metabolism (*e.g.*, a promoter from a UDPgalactose 4-epimerase gene (*galE*)). The dual *tac-gal* promoter, which is described in PCT Patent Application Publ. No. WO98/20111.

[0228] Another inducible promoter is the *cspA* promoter, which is highly induced at low temperatures in *E. coli*. See, *e.g.*, Sorensen and Mortensen, *BioMed Central*, www.microbialcellfactories.com/content/4/1/1 and Mujacic *et al. Gene* 238:325-3332 (1999).

[0229] A construct that includes a polynucleotide of interest operably linked to gene expression control signals that, when placed in an appropriate host cell, drive expression of the polynucleotide is termed an "expression cassette." Expression cassettes that encode the fusion proteins of the invention are often placed in expression vectors for introduction into the host cell. The vectors typically include, in addition to an expression cassette, a nucleic acid sequence that enables the vector to replicate independently in one or more selected host cells. Generally, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria. For instance, the origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria. Alternatively, the vector can replicate by becoming integrated into the host cell genomic complement and being replicated as the cell undergoes DNA replication. A preferred expression vector for expression of the enzymes in bacterial cells is pTGK, which includes a dual *tac-gal* promoter and is described in PCT Patent Application Publ. NO. WO98/20111. Another useful cloning vector is pCWin2-MBP or a version of pCWin2 with a modified 5' UTR. *See, e.g.*, PCT/US05/00302, filed January 6, 2005, which is herein incorporated by reference for all purposes.

[0230] The construction of polynucleotide constructs generally requires the use of vectors able to replicate in bacteria. A plethora of kits are commercially available for the purification of plasmids from bacteria (*see*, for example, EasyPrepJ, FlexiPrepJ, both from Pharmacia Biotech; StrataCleanJ, from Stratagene; and, QIAexpress Expression System, Qiagen). The isolated and purified plasmids can then be further manipulated to produce other plasmids, and used to transfect cells. Cloning in *e.g.*, *E. coli*, *Streptomyces* or *Bacillus* is possible.

[0231] Selectable markers are often incorporated into the expression vectors used to express the polynucleotides of the invention. These genes can encode a gene product, such as a protein, necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, such as ampicillin, neomycin, kanamycin, chloramphenicol, or tetracycline. Alternatively, selectable markers may encode proteins that complement auxotrophic deficiencies or supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli*. Often, the vector will have one selectable marker that is functional in, *e.g.*, *E. coli*, or other cells in which the vector is

replicated prior to being introduced into the host cell. A number of selectable markers are known to those of skill in the art and are described for instance in Sambrook *et al.*, *supra*. An auxotrophic expression system is known for *Pseudomonas* species. See, e.g., U.S. Patent Application Publication No. US 2005/0186666, published August 25, 2005, which is herein
5 incorporated by reference for all purposes.

[0232] Construction of suitable vectors containing one or more of the above listed components employs standard ligation techniques as described in the references cited above. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required. To confirm correct sequences in plasmids constructed, the
10 plasmids can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequence analysis according to known methods. Molecular cloning techniques to achieve these ends are known in the art. A wide variety of cloning and *in vitro* amplification methods suitable for the construction of recombinant nucleic acids are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct
15 persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Volume 152, Academic Press, Inc., San Diego, CA (Berger); and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement) (Ausubel).

[0233] A variety of common vectors suitable for use as starting materials for constructing the expression vectors of the invention are well known in the art. For cloning in bacteria, common vectors include pBR322 derived vectors such as pBLUESCRIPT™, and λ -phage derived vectors.

[0234] The methods for introducing the expression vectors into a chosen prokaryotic
25 microorganism are not particularly critical, and such methods are known to those of skill in the art. For example, the expression vectors can be introduced into prokaryotic cells, including *E. coli*, by calcium chloride transformation, and into eukaryotic cells by calcium phosphate treatment or electroporation. Other transformation methods are also suitable.

[0235] Translational coupling may be used to enhance expression. The strategy uses a
30 short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the

termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. See Squires, et. al. (1988), *J. Biol. Chem.* 263: 16297-16302.

5 [0236] The soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides are preferably expressed intracellularly. Intracellular expression often results in surprisingly high yields. In another embodiment, the soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides are fused to a subsequence of protein A, a maltose binding protein, a starch binding protein, or bovine serum albumin
10 (BSA), for example, to facilitate purification, secretion, or stability.

[0237] The soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides of the invention can also be further linked to other bacterial proteins. This approach often results in high yields, because normal prokaryotic control sequences direct transcription and translation. In *E. coli*, *lacZ* fusions are often used to express heterologous
15 proteins. Other examples are discussed below. Suitable vectors are readily available, such as the pUR, pEX, and pMR100 series (*see, e.g.,* Sambrook *et al., supra.*). For certain applications, it may be desirable to cleave the non-glycosyltransferase amino acids from the fusion protein after purification. This can be accomplished by any of several methods known in the art, including cleavage by cyanogen bromide, a protease, or by Factor X_a (*see, e.g.,*
20 Sambrook *et al., supra.*; Itakura *et al., Science* (1977) 198: 1056; Goeddel *et al., Proc. Natl. Acad. Sci. USA* (1979) 76: 106; Nagai *et al., Nature* (1984) 309: 810; Sung *et al., Proc. Natl. Acad. Sci. USA* (1986) 83: 561). Cleavage sites can be engineered into the gene for the fusion protein at the desired point of cleavage.

[0238] More than one recombinant protein may be expressed in a single host cell by
25 placing multiple transcriptional cassettes in a single expression vector, or by utilizing different selectable markers for each of the expression vectors which are employed in the cloning strategy. For example, multiple proteins can be expressed in a single cell, *e.g.,* accessory enzymes, therapeutic proteins and glycosyltransferases that direct O-linked glycosylation.

30 [0239] A suitable system for obtaining recombinant proteins from *E. coli* which maintains the integrity of their N-termini has been described by Miller *et al. Biotechnology* 7:698-704 (1989). In this system, the gene of interest is produced as a C-terminal fusion to the first 76

residues of the yeast ubiquitin gene containing a peptidase cleavage site. Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal residue.

[0240] In some embodiments, multiple heterologous proteins are expressed in a single cell.

5 Many methods are available to those of skill to accomplish this expression. Some non-limiting examples follow.

[0241] In some embodiments, multiple heterologous proteins are expressed from a polycistronic expression construct. A polycistronic expression cassette has the following general form [Promoter] { [enhancer/ribosome binding site (RBS)] [open reading frame] }X
10 n [terminator]; where 'n' is the number of open reading frames in the polycistronic expression cassette and can be varied depending on the needs of the user. The [enhancer/RBS] element can also be varied, *e.g.*, either an extended enhancer/RBS combination or a simple RBS. Other elements that can vary within the polycistronic expression cassette include 5' UTRs.

[0242] The cellular location of any expression cassette, *e.g.*, independent promoter or
15 polycistronic expression cassette, can be varied. For example, expression cassettes can be either carried on one or multiple plasmids, or integrated into the genome (either directed or random integration, with a selection marker or regenerated into a stable integrant. *See, e.g.*, Link *et al.* *J. Bact.* 179:6228-6237 (1997); Muyrers *et al.*, *TIBS* 26:325-331 (2001); and Court *et al.*, *Annu. Rev. Genet.* 36:361-388 (2002). Integrated expression cassettes can be
20 regulated by constructed promoters (*e.g.* lac, pho, tac, T7, etc, taken from a plasmid), or by integrating a polycistronic array behind an endogenous (natural) genomic promoter. Genomic promoters can be regulated (*e.g.* integrate into the galE locus, and use the endogenous gal promoter) or constitutive (integrate behind a constitutive promoter; or screen for random integrants that express the enzymes). In one embodiment, all heterologous
25 glycosyltransferase, accessory enzyme, and therapeutic protein nucleic acid are expressed from plasmid constructs. In another embodiment, either all heterologous glycosyltransferase and accessory enzyme nucleic acid or a therapeutic protein nucleic acid are integrated into a bacterial genome and the other is expressed from a plasmid construct. In yet another embodiment, all heterologous glycosyltransferase, accessory enzyme, and therapeutic protein
30 nucleic acid are integrated into the bacterial genome.

[0243] Expression of heterologous proteins can be regulated in a temporal fashion, using *e.g.*, different inducible promoters to turn expression of specific genes on and off as desired

by the user. For example, expression of the nucleic acid could be induced prior to the production of the glycosyltransferases. Alternately, the cell could be prepared with prior induction of glycosyltransferases followed by the expression of the therapeutic protein.

XI. Purification and *in vitro* modification of soluble O-glycosylated therapeutic proteins

[0244] The soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides of the present invention are preferably expressed as intracellular proteins and can be used in this form, in the methods of the present invention. For example, permeabilized cells or a crude cellular extract containing the expressed intracellular s soluble, active O-glycosylated therapeutic proteins and glycosyltransferase polypeptides can used in the methods of the present invention or to *e.g.*, assay the activity of the soluble, active O-glycosylated therapeutic proteins and/or glycosyltransferase polypeptides.

[0245] Alternatively, the soluble, active O-glycosylated therapeutic protein can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see, generally, R. Scopes, Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.*, Academic Press, Inc. N.Y. (1990)). Substantially pure compositions of at least about 70, 75, 80, 85, 90% homogeneity are preferred, and 92, 95, 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.*, as immunogens for antibody production.

[0246] To facilitate purification and expression of the soluble, active O-glycosylated therapeutic protein of the invention, the nucleic acids that encode the proteins can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available, *i.e.* a purification tag. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of fusion proteins having these epitopes are commercially available (*e.g.*, Invitrogen (Carlsbad CA) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells). Additional expression vectors suitable for attaching a tag to the glycosyltransferases of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (*e.g.*, FLAG" (Kodak, Rochester NY). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines are used, although one can use more or less

than six. Suitable metal chelate affinity ligands that can serve as the binding moiety for a polyhistidine tag include nitrilo-tri-acetic acid (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" In Genetic Engineering: Principles and Methods, J.K. Setlow, Ed., Plenum Press, NY; commercially available from Qiagen (Santa Clarita, CA)). Other purification or epitope tags include, *e.g.*, AU1, AU5, DDDDK (EC5), E tag, E2 tag, Glu-Glu, a 6 residue peptide, EYMPME, derived from the Polyoma middle T protein, HA, HSV, IRS, KT3, S tag, S1 tag, T7 tag, V5 tag, VSV-G, β -galactosidase, Gal4, green fluorescent protein (GFP), luciferase, protein C, protein A, cellulose binding protein, GST (glutathione S-transferase), a step-tag, Nus-S, PPI-ases, Pfg 27, calmodulin binding protein, dsb A and fragments thereof, and granzyme B. Epitope peptides and antibodies that bind specifically to epitope sequences are commercially available from, *e.g.*, Covance Research Products, Inc.; Bethyl Laboratories, Inc.; Abcam Ltd.; and Novus Biologicals, Inc.

[0247] Purification tags also include maltose binding domains and starch binding domains. Proteins comprising purification tags can be purified using a binding partner that binds the purification tag, *e.g.*, antibodies to the purification tag, nickel or cobalt ions or resins, and amylose, maltose, or a cyclodextrin. Purification tags also include starch binding domains, *E. coli* thioredoxin domains (vectors and antibodies commercially available from *e.g.*, Santa Cruz Biotechnology, Inc. and Alpha Diagnostic International, Inc.), and the carboxy-terminal half of the SUMO protein (vectors and antibodies commercially available from *e.g.*, Life Sensors Inc.). Starch binding domains, such as a maltose binding domain from *E. coli* and SBD (starch binding domain) from an amylase of *A. niger*, are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacyclodextrin (BCD)-derivatized resin is described in WO 2005/014779, published February 17, 2005, herein incorporated by reference in its entirety. In some embodiments, a soluble, active O-glycosylated therapeutic protein comprises more than one purification or epitope tag.

[0248] Other haptens that are suitable for use as tags are known to those of skill in the art and are described, for example, in the Handbook of Fluorescent Probes and Research Chemicals (6th Ed., Molecular Probes, Inc., Eugene OR). For example, dinitrophenol (DNP), digoxigenin, barbiturates (see, *e.g.*, US Patent No. 5,414,085), and several types of fluorophores are useful as haptens, as are derivatives of these compounds. Kits are commercially available for linking haptens and other moieties to proteins and other

molecules. For example, where the hapten includes a thiol, a heterobifunctional linker such as SMCC can be used to attach the tag to lysine residues present on the capture reagent.

[0249] One of skill would recognize that modifications can be made to the catalytic or functional domains of the soluble, active eukaryotic glycosyltransferase polypeptide without diminishing their biological function. Some modifications may be made to facilitate the cloning, expression, or incorporation of the catalytic domain into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, the addition of codons at either terminus of the polynucleotide that encodes the catalytic domain to provide, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located restriction enzyme sites or termination codons or purification sequences.

XII. Production of soluble, active, O-glycosylated therapeutic proteins

[0250] Production of soluble, active O-glycosylated therapeutic protein in microorganisms that have oxidizing environments provides a bacterial system with many of the characteristics of, *e.g.*, protein production in Chinese hamster ovary cells. One difference is that the bacterial cells can be grown on standard microbial media, *e.g.*, a minimal, defined medium or a rich medium such as LB.

[0251] Production of soluble, active O-glycosylated therapeutic protein in microorganisms that have oxidizing environments provides efficiency by eliminating one or more time-consuming production steps, such as solubilization, denaturation, and refolding of the O-glycosylated therapeutic protein and/or the enzymes used for glycosylation.

[0252] Production of soluble, active O-glycosylated therapeutic protein in microorganisms that have oxidizing environments provides flexibility when designing a protein production protocol. For example, glycosyltransferases can be individually produced in a microorganism that has an oxidizing environment and then stored until needed. If desired, the microorganisms can also produce protein(s) involved in production of a nucleotide sugar that is a substrate for the glycosyltransferase. Similarly, therapeutic proteins can be individually produced in a microorganism that has an oxidizing environment. Lysates containing an appropriate glycosyltransferase, including a nucleotide sugar : polypeptide glycosyltransferase protein, can be mixed with a lysate containing the soluble therapeutic protein under conditions that promote O-glycosylation of the therapeutic protein. The

glycosyltransferases can be added sequentially to the therapeutic protein or can be added as a mixture of glycosyltransferases. With the availability of multiple lysates containing glycosyltransferases, the user can select one or more glycosylation patterns for a therapeutic protein for, *e.g.*, simultaneous production and testing. In some embodiments, purified or partially purified glycosyltransferases are used in combination with lysates containing glycosyltransferases and the therapeutic protein of interest. The purified or partially purified glycosyltransferases can be produced in, *e.g.*, a microorganism that has an oxidizing environment; a microorganism that has a reducing environment, such as wild-type *E. coli*; or a eukaryotic expression system, such as yeast, including *Saccharomyces*, *Schizosaccharomyces*, and *Pichia*, Sf9 cells, and mammalian cells, including CHO cells and NIH3T3 cells.

[0253] In another example, appropriate glycosyltransferases, including a nucleotide sugar : polypeptide glycosyltransferase protein, are expressed together in a microorganism that has an oxidizing environment. If desired, the microorganisms can also produce protein(s) involved in production of a nucleotide sugar that is a substrate for the glycosyltransferase(s).

Therapeutic proteins of interest are produced individually in a microorganism that has an oxidizing environment. The mixed glycosyltransferase lysate is combined with a lysate containing the soluble therapeutic protein under conditions that promote O-glycosylation of the therapeutic protein. As above, purified or partially purified glycosyltransferases from a variety of sources can also be added to the reaction mixtures.

[0254] In another example, one or more glycosyltransferases are expressed together in a prokaryotic microorganism that has an oxidizing environment and other glycosyltransferases are expressed individually in a prokaryotic microorganism that has an oxidizing environment. Therapeutic proteins of interest are produced individually in a microorganism that has an oxidizing environment. The appropriate glycosyltransferase lysates are combined with the lysate containing the soluble therapeutic protein under conditions that promote O-glycosylation of the therapeutic protein. As above, purified or partially purified glycosyltransferases from a variety of sources can also be added to the reaction mixtures.

[0255] In another example, appropriate glycosyltransferases, including a nucleotide sugar : polypeptide glycosyltransferase protein and a therapeutic protein, are expressed together in a microorganism that has an oxidizing environment. If desired, the microorganism can also produce protein(s) involved in production of a nucleotide sugar that is a substrate for the glycosyltransferase(s). The soluble therapeutic protein is O-glycosylated within the

microorganism and harvested from a cell lysate by the user. In a further embodiment, the cell lysate is combined with additional sugar precursors to enhance production of O-glycosylated therapeutic protein.

[0256] Any of the O-glycosylation reactions described above can be further enhanced by addition of a nucleotide sugar donor substrate to the reaction mix containing cell lysates or cell lysates and purified or partially purified glycosyltransferases.

[0257] Any of the O-glycosylation reactions described above can be further enhanced by addition of a modified nucleotide sugar donor substrate to the reaction mix containing cell lysates or cell lysates and purified or partially purified glycosyltransferases. Methods for conjugation of modified sugars to peptides or proteins are found *e.g.*, in USSN 60/328,523 filed October 10, 2001; USSN 60/387,292, filed June 7, 2002; USSN 60/391,777 filed June 25, 2002; USSN 60/404,249 filed August 16, 2002; and PCT/US02/32263; each of which are herein incorporated by reference for all purposes. Preferred modified sugars include, *e.g.*, CMP-sialic acid PEG.

[0258] It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and equivalents thereof known to those skilled in the art, and so forth.

[0259] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed. Citations are incorporated herein by reference.

EXAMPLES

Example 1: Production of glycosylated and glycoPEGylated human growth hormone using mixtures of crude lysate and/or purified reagents

[0260] Storage of Cell Pellets. Unlysed cell pellets from *trxB* gor supp mutant *E. coli* fermentation harvests of hGH (SEQ ID NO: 7), MBP-GalNAcT2, MBP-Core-1-GalT1, MBP-ST3Gal1, and GalNAcT2 coexpressed with GalE were stored at -80°C.

[0261] Cell Lysate of hGH. A human growth hormone mutant (hGH P254; T135INT, Y43A, Y144A, F140A, K141N (SEQ ID NO: 7) was expressed in a trxB, gor, supp *E. coli* strain using a pCWin2-derived vector with an alternate 5' UTR. Cell pellet was resuspended in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4. Upon resuspension, the cells were
5 homogenized with three passes over a microfluidizer set at 16,000-18,000 PSI. The cell homogenate was immediately centrifuged at 12000 x G for 50 minutes on a centrifuge equipped with a fixed angle rotor at 4°C. The supernatant was filtered using a 0.2 micron cellulose acetate filter. The filtered homogenate was stored only briefly at 4°C prior to enzymatic reaction.

[0262] Cell Lysate of MBP-GalNAcT2 and MBP-Core-1-GalT1. Cell pellet from a trxB, gor, supp *E. coli* culture expressing MBP-GalNAcT2 (SEQ ID NO:28) or MBP-Core-1-GalT1 (SEQ ID NO:46) was resuspended in 50 mM Tris-HCl, 20 mM NaCl, 5 mM EDTA, 0.02% NaN₃, pH 7.4. Upon resuspension, the cells were homogenized with three passes over a microfluidizer set at 16,000-18,000 PSI. The cell homogenate was immediately centrifuged
15 at 12000 x G for 30 minutes on a centrifuge equipped with a fixed angle rotor at 4°C. The supernatant was filtered using 0.2 micron cellulose acetate filters. A portion of the filtered homogenate was concentrated and buffer exchanged with 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4 in a 5 kDa MWCO centrifugal filter. The concentrated lysate was stored briefly at 4°C prior to enzymatic reaction. On a second occasion, the concentrated
20 lysate of MBP-Core-1-GalT1 was mixed 1:1 with glycerol and stored at -20°C prior to enzymatic reactions.

[0263] Cell Lysate Containing GalNAcT2 and GalE, or MBP-ST3Gal1. Cultures of trxB, gor, supp *E. coli* co-expressing GalNAcT2 (SEQ ID NO:27) with GalE, (SEQ ID NO:79) or expressing MBP-ST3Gal1 (SEQ ID NO:57) were prepared and resuspended in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4.. Cells were homogenized with three passes over a microfluidizer set at 16,000-18,000 PSI. The cell homogenate was immediately centrifuged at
25 12000-19000 x G for 20-50 minutes on a centrifuge equipped with a fixed angle rotor at 4°C. The supernatant was filtered using a 0.2 micron cellulose acetate filter. The filtrate was concentrated in a 5 kDa MWCO centrifugal filter, mixed 1:1 with glycerol and stored at -
30 20°C prior to enzymatic reaction.

[0264] Preparation of hGH-GalNAc-Gal-SA-Cys-PEG-40 kDa using *E.coli* lysates of hGH, GalNAcT2 with GalE, MBP-Core-1-GalT1 and MBP- ST3Gal1. The hGH lysate (SEQ ID

NO:7) (200 mcg hGH, 0.025 mL, 9.1 nanomoles) was combined with 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, 0.01% polysorbate 80, pH 7.4 (4.8 mL) and a solution of UDP-GlcNAc (91 nanomoles, 5.8 mL in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, 0.01% polysorbate 80, pH 7.4, 10 mg/mL). A solution of MnCl₂ (200 mM MnCl₂ in water, 4 mL) was added to the hGH lysate solution. The crude GalNAcT2/GalE enzyme lysate (0.75 mU GalNAcT2, 2 mL) was added to the hGH mixture and mixed very gently. The reaction mixture was centrifuged (1 min, 13,000 rpm) to settle particulate matter and the supernatant was removed and incubated at 32°C with gentle shaking for 17 hours. As shown in Figure 1, the addition of GalNAc was determined to be complete by MALDI analysis of the reaction mixture. The hGH-GalNAc solution (6.8 nmol, 30 mL) was combined with a solution of UDP-Gal (34 nmol, 2.1 mL in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, 0.01% polysorbate 80, pH 7.4, 10 mg/mL). The MBP-Core-1-GalT1 enzyme lysate (20 mU, 3.75 mL, 5.4 U/mL) was added to the hGH-GalNAc reaction mixture with gentle mixing. The reaction mixture was centrifuged (1 min, 13,000 rpm) to settle particulate matter and the supernatant was removed and incubated for 22 hrs at 32°C. As shown in Figure 2, the addition of Gal was determined to be complete by MALDI analysis of the reaction mixture. A solution of CMP-SA-cys-PEG-40 kDa (5.84 nanomoles, 4.56 mL in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, 0.01% polysorbate 80, pH 7.4, 1.25 mM) was added the hGH-GalNAc-Gal reaction mixture (15 mL, 2.92 nanomoles). The MBP-ST3Gal1 lysate (5.9 mU, 4.3 mL, 1.37 U/mL) and MnCl₂ (15 mL of 200 mM solution in water) were added with gentle mixing. The reaction mixture was incubated at 32°C for 62 hrs. The reaction was monitored for extent of PEGylation by SDS-PAGE and RP-HPLC (Figure 3). The hGH-GalNAc-Gal-SA-cys-PEG-40 kDa product was detected by both SDS-PAGE analysis and RP-HPLC analysis.

[0265] Preparation of hGH-GalNAc-Gal-SA-Cys-PEG-40 kDa using E.coli lysates of hGH, MBP-GalNAcT2, MBP-Core-1-GalT1 mixed with purified ST3Gal1. The hGH lysate (SEQ. ID NO:7) (22 mg hGH, 2.8 mL, 1 micromole) was diluted with 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃ (0.61 mL). A solution of UDP-GalNAc (3.26 mg, 5 micromoles, 0.33 mL in 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃, 10 mg/mL) was added to the hGH lysate. The resulting solution was adjusted to 0.01% polysorbate-80 (0.04 mL of 1% solution in 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃) and 0.22 mL of a 200 mM MnCl₂ solution in water. The crude MBP-GalNAcT2 enzyme lysate (220 mU, 0.44 mL) was added to the hGH mixture, mixed very gently and incubated at room temperature for 16

hours. The addition of GalNAc was determined to be complete by MALDI analysis of the reaction mixture. A solution of UDP-Gal (3.05 mg, 5 micromoles, 0.31 mL in 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃, 10 mg/mL) was added to the hGH-GalNAc solution and the resulting mixture was readjusted to 0.01% polysorbate-80 (0.01 mL of 1% solution in 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃). The MBP-Core-1-GalT1 enzyme lysate (2.8 U, 0.44 mL) was added to the hGH-GalNAc reaction mixture with gentle mixing and the resulting solution was incubated for 24 hrs at room temperature. The addition of Gal was determined to be complete by MALDI analysis of the reaction mixture. A solution of CMP-SA-cys-PEG-40 kDa (27.6 mg, 0.69 micromoles, 0.55 mL in 50 mM Tris-HCl, pH 7.4, 20 mM NaCl, 0.02% NaN₃) was added to the hGH-GalNAc-Gal reaction mixture (5 mg, 1.2 mL, 0.23 micromoles). Purified ST3Gal1 enzyme (0.5 U, 0.77 mL) and MnCl₂ (0.07 mL of 200 mM solution in water) were added with gentle mixing. The reaction mixture was incubated at 20 and 32°C for up to 80 hours. The reaction was monitored for extent of PEGylation by SDS-PAGE and RP-HPLC (Figures 4A and 4B). The retention time of the product hGH-GalNAc-Gal-SA-cys-PEG-40 kDa is 23.0 minutes. The retention time of the hGH starting material is 25.5 minutes (Figure 4C). Zorbax 300SB-C3 150 x 2.1 mm, 5 micron column. Over the course of the incubation, the majority of the unglycosylated hGH starting material was converted to the glycoPEGylated product, with better yield at 32°C (Figure 4D).

[0266] Purification of hGH-GalNAc-Gal-SA-cys-PEG-40 kDa produced by crude lysate mixing. hGH-GalNAc-Gal-SA-cys-PEG-40 kDa was purified from a crude lysate after glycosylation and glycoPEGylation with the glycosyltransferase-containing lysates from a trxB, gor, supp *E. coli* strain. The first step was ion exchange chromatography using a strong anion exchanger, e.g., a Q-sepharose fast flow column. The second step was size exclusion chromatography using a Superdex 200 column. A hGH-GalNAc-Gal-SA-cys-PEG-40 kDa-containing fraction was eluted from the size exclusion column. The protein was demonstrated by RP-HPLC to run mostly as a single peak (Figure 5).

[0267] Production of hGH-GalNAc-Gal-SA-Cys-PEG-40 kDa using mixed *E. coli* lysates. hGH-GalNAc-Gal-SA-cys-PEG-40 kDa can also be produced in a one pot reaction using the following reagents. GalNAc-T2 (Δ 445), GalE or GNE, Core-1-GalT1, and ST3Gal1, which are all co-expressed in a single trxB, gor, supp *E. coli* cell. The enzymes are expressed from multiple plasmids or are expressed from a single plasmid. Cell lysates containing the co-expressed enzymes are used to glycoPEGylate hGH expressed in trxB, gor, supp *E. coli* in the

presence of the following reaction components: CMP-SA-Cys-PEG-40 kDa, UDP-GlcNAc, and UDP-Glucose. The hGH can be added as part of a cell lysate or can be further purified before addition to the reaction mixture. The reaction scheme is presented in Figure 6.

[0268] Preparation of hGH-GalNAc-Gal-SA-Cys-PEG-40 kDa using purified

5 glycosyltransferases. The hGH (SEQ ID NO:7) from trxB, gor, supp *E. coli* fermentation lysates was purified by anion exchange chromatography (DEAE) and size exclusion chromatography (Superdex 75). The purified hGH (53.6 mg, 2.4 micromoles) was adjusted to 0.01% Polysorbate 80 in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4 and concentrated to a volume of 17.8 mL using a centrifugal filter (5 kDa MWCO). The UDP-

10 GalNAc (20 micromoles) was added as a solution in 50 mM Tris, 20 mM NaCl, 0.001% Polysorbate 80, 0.02% NaN₃, pH 7.4 (0.1 mL). A solution of MnCl₂ (0.1 M MnCl₂ in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4, 12.8 microliters) was added to GalNAcT2 (produced in baculovirus) enzyme (270 mU) and the resulting solution was added to the hGH solution and mixed very gently. The reaction mixture was incubated at room temperature for

15 17 hours at which time additional GalNAcT2 enzyme (270 mU) and MnCl₂ solution (0.1 M MnCl₂ in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4, 12.8 microliters) were added. The addition of GalNAc was determined to be complete by MALDI analysis of the reaction mixture after incubation for a total of 27.5 hours. A solution of MnCl₂ (0.1 M MnCl₂ in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4, 19 microliters) was added to Core-1-

20 GalT1 enzyme (produced in baculovirus) (0.238 U). The Core-1-GalT1 enzyme solution and a solution of UDP-Gal in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4 (24 micromoles, 0.1 mL) were added to the hGH-GalNAc reaction mixture with gentle mixing and the resulting solution was incubated for 15 hrs at room temperature. At this time additional Core-1-GalT1 enzyme (0.238 U) and MnCl₂ solution (0.1 M MnCl₂ in 50 mM

25 Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4, 19 microliters) were added to the hGH reaction mixture. The reaction mixture was incubated at room temperature for an additional 17.5 hrs and the addition of Gal was determined to be complete by MALDI analysis. The CMP-SA-Cys-PEG-40 kDa (4.85 micromoles) was added to the hGH-GalNAc-Gal reaction mixture as a solution in 50 mM Tris-HCl, 20 mM NaCl, 0.02% NaN₃, pH 7.4 (1 mL) and ST3Gal1

30 enzyme (produced in baculovirus) (4.26 U) was added with gentle mixing. The reaction mixture was incubated at room temperature for 50 hrs and was monitored for extent of PEGylation by SDS PAGE and RP-HPLC. The product, hGH-GalNAc-Gal-SA-cys-PEG-40 kDa, was purified using SP Sepharose, SEC (Superdex 200), and Phenyl Sepharose

chromatography. The purified hGH-GalNAc-Gal-SA-cys-PEG-40 kDa was concentrated and then formulated. The product was analyzed by a BCA protein assay, SDS-PAGE gels (silver stain), RP-HPLC, and for cell proliferation. This process yielded 6.11 mg of active hGH-GalNAc-Gal-SA-cys-PEG-40 kDa in >96% purity.

5 Example 2: Expression of soluble, glycosylated human growth hormone in bacteria.

[0269] hGH was expressed from a pACYC184-derived vector carrying the expression cassette from pCWin2 with a modified 5' UTR, in a *trxB*, *gor*, *supp* *E. coli* strain. The *E. coli* cells also included a vector that expressed GalNAc-T2 ($\Delta 51 \Delta 445$) with GalE (SEQ ID NO:79) or another epimerase, GNE (SEQ ID NO:80), from pCWin-2 derived expression
10 cassettes with modified 5' UTRs. Thus, the *trxB*, *gor*, *supp* *E. coli* strain expressed hGH P254 at relatively low expression levels, and GalNAc-T2 ($\Delta 51 \Delta 445$) and a dual specificity Glc/GlcNAc epimerase at relatively high expression levels. Cells were grown in medium in shake flasks. RP-HPLC was used to monitor the reaction. Results are shown in Figure 7. A chromatogram of cell lysate from a control *E. coli* culture that expressed only hGH (peak
15 26.2) is shown in Figure 7A. A chromatogram of the cell lysate from hGH (split peak circa 26.0) and enzyme-expressing *E. coli* strain is shown in Figure 7B. The hGH elution profile is different for the two cultures, indicating that hGH is glycosylated when grown in the presence of the GalNAc-T2 ($\Delta 51 \Delta 445$) protein. The table in Figure 8 summarized the expressed yield and percent conversion of hGH to hGH-GalNAc in *E. coli* cells that expressed
20 either GalE or GNE with the GalNAc-T2 ($\Delta 51 \Delta 445$) protein. The percent conversions ranged from 31.2% to 41.2%, with a slightly higher conversion observed with expression of the GalE protein. The presence of hGH-GalNAc in lysates from *E. coli* cells that expressed hGH, GalNAc-T2 ($\Delta 51 \Delta 445$), and an epimerase was verified by LC/MS. Results are shown in Figure 9.

25 [0270] Additional glycosyltransferases of the O-linked glycosylation pathway can be co-expressed with GalNAc-T2 ($\Delta 445$) and hGH or a therapeutic protein of interest in a *trxB*, *gor*, *supp* *E. coli* strain. An exemplary reaction scheme is presented in Figure 10.

[0271] For example, glycosylated hGH was produced by the coexpression of hGH with GalNAcT2 ($\Delta 53 \Delta 445$), Core-1-GalT1 and GNE in *trxB*, *gor*, *supp* *E. coli* from a
30 polycistronic expression cassette on a single plasmid. Expression was induced in shake flasks at 20°C for approximately 16 hours. Two cassettes were tested; in the first, an extended 5' UTR precedes each cistron. In the second cassette, an extended 5' UTR only

precedes the first two cistrons, while simple Shine-Dalgarno sequences precede the last two cistrons. Crude cell lysates from induced cells bearing either of these two polycistronic constructs was assayed for hGH levels and glycosylation by RP-HPLC and LC/MS. As summarized in Figure 11, total hGH expression was approximately 50-60 mg/L, with 30% conversion yielding approximately 15-18 mg/L hGH-GalNAc. In a separate experiment, GalNAcT2 and Core-1-GalT1 activity was measured in lysates similarly prepared from these strains (Figure 12). Both enzymes were active. In addition, hGH in lysate from *trxB*, *gor*, *supp E. coli* expressed from polycistronic expression cassette 2 was further glycosylated in vitro in a reaction of clarified lysate with 50mM Tris pH 7, 20mM NaCl, 0.2% NaN₃, 0.01% Tween-80, 20mM MnCl₂, and 10mM of the indicated sugar nucleotides (Figure 13). The reaction was carried out for five hours at 32°C, and conversion of hGH monitored by LC/MS. As plotted in Figure 13, yield of hGH-GalNAc was increased by the addition of UDP-GlcNAc. In turn, hGH-GalNAc was readily converted to hGH-GalNAc-Gal with the addition of UDP-Gal or UDP-Glc. The addition of both UDP-GlcNAc and UDP-Glc resulted in the conversion of over 60% of total hGH into hGH-GalNAc-Gal. Finally, *trxB*, *gor*, *supp E. coli* cells bearing polycistronic expression cassette 2 were induced up to 40 hours in shake flasks at 20°C. The distribution of hGH glycosylation in crude extract from these cells was determined by LC/MS, and is plotted in Figure 14A. In a separate experiment, cells bearing polycistronic expression cassette 2 were induced in shake flasks at 20°C for 66 hours. The distribution of hGH glycosylation in crude extract from these cells was determined by LC/MS as shown in Figure 14B. After the 66 hour induction, the hGH-GalNAc-Gal product was detected.

Example 3: Co-expression of glycosyltransferases with a protein disulfide isomerase.

[0272] MBP-tagged human GalNAcT2 Δ51 (SEQ ID NO:28) and MBP-tagged rat ST3Gal3 (Δ72, SEQ ID NO:78) were co-expressed in *trxB gor supp* mutant *E. coli* cells with a protein disulfide isomerase. *E. coli* DsbC (D22H Δ20) was cloned by PCR into a ampicillin-selectable vector derived from pACYC177 bearing the expression cassette from pCWin2 with a modified 5' UTR. Numbering of the N-terminal deletion and the D22H mutation is based on full length DsbC, SEQ ID NO:18. The deletion of DsbC residues 1-20 removes its signal sequence and is predicted to result in the cytoplasmic expression of DsbC. Methods, including protein induction, lysate preparation, and activity analyses, were essentially as described in Examples 4 and 6.

[0273] Either GalNAcT2 alone or DsbC and GalNAcT2 expression plasmids together were transformed into *trxB gor* supp mutant cells, and induced for expression overnight at 20°C in shake flask cultures. As shown in Table 2, MBP-GalNAc-T2 was solubly expressed at higher levels when coexpressed with DsbC. Similarly, coexpression of MBP-ST3Gal3 with DsbC in *trxB gor* supp mutant cells improved expressed activity levels relative to the MBP-ST3Gal3-only cells.

Table 2: Yields based on observed enzyme activity of glycosyltransferases coexpressed with a protein disulfide isomerase

A summary of the highest observed activities in lysate samples for the indicated glycosyltransferases expressed in *trxB gor* supp mutant *E. coli* with or without coexpression of a protein disulfide isomerase.

Expressed Glycosyltransferase	Expressed in <i>trxB gor</i> supp mutant cells Enzyme Activity (U/L)	Coexpressed in <i>trxB gor</i> supp mutant cells with a protein disulfide isomerase, DsbC Enzyme Activity (U/L)
MBP-GalNAcT2	6.4	25
MBP-ST3Gal3	4	20

Example 4: Expression of eukaryotic glycosyltransferase mutants as soluble proteins in bacteria.

Eukaryotic glycosyltransferases from the O-linked glycosylation pathway

[0274] GalNAcT2 $\Delta 51 \Delta 445$ wildtype and mutant constructs designed to alter putative solvent exposed residues (Q130E, W131S, Y177K, F316D, and F395D, mutation positions indicated based on full length wildtype human GalNAcT2) were expressed in a *trxB, gor, supp E. coli* strain. Cells were grown in shake flasks, and induced for at least 16 hours at 20°C with IPTG. Crude lysates were prepared and tested for expressed GalNAc transferase activity in a reaction with 20 mM Tris pH 7, 10 mM $MnCl_2$, 1.5 mM UDP-GalNAc, 1 mM synthetic peptide acceptor. Following a 30 minute incubation at 37°C, the reaction was quenched with 0.01N HCl, and the peptide acceptor was separated from the reaction mix by centrifugation through a 10,000 MWCO concentrator. Peptide and GalNAc-peptide were detected and quantified by RP-HPLC. One of the GalNAcT2 substitution mutants, F316D, improved enzyme activity expressed in *trxB, gor, supp E. coli* by 50% (Figure 15). Only the W131S mutation exhibited reduced enzymatic activity as compared to the unmutated

GalNAcT2 protein. Activities of the other three mutants were about equal to those of the unmutated GalNAcT2 protein.

[0275] Core-1-GalT1 $\Delta 50$ and $\Delta 31$ were expressed in *trxB*, *gor*, *supp E. coli* shake flask cultures induced with IPTG at 20°C for at least 16 hours. Clarified cell lysates were tested for relative galactosyltransferase activities in a reaction with a synthetic acceptor glycopeptide containing a GalNAc-Thr residue. GalNAc-peptide and Gal-GalNAc-peptide were detected and quantified by RP-HPLC. The shorter $\Delta 50$ form of untagged Core-1-GalT1 solubly expressed at 5.1 U/L, whereas the longer $\Delta 31$ form solubly expressed at 1.7 U/L.

Eukaryotic glycosyltransferases from the N-linked glycosylation pathway

[0276] MBP-tagged ST3Gal3 $\Delta 72$ and untagged ST3Gal3 $\Delta 72$ constructs were expressed in a *trxB*, *gor*, *supp E. coli* strain. Both were grown in shake flasks, and untagged ST3Gal3 was also grown in a 10L batch fermentation. Clarified lysates were prepared and tested for expressed sialyltransferase activity in a reaction with 20 mM MOPS pH 6.5, 0.1 mg/ml BSA, 10 mM $MnCl_2$, 2 mM CMP-NAN, and 30 mM lacto-n-neotetraose. Following a two hour incubation at 30°C, the reaction was stopped by heat inactivation, and the reaction substrate and product detected and quantified by HPLC. In shake flasks, MBP-ST3Gal3 and untagged ST3Gal3 solubly expressed at 4.6 and 1 U/L, respectively. In the fermentor, soluble expression of untagged ST3Gal3 reached 2 U/L by 48 hours post-induction.

[0277] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

INFORMAL SEQUENCE LISTING

SEQ ID NO:1.

Human FGF-20

5 MAPLAEVGGFLGGLEGLGQQVGSFLLPPAGERPPLLGERRSAAERSARGGPGAAQLAHLHGILRRRQLYCRTGF
HLQILPDGSGVQGTRQDHSLEFGILEFISVAVGLVSIRGVDSGLYLGMNDKGELYGSEKLTSECI FREQFEENWYNT
YSSNIYKHGDTGRRYFVALNKDGTPRDGARSKRHQKFTHFLPRPVDPERVPELYKDLLMYT

SEQ ID NO:2.

10 Human mature FGF-21

MDSDETGFEHSGLWVSVLAGLLLGACQAHPIPDSSPLLQFGGQVRQRYLYTDDAQQTEAHLEIREDGTVGGAADQ
SPESLLQLKALKPGVIQILGVKTSRFLCQRPDGAlyGSLHFDPEACSFRELLLEDGYNVYQSEAHGLPLHLPGNK
SPHRDPAPRGPARFLPLPGLPPALPEPPGILAPQPPDVGSSDPLSMVGPSQGRSPSYAS

15 SEQ ID NO:3.

Human Glucocerebrosidase aa 38-536

MGARPCIPKSFEGYSSVVCNATYCDSDPPTFPALGTFSTRYESTRSGRRMELSMGPIQANHTGTGLLLTLQPEQ
KFQKVKGFGGAMTDAAALNILALSPPAQNLLLSYFSEEGIGYNIIRVPMASCDFSIRTYTYADTPDDFQLHNFS
LPEEDTKLKIPLIHRALQLAQRVSLASPWTSPTWLKTNGAVNGKSLKGQPGDIYHQTWARYFVKFLDAYAEH
20 KLQFWAVTAENEPSAGLLSGYPFQCLGFTPEHQDFIARDLGPTLANSTHHNVRLMLDDQRLLLPHWAKVVLT
PEAAKYVHGIAVHWYLDFLAPAKATLGETHRLFPNTMLFASEACVGSKFWEQSVRLGSWDRGMQYSHSIITNLLY
HVVGWTDWNLALNPEGGPNNVRFVDSPIIVDITKDTFYKQPMFYHLGHFSKFIPEGSRVGLVASQKNDLDAVA
LMHPDGSVVVVVLRSSKDVPLTIKDPVGFLETISPGYSIHTYLWRRQ

25 SEQ ID NO:4.

Human NT-3

MYAEHKSHRGEYSVCDSESLWVTDKSSAIDIRGHQVTVLGEIKTGNMSPVKQYFYETRCKEARPVKNGCRGIDDKH
WNSQCKTSQTYVRALTSENNKLVGWRWIRIDTSCVCALSRKIGRT

SEQ ID NO:5.

MBP-tagged human pro-NT-3

MKTEEGKLVIWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIIFWAHDRFGGYAQSG
 LLAEITPDKAFQDKLYPFTWDVRYNGKLIAYPIAVEALSIIYNKDLLPNFPKTWEEIPALDKELKAKGKSALMF
 5 NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAM
 TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
 GAVALKSYYYEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
 NNNNNNLGPGAAHYVEFGSHMQGNNMDQRSLPEDSLNSLI IKLIQADILKNKLSKQMVDVKENYQSTLPKAEAPR
 EPERGGPAKSAFQPVIAMDTLLRQQRRYNSPRVLLSDSTPLEPPPLYLMEDYVGSPPVANRTSRRKRYAEHKSH
 10 RGEYSVCDSESLWVTDKSSAIDIRGHQVTVLGEIKTGNSPVKQYFYETRCKEARPVKNGCRGIDDKHWSQCKTS
 QTYVRALTSENNKLVGWRWIRIDTSCVCALSRKIGRT

SEQ ID NO:6.

Mutant human GH #1

MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPQTSLCFSES IPTPSNREETQQKSNLE
 LLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSP TINTIFKQTYSKFDTN
 SHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF

SEQ ID NO:7.

20 Mutant human GH #2

MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKASFLQNPQTSLCFSES IPTPSNREETQQKSNLE
 LLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSP TINTIANQTASKFDTN
 SHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF

25 SEQ ID NO:8.

dnaK E. coli EG10241 HSP-70-type molecular chaperone, heat-inducible

MGKIIGIDLGTNSCAVIMDGTTPRVLENAEGDRTPSIIAYTQDGETLVGQPAKRQAVTNPQNTLFAIKRLIGR
 RFQDEEVQRDVSIMPFKIIAADNGDAWVEVKGQKMAPQISAEVLKMKKTAEDYLGEFPVTEAVITVPAYFNDAQ
 RQATKDAGRIAGLEVKRIINEPTAAALAYGLDKGTGNRTIAVYDLGGGTFDISIIIDEVDGEKTFEVLATNGDT
 30 HLGGEDEFDSRLINYLVEEFKKDQGIDLRNDPLAMQRLKEAAEKAKIELSSAQQTVDNLPYITADATGPKHMNIKV
 TRAKLESLEDLVNRSIEPLKVALQDAGLSVSDIDDVILVGGQTRMPMVQKKVAEFFGKEPRKDVNPDEAVAIGA
 AVQGGVLTGDVKDVLVLLDVTPLSLGIETMGGVMTTLIAKNTTIPTKHSQVFSTAEDNQSAVTIHLVQGERKRAAD
 NKS LGQFNLDGINPAPRGMPQIEVTFDIDADGILHVS AKDKNSGKEQKITIKASSGLNEDEIQKMVRDAEANA EA
 DRKFEELVQTRNQGDHLLHSTRKQVEEAGDKLPADDKTAIESALTALETALKGEDKAAIEAKMQELAQVSQKLME
 35 IAQQQHAQQQTAGADASANNAKDDDVDAEFEEVKDKK

SEQ ID NO:9.

clpB, E. coli EG10157 ClpB protease, ATP dependent

MRLDRLTNKFQLALADAQSLALGHDNQFIEPLHLMSALLNQEGGSVSPLLTSAGINAGQLRTDINQALNRLPQVE
 40 GTGGDVQPSQDLVRVLNLCDKLAQKRGDNFISSELFVLAALLESRGTLADILKAAGATTANITQAIEQMRGGESVN
 DQGAEDQRQALKKYTIDLTERAEQGLDPVIGRDEEIRRTIQVLQRRTKNNPVLIGEPGVGKTAIVEGLAQRIIN

GEVPEGLKGRRLALDMGALVAGAKYRGEFEERLKGVLNDLAKQEGNVILFIDELHTMVGAGKADGAMDAGNMLK
 PALARGELHCVGATTLD EYRQYIEKDAALERRFQKV FVAEPSVEDTIAILRGLKERYELHHHVQITDPAIVAAAT
 LSHRYIADRQLPKAIDLIDEAASSIRMQIDSKPEELDRLDRRI IQLKLEQQALMKESDEASKKRDLMLNEELSD
 KERQYSELEEEWKAEKASLSGTQTITKAELEQAKIAIEQARRVGD LARMSELQYKG IPELEKQLEAATQLEGKTMR
 5 LLRNKVTDAEIAEVLARWTGIPVSRMMESEREKLLRMEQELHHRVIGQNEAVDAVSNAIRRSRAGLADPNRPIGS
 FLFLGPTGVGKTELCKALANFMFDSDEAMVRIDMSEFMEKHSVSR LVGAPPGYVGYEEGGYLTEAVRRRPYSVIL
 LDEVEKAHPDVFNILLQVLDDGRLTDGQGRTVDFRNTVVM TSNLGSDLIQERFGELDYAHMKELVLGVVSHNFR
 PEFINRIDEVVVFHPLGEQHIA SIAQIQLKRLYKRLEERGYEIHISDEALKLLSENGYDPVYGARPLKRAIQQQI
 ENPLAQQILSGELVPGKVIRLEV NEDRIVAVQ

10

SEQ ID NO:10.

GroEL, *E. coli*

AAKDVKFGNDARVKMLRGVNLADAVKVTLGPKGRNVLDK SFGAPTITKDGVS VAREIELEDKFENMGAQMVK E
 VASKANDAAGDGT TTTATVLAQAIITEGLKAVAAGMNPMDLKR GIDKAVTA AVEELKALSVP CSDSKAIAQVGTIS
 15 ANSDETVGKLIAEAMD KVGKEGVITVEDGTGLQDELDVVEGMQFDRGYLSPYFINKPETGAVELES PFILLADKK
 ISNIREMLPVLEAVAKAGKPLLI I AEDVEGEALATLVVNTMRGIVKVA AVKAPGFGDRRKAMLQDIATLTGGTVI
 SEEIGMELEKATLEDLGQAKRVVINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQERVAKLGGV
 AVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALIRVASKLADLRGQNE DQNVGIKVALRAMEAPL
 RQIVLNCGEEPSVANTVKG DGNYGYNAATEEYGNMIDMGILDPTKVTR SALQYAASVAGLMITTECMVTDLPK
 20 NDAADLGAAGGMGGM

SEQ ID NO:11.

Cpn10, *Oleispira antarctica*

MKIRPLHDRVRRKEETATAGGIILPGAAAEKPNQGVVISVGTGRILDNGSVQALAVNEG DVVFGKYSQNT
 25 IDIDGELLILNESDIYGVLEA

SEQ ID NO:12.

Cpn 60, *Oleispira Antarctica*

MAAKDVLFGDSARAKMLVGVN ILADAVRVTLGPKGRNVVIEK SFGAPIITKDGVS VAREIELKDKFENMGAQMV K
 30 EVASQANDQAGDGT TTTATVLAQAI ISEGLKSVAAGMNPMDLKR GIDKATAAVVA AIKEQAQPCLDTKAIAQVGTI
 SANADETVGRLIAEAMEKVGKEGVITVEEGKGLEDEL DVVEGMQFDRGYLSPYFINNQEKMTVEMENPLILLVDK
 KIDNLQELLPILENVAKSGRPLLI VAEDVEGQALATLVVNNLRGT FKVA AVKAPGFGDRRKAMLQDLAILTGGQV
 ISEELGMSLETADPSSLGTASKVVIDKENTVIVDGAGTEASVNTRVDQIRAEIESSTS DYDIEKLQERVAKLGG
 VAVIKVGAGSEMEMKEKKDRVDDALHATRAAVEEGVVAGGGVALIRALSSVTTVVGDNEDQNVGIALALRAMEAPI
 35 RQIAGNAGAEGSVVVDKVKSGTGSFGFNASTGEYGD MIAMGILDPKVTRSSLQAAASIAGLMITTEAMVADAPV
 EEGAGGMPDMGGMGGMGGMGMPGMM

SEQ ID NO:13.

Rat PDI protein, PDIA1_RAT

MLSRALLCLALAWAARVGADALEEEDNVLVLKSNFAEALAAHNYLLVEFYAPWCGHCKALAPEYAKAAAKLKAE
 40 GSEIRLAKVDATEESDLAQYGV RGYPTIKFFKNGDTASPK EYTAGREADDIVNWLKKRTGPAATTLSDTAAAES

LVDSSEVTVIGFFKDAGSDSAKQFLLAAEAVDDIPFGITSNSDVFSKYQLDKDGVVLFKKFDEGRNNFEGEITKE
KLLDFIKHNQLPLVIEFTEQTAPKIFGGEIKTHILLFLPKSVSDYDGKLSNFKKAAEGFKGKILFIFIDSDHTDN
QRILEFFGLKKEECPAVRLITLLEEEMTKYKPESDELTAEKITQFCHHFLEGKIKPHLMSQELPEDWDKQPVKVLV
GKNFEEVAFDEKKNVFVEFYAPWCGHCKQLAPIWDKLGETYKDHENIVIAKMDSTANEVEAVKVHSFPTLKFFPA
5 SADRTVIDYNGERTLDGFKKFLESQGDGAGDNDLDLEEALPDMEEDDDQKAVKDEL

SEQ ID NO:14.

Erolp Sacchromyces

MRLRTAIATLCLTAFTSATSNNSYIATDQTQNAFNDTHFCKVDRNDHVSPSCNVTFNELNAINENIRDDLSALLK
10 SDDFFKYFRDLKYQCSFWDANDGLCLNRACSVDVVEDWDTLPEYWQPEILGSFNNDTMKEADDSDECKFLDQLC
QTSKKPVDIEDTINYCDVNDFNGKNAVLIDLTANPERFTGYGGKQAGQIWSTIYQDNCFTIGETGESLAKDAFYR
LVSGFHASIGTHLSKEYLNTKTGKWEPNLDLFMARIGNFPDRVTNMYFNYAVVAKALWKIQPYLPEFSFCDLVNK
EIKNKMNDVISQLDTKIFNEDLVFANDLSLTLDKDEFRRFKNVTKIMDCVQCDRCRLWGKIQTGTGYATALKILFE
INDADEFTKQHIVGKLTKEYELIALLOTTFGRLSIESVNMFEKMYGKRLNGSENRLSSFFQNNFFNILKEAGKSI
15 RYTIENTINSTKEGKKKTNNQSQSHVFDDLKMPKAEIVPRPSNGTVNKKWKAWNTEVNNVLEAFRFIYRSYLDLPRN
IWELSLMKVYKFWNKFIGVADYVSEETREPISYKLDIQ

SEQ ID NO:15.

Pdilp Sacchromyces

MKMNLRKRLVVTFSCITFLLKFTIAAAEPPEGFPEPLNPTNFKEELSKGLHIIDFYSPYCPHCKHLAPVWMETWE
20 EFKEESKTLNITFSQVNCIESADLCGDENIEYFPEIRLYNP SGYIKSFTETPRTKESLIAFARRESMDPNNLDTD
LDSAKSESQYLEGFDLELIAGKATRPHLVSWFPTKDMKNSDDSLEFKNCDKCHEFORTWKIISRQLAVDDINTG
HVNCESNPTICEELGFGDLVKITNHRADREP KVALVLPNKTSNNLFDYPNGYSAKSDGYVDFARRTFTNSKFPNI
TEGELEKKANRDIDFLQERGRVTNNDIHLVFSYDPETVVEDFDILEYLIIEPLSKI PNIIYLHQIDKNLINLSRNL
25 FGRMYEKINYDASQTQKVFNKEYFTMNTVTQLPTFFMEFKDGDPISYVFPGYSTTEMRNIDAIMDWVKYSNPLVT
EVDSSNLKKLISFQTKSYDLAIQLISSTDHKKHIGKSNKLIKNNLLASWEYEHIRMENNFEEINERRARKADGIK
KIKEKKAPANKIVDKMREEIPHMDQKLLGLYLDISKEKNFRKYGITGEYKIGDVIIDKSNYYYNNKDNFGNS
LTSNNPQLLREAFVSLNIPSKALYSSKLKGRLLNSPFHNVLSFLDIHGNMGPGYLIVIVLFIAILKGPSIYRRY
KVRKHYRAKRNAVGILGNMEKKKNQD
30

SEQ ID NO:16.

DsbB protein, E. coli

MLRFLNQCSQGRGAWLLMAFTALALELTALWFQHVMLLKPCVLCIYERCALFGVLGAALIGAIAPKTPLRYVAMV
IWLYSAFRGVQLTYEHTMLQLYPSPFATCDFMVRFPFWLPLDKWVPQVFVASGDCAERQWDFLGLEMPQWLLGIF
35 IAYLIVAVLVVISQPFKAKKRDLFGR

SEQ ID NO:17.

DsbA protein, E. coli

MKKIWLALAGLVLAFSASAAQYEDGKQYTTLEKPVAGAPQVLEFFSFFCPHCYQFEEVLHISDNVKKKLPEGVKM
40 TKYHVNFMGGDLGKDLTQAWAVAMALGVEDKVTVPFLFEGVQKTQTI RSASDIRDVFINAGIKGEEYDAAWNSFVV
KSLVAQQEKAADVQLRGVPAMFVNGKYQLNPQGMDSNMDVFVQYADTVKYLSEKK

SEQ ID NO:18.

DsbC protein, *E. coli*

5 MKKGFMLFTLLAAFSGFAQADDAAIQQTAKMGIKSSDIQPAPVAGMKTVLTNSGVLYITDDGKHIIQGPMYDVS
GTAPVNVNTNKMLLKQLNALEKEMIVYKAPQEKHVITVFTDITCGYCHKLHEQMADYNALGITVRYLAFFRQGLDS
DAEKEMKAIWCAKDKNKAFDDVMAGKSVAPASCDVDIADHYALGVQLGVSGTPAVVLSNGTLVPGYQPPKEMKEF
LDEHQKMTSGK

SEQ ID NO:19.

10 DsbD protein, *E. coli*

MAQRIFTLILLLLCSTSVFAGLFDAPGRSQFVPADQAFADFQQNQHDNLNTWQIKDGYLYLRKQIRITPEHAKIA
DVQLPQGVWHEDEFYKSEIYDRRLTLPVTINQASAGATLTVTYQGCADAGFCYPPETKTVPLSEVVANNAAPQP
VSVFPQQEQPTAQLPFSALWALLIGIGIAFTPCVLPMPYPLISGIVLGGKQRLSTARALLLTFIYVQGMALTYTALG
LVVAAAGLQFQAALQHPYVLIGLAIVFTLLAMSMFGLFTLQLPSSLQTRLTLMSNRQGGSPGGVFVMGAIAGLI
15 CSPCTTAPLSAILLYIAQSGNMWLGGGTLYLYALGMGLPLMLITVFGNRLLPKSGPWMEQVKTAFGFVILALPVF
LLERVIGDVWGLRLWSALGVAFFGWAFITSLQAKRGWMRIVQIILLAAALVSVRPLQDWAFGATHTAQTQTHLNF
TQIKTVDELNQALVEAKGKPVMLDLYADWCVACKEFEKYTFSDPQVQKALADTVLLQANVTANDAQDVALLKHLN
VLGLPTILFFDGQGEHPQARVTGFMDAETFSAHLRDRQP

20 SEQ ID NO:20.

DsbG protein, *E. coli*

MTVIGYAFYSTFALTEKDKMLKILLALLPAIAFAEELPAPVKAIEKQGITI IKTDFDAPGGMKGYLGKYQDMG
VTIYLTPDGKHAISGYMYNEKGENLSNTLIEKEIYAPAGREMWRMEQSHWLLDGKKDAPVIVYVFADPFCPYCK
QFWQQARPWVDSGKVQLRTLVLGVKIPESPATAAAILASKDPAKTWQQYEASGGKLLNVPANVSTEQMKVLSDN
25 EKLMDDLGANVTPAIYYMSKENTLQQAVGLPDQKTLNIIMGNK

SEQ ID NO: 21.

human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2)

30 MRRRSRMLLCFAFLWVLGIAYMYSGGGSALAGGAGGGAGRKEDWNEIDPIKKKDLHHSNGEKAQSMETLPPGK
VRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARS
ALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDS
HCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDFQYVGASADLKGGFOWNLVFKWDYMTPEQRRSRQGNPVAP
IKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIPC SRVGHVFRKQHPYTFPGGSGT
VFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPFWYLENVYPDLRVPDHDIAFGAL
35 QQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQI
EGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ

SEQ ID NO: 22.

Drosophila UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2)

40 MRRNIKLIVFVSIIWMFVMVYFQSSTEKVENRALRLREVATAMQQYQDDSSSSAAASTARQWAPAGGGAGPGAA
AGAAGSGADDPGGNVILIGSVKDFERNAVHGLKLNIGIVALEETSQGLSGGTGGPGGRLPVAPSGRGTEVEYFNEA

GYIRAGALRNGEDPYIRNRFNQEASDALPSNRDIPDTRNPMCRTKKYREDLPETSVIITFHNEARSTLLRTIVSV
 LNRSPHLIREIVLVDDYSDHPEDGLELAKIDKVRVIRNDKREGLVRSRVKGADAAVSSVLTFLDHSHVECNEMWL
 EPLLERVREDPTRVVCVIDVISMDNFQYIGASADLRGGFDWNLIKFWEYLSPSERAMRHNDPTTAIRTPMIAGG
 LFVIDKAYFNKLGKYDMKMDVWGGENLEISFRVWQCGGSLEIIPCSRVGHVFRKRHPYTFPGSGNVFARNTRRA
 5 AEVWMDDYKQHYNAVPLAKNIPFGNIDDRALKEKLHCKPFWYLENVYPDLQAPDPQEVGQFRQDSTECLDTM
 GHLIDGTVGIFPCHNTGGNQEWAFTRKEIKHDDLCLTLVTFARGSQVVLKACDDSENQRWIMREGGLVRHYKIN
 VCLDSRDQSQQGVSAQHCHNSALGTQRWSFGKYA

SEQ ID NO: 23.

10 mouse UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2)
 MRRRSRMLLCFALLWVLGIAYMYSGGGSALAAGGGGAGRKGWDNDIDSIIKKDLHHSRGDEKAQGVETLPPGKV
 RWPFDNFQEAYVGGTMVRSGQDPYARNKFNQVESDKLHMDRGIPTDRHDQCQRKQWRVDLPATSVVITFHNEARSA
 LLRTTVSVLKRSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSH
 CECNERWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSRQGNPVAPI
 15 KTPMIAGGLFVMDKLYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIPCSRVGHVFRKQHPYTFPGSGTV
 FARNTRRAAEVWMDEYKHFYAAVPSARNVPYGNIQSRLELRKKLGCKPFWYLDNVYPELRVPDHDQDIAFGALQ
 QGTNCLDTLGHFADGVVGIYECHNAGGNQEWALTKEKSVKHMDLCLTVVDRSPGSLIRLQGCRENDSRQKWEQIE
 GNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFSNLQQ

20 SEQ ID NO: 24.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ51)
 KKKDLHHSNGEEKAQSMETLPPGKVRWPFDNFQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQC
 QRKQWRVDLPATSVVITFHNEARSALLRTTVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRR
 EGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDW
 25 NLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKLYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEI
 IPCSRVGHVFRKQHPYTFPGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPF
 KWYLENVYPELRVPDHDQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVD
 RAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNQQ

30 SEQ ID NO: 25.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ1-51
 Δ445-571)
 KKKDLHHSNGEEKAQSMETLPPGKVRWPFDNFQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQC
 QRKQWRVDLPATSVVITFHNEARSALLRTTVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRR
 35 EGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDW
 NLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKLYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEI
 IPCSRVGHVFRKQHPYTFPGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPF
 KWYLENVYPELRVPDHDQD

SEQ ID NO: 26.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ 51)
alternate form

MSKKKDLHHSNGEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHD
5 QCQRKQWRVDLPATSVVITFHNEARSALLRTTVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRND
RREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGF
DWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSL
EIIPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCK
10 PFKWYLENVYPELRVPDHQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTV
VDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLOQ

SEQ ID NO: 27.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ 1-51
 Δ 445-571) alternate form

MSKKKDLHHSNGEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHD
15 QCQRKQWRVDLPATSVVITFHNEARSALLRTTVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRND
RREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGF
DWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSL
EIIPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCK
20 PFKWYLENVYPELRVPDHQD

SEQ ID NO: 28.

Maltose binding protein-tagged truncated human UDP-N-
acetylgalactosaminyltransferase 2 (MBP-GalNAc-T2 Δ 51)

MKIEEGKLVIIWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEFQVAATGDGPDIIFWAHDREGGYAQSG
25 LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFVLVDLIKHKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSYYYEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
30 NNNNNNLGIEGRISEFGSKKKDLHHSNGEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVE
SDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTTVSVLKKSPPHLIKEIILVDDYSNDPED
GALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINM
DNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGG
ENLEISFRVWQCGGSLEIIPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPY
35 GNIQSRLELRKKLSCKPFKWYLENVYPELRVPDHQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWA
LTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPA
LSQQWKFTLNLOQ

SEQ ID NO: 29.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ53)

KDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQCQR
KQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRREG
5 LMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNL
VFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIP
CSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPFW
YLENVYPELRVPDHDQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVDRA
PGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ

SEQ ID NO: 30.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ1-53
Δ445-571)

KDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQCQR
15 KQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRREG
LMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNL
VFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIP
CSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPFW
YLENVYPELRVPDHDQD

SEQ ID NO: 31.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ53)
alternate form

MSKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQC
25 QRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRR
EGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDW
NLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEI
IPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPF
KWYLENVYPELRVPDHDQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVD
30 RAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ

SEQ ID NO: 32.

Truncated human UDP-N-acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ1-53
Δ445-571) alternate form

MSKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQC
35 QRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGKIEKVRVLRNDRR
EGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDW
NLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEI
IPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPYGNIQSRLELRKKLSCKPF
40 KWYLENVYPELRVPDHDQD

SEQ ID NO: 33.

Maltose binding protein-tagged truncated human UDP-N-

acetylgalactosaminyltransferase 2 (MBP-GalNAc-T2 Δ1-53 Δ445-571)

MKIEEGKLVIWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
 5 LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEI PALDKELKAKGKSALMF
 NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKKNHMNADTDYSIAEAAFNKGETAM
 TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
 GAVALKSYYYEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
 NNNNNNLGIEGRISEFGSHMSKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQV
 10 ESKLRLMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPE
 DGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVIN
 MDNFQYVGASADLKGFDWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWG
 GENLEISFRVWQCGGSLEIIPCVRGHVFRKQHPYTFPGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVP
 YGNIQSRLELRKKLSCKPFWYLENVYPELRVPDHDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEW
 15 ALTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGP
 ALSQQWKFTLNQQ

SEQ ID NO: 34.

Maltose binding protein-tagged truncated human UDP-N-

acetylgalactosaminyltransferase 2 (MBP-GalNAc-T2 Δ53)

MKIEEGKLVIWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
 LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEI PALDKELKAKGKSALMF
 NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKKNHMNADTDYSIAEAAFNKGETAM
 TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
 25 GAVALKSYYYEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
 NNNNNNLGIEGRISEFGSHMSKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQV
 ESKLRLMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPE
 DGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVIN
 MDNFQYVGASADLKGFDWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKIFYFEELGKYDMMMDVWG
 30 GENLEISFRVWQCGGSLEIIPCVRGHVFRKQHPYTFPGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVP
 YGNIQSRLELRKKLSCKPFWYLENVYPELRVPDHDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEW
 ALTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGP
 ALSQQWKFTLNQQ

SEQ ID NO: 35.

Maltose binding protein-tagged truncated human UDP-N-

acetylgalactosaminyltransferase 2 (GalNAc-T2 Δ1-51 Δ445-571)

mkieegklviwingdkgynglaevgkkfekdtgikvtvehpdkleekfpqvaatgdgpdii fwahdrfggyaqsg
 llaeitpdka fqd klypftwdavryngkliaypiavealsliynkdllpnppktweeipaldkelkakgksalmf
 40 nlqepyftwpliaadgg yafkyengkydikdv gvdnagakagltflvdliknkhmnadtdysia eaa fnkgetam
 tingpwawsnidtskvnygv tvlptfkgqpskpfvgvlsaginaaspnkelakeflenylltdegleavnkdkpl

gavalksyeeelakdpriaatmenaqkgeimpnipqmsafwyavrtavinaasgrqtvdealkdaqtnssssnnnn
 nnnnnnlgiegrisefgsKKKDLHHSNGEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVE
 SDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPED
 GALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINM
 5 DNFQYVGASADLKGGFOWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGG
 ENLEISFRVWQCGGSLEIIPCVRVGHVFRKQHPYTFPGSGTVFARNTRRAAEVWMDEYKNFYAAVPSARNVPY
 GNIQSRLELRKKLSCKPFWYLENVYPELRVPDHQD

SEQ ID NO: 36.

10 Human UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
 galactosyltransferase (Core-1-Gal-T1)
 MASKSWLNFLTFLCGSAIGFLLCSQLFSILLGEKVDTQPNVLHNDPHARHSDDNQGNHLEGQMNFNADSSQHKDE
 NTDIAENLYQKVRILCWVMTGPNLEKKAKHVKATWAQRCNKVLFMSSEENKDFPAVGLKTKEGRDQLYWKTICA
 FQYVHEHYLEDADWFLKADDDTYVILDNLRWLLSKYDPEEPIYFGRRFKPYVKQGYMSGGAGYVLSKEALKRFVD
 15 AFKTDKCTHSSSIEDLALGRCMEIMNVEAGDSRDTIGKETFHFPVPEHHLIKGYLPRTFWYWNYNYPPEGPGC
 CSDLAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPTLPERILKEISQANKNEDTKVKLGNP

SEQ ID NO: 37.

20 Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
 galactosyltransferase (Core-1-Gal-T1)
 MTANSLIGRSILNEGRSNKRSFVSLIVGLIVGFCLAEFLVYSTPERSEFMPYDGHHRGDVNDAAHSHDMMEMSGP
 EQDVGGHEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGKRCNKLIIFMSSAKDDELDAVALPVGE
 GRNNLWGKTKEAYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPSPETPVYFGCKFKPYVKQGYMSGGAGY
 VLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLGDSRDSNGRGRFFFPVPEHHLIPSHTDKKFW
 25 YWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTS
 DGVSKRSAETKTQ

SEQ ID NO: 38.

30 mouse UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
 galactosyltransferase (Core-1-Gal-T1)
 MASKSWLNFLVFLCGSAIGFFLCSQLLSILLREEAAIQPNMLHNDPHARHSDDNHSHLKGQMNFNADSSQHKDE
 NIDVAENLYQVKILCWVMTSPQNLEKKAKHVKATWAQRCNKVLFMSSEENQDFPTVGLKTKEGREQLYWKTICA
 FQYVHDHYLEDADWFMKADDDTYVIVDNLRWLLSKYDPEQPIYFGRRFKPYVKQGYMSGGAGYVLSKEALRRFVN
 AFKTEKCTHSSSIEDLALGRCMEIINVEAGDSRDTIGKETFHFPVPEHHLIKGYLPKTFWYWNYNYPPIEGPGC
 35 CSDIAVSFHYVDGTTMYELEYLVYRLRPYGCYLYRYQPALPENILKEINQVNRKEDTKIKLGNP

SEQ ID NO: 39.

40 rat UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
 galactosyltransferase (Core-1-Gal-T1)
 MASKSWLNFLTFLCGSAIGFFLCSQLLNILLQEQADVQPNMLHNDPHARHSDDSGHNHLKGQMDFNADSSQHKDE
 NTDVAENLYQVKVLCWVMTSPQNLEKKAKHVKATWAQRCNKVLFMSSEENKDFPTVGLETKEGREQLYWKTICA

FQYVHDHYLEDADWFMKADDDTYVILDNLRWLLSKYNPEQPIYFGRRFKPYVKQGYMSGGAGYVLSKEALRRFVD
AFKTEKCTHSSSIEDLALGRCMEIIEAGDSRDPTGKETFHFPVPEHHLIKGYLPKTFWYWNYNYPPEGPGC
CSDIAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPALPENILKEINQVNKKEDTKIKLGNP

5 SEQ ID NO: 40.

Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
galactosyltransferase (Core-1-Gal-T1) A383T mutant

MTANSLGRSILNEGRSNKRSFVSLIVGLIVGFCLAEFLVYSTPERSEFMPYDGHHRGVDNDAHSHDMMMSGP
EQDVGGHEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWKGKRCNKLI FMSSAKDDELDAVALPVGE
10 GRNNLWGKTKEAYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGY
VLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLAGDSRDSNGRGRFFPFVPEHHLIPSHTDKKFW
YWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTS
DGVSKRSTETKTQ

15 SEQ ID NO: 41.

Truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
galactosyltransferase (Core-1-Gal-T1 Δ31) A383T mutant

GFCLAEFLVYSTPERSEFMPYDGHHRGVDNDAHSHDMMMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIM
TNPSNHQKKARHVKRTWKGKRCNKLI FMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKAD
20 DDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDV
EIGKCLQNVNVLAGDSRDSNGRGRFFPFVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQ
YVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSTETKTQ

SEQ ID NO: 42.

25 Truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
galactosyltransferase (Core-1-Gal-T1 Δ31) A383T mutant alternate form

MGFCLAEFLVYSTPERSEFMPYDGHHRGVDNDAHSHDMMMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWI
MTNPSNHQKKARHVKRTWKGKRCNKLI FMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKA
DDDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAED
30 VEIGKCLQNVNVLAGDSRDSNGRGRFFPFVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQ
MYVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSTETKTQ

SEQ ID NO: 43.

35 Truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-
galactosyltransferase (Core-1-Gal-T1 Δ50)

PYDGHHRGVDNDAHSHDMMMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWKG
RCNKLI FMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPYSP
ETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLAGDSRDS
NGRGRFFPFVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIINTP
40 DALPNKLAVGELMPEIKEQATESTSDGVSKRSAETKTQ

SEQ ID NO: 44.

Truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-galactosyltransferase (Core-1-Gal-T1 Δ50) A383T mutant

5 PYDGHRRHGDVNDAAHSHDMMEMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGK
RCNKLIFMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPYSP
ETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLAGDSRDS
NGRGRFFFPVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIINTP
DALPNKLAVGELMPEIKEQATESTSDGVSKRSTETKTQ

10 SEQ ID NO: 45.

Truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-galactosyltransferase (Core-1-Gal-T1 Δ50) A383T mutant alternate form

15 MSPYDGHRRHGDVNDAAHSHDMMEMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTW
GKRCNKLIFMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPY
SPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLAGDSR
DSNGRGRFFFPVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIIN
TPDALPNKLAVGELMPEIKEQATESTSDGVSKRSTETKTQ

SEQ ID NO: 46.

20 Maltose binding protein-tagged truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-galactosyltransferase (MBP-Core-1-Gal-T1 Δ50) A383T mutant

25 MKIEEGKLVIIWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNKDLLPNPKTWEIIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAFNKGETAM
TINGPWAWSNIDTSKVNIGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSIEEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
NNNNNNLIEGRISEFGSPYDGHRRHGDVNDAAHSHDMMEMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIMT
30 NPSNHQKKARHVKRTWGKRCNKLIFMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKADD
DTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDVE
IGKCLQNVNVLAGDSRDSNGRGRFFFPVPEHHLIPSHTDKKFWYWQYIFYKTDEGLDCCSDNAISFHYVSPNQMY
VLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSTETKTQ

SEQ ID NO: 47.

35 Human Core-1-Gal-T1 chaperone 1/COSMC

MLSESSSFLKGVMLGSIFCALITMLGHIRIGHGNRMHHHEHHHLQAPNKEDILKISEDERMELSKSFRVYCIILV
KPKDVSLSAAVKETWTKHCDKAEFFSSENVKVFESINMDTNDMWLMMRKAYKYAFDKYRDQYNWFFLARPTTFAI
IENLKYFLLKKDPSQPFYLGHITKSGDLEYVGMEGGIVLSVESMKRLNSLLNIPEKCPEQGGMIWKISEDKQLAV
CLKYAGVFAENAEDADGKDVFNKTSVGLSIKEAMTYHPNQVVEGCCSDMAVTFNGLTPNQMHVMMYGVYRLRAFG
40 HIFNDALVFLPPNGSDND

SEQ ID NO: 48.

Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-galactosyltransferase (Core-1-Gal-T1 Δ31)

GFCLAELFVYSTPERSEFMPYDGHHRGVDNDAHSHDMMMSGPEQDVGGHEHVHENSTIAERLYSEVRVLCWIM
 5 TNPSNHQKKARHVKRTWGKRCNKLI FMSSAKDDELDAVALPVGEGRNNLWGKTKEAYKYIYEHHINDADWFLKAD
 DDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFVVEALPNPKLCKSDNSGAEDV
 EIGKCLQNVNVLGDSRDSNGRGRFFFPVPEHHLIPSHTDKKFWYQYIFYKTDEGLDCCSDNAISFHYVSPNQM
 YVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSAETKTQ

10 SEQ ID NO: 49.

Maltose binding protein-tagged truncated Drosophila UDP-galactose:N-acetylgalactosamine-alpha-R beta-1,3-galactosyltransferase (MBP-Core-1-Gal-T1 Δ31) A383T mutant

MKIEEGKLVWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
 15 LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALS LIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
 NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIKKNHMNADTDYSIAEAAFNKGETAM
 TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
 GAVALKSYYYEELAKDPRIAATMENAQKEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNN
 NNNNNNLGIEGRISEFGSHMGFCLAELFVYSTPERSEFMPYDGHHRGVDNDAHSHDMMMSGPEQDVGGHEHVH
 20 ENSTIAERLYSEVRVLCWIMTNPSNHQKKARHVKRTWGKRCNKLI FMSSAKDDELDAVALPVGEGRNNLWGKTKE
 AYKYIYEHHINDADWFLKADDDTYTIVENMRYMLYPYSPETPVYFGCKFKPYVKQGYMSGGAGYVLSREAVRRFV
 VEALPNPKLCKSDNSGAEDVEIGKCLQNVNVLGDSRDSNGRGRFFFPVPEHHLIPSHTDKKFWYQYIFYKTDE
 GLDCCSDNAISFHYVSPNQMYVLDYLIYHLRPYGIINTPDALPNKLAVGELMPEIKEQATESTSDGVSKRSTETK
 TQ

25

SEQ ID NO: 50.

human Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1)

MVTLRKRTLKVLTLFLVLFIFLTSSFFLNYSHTMVATTWFPKQMVLELSENLRLIKHRPCTCTHCIGQRKLSAWFD
 ERFNQTMQPLLT AQNALLEDDTYRWLRLQREKKPNNLNDTIKELFRVVPGNVDPMLEKRSVGCRRCAVVGNSGN
 30 LRESSYGPEIDSHDFVLRMNKAPTAGFEADVGTKTTHHLVYPESFRELGDNVSMILVPFKTIDLEWVVSATITGT
 ISHTYIPVPAKIRVKQDKILIYHPAFIKYVFDNLQGHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHY
 WENNPSAGAFRKTGVHDADFESNVTATLASINKIRIFKGR

SEQ ID NO: 51.

35 porcine Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal1)

MAPMRKKSTLKLTLVLVLFIFLTSSFFLNYSHTVVTAWFPKQMVLELSENFKKLMKYPYRPCTCTRCIEEQRVSA
 WDERFNRSMQPLLTAKNAHLEEDTYKWWLRLQREKQPNNLNDTIRELFQVVPGNVDPLLEKRLVSCRRCAVVG
 SGNLKESYYGPQIDSHDFVLRMNKAPTEGFEADVGSKTTHHFVYPESFRELAQEVSMILVPFKTTDLEWVISATT
 TGTISHTYVVPVPAKIKVKEKILIYHPAFIKYVFDRLQGHGRYPSTGILSVIFSLHICDEVDLYGFGADSKGNW
 40 HHYWENNPSAGAFRKTGVHDGDFESNVTILASINKIRIFKGR

SEQ ID NO: 52.

mouse Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1)
 MRRKTLKYLTFLLFIFLTSEVLNYSNTGVPSAWFPKQMLLESENFRRFIKSQPCTCRHCISQDKVSYWFDQRF
 NKTMQPLLTVHNALMEEDTYRWWLRLQREKPNNLSDTVKELFRLVPGNVDPMLNKRLVGCRRCAVVGNSGNLKD
 5 SSYGPEIDSHDFVLRMNKAPTGVFEADVGSRTTHHLVYPESFRELGENVMVLVPFKTTDLQWVISATTTGTITH
 TYVPVPPKIKVKQEKILIYHPAFIKYVFDNWLQGHGRYPSTGILSIIFSIIHICDEVDLYGFGADSKGNWHHYWEN
 NPSAGAFRKTGVHDGDFEYNITTTLAANKIRIFKGR

SEQ ID NO: 53.

10 rat Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1)
 MVNMRKRTLKYLTFLLFIFLTSEVLNYSNSGVPSAWFPKQMVLEFSENFRKFIKSQPCTCRHCISQGVSYWFD
 QRFNKTMQPLLTVHNALMEEDTYRWWLRLQREKPNNLSDTVKELFRLVPGNVDPMLNKRLVGCRRCAVVGNSGN
 LKDSYSGPEIDSHDFVLRMNRAPTGVFEADVGSRTTHHLVYPESFRELGENVMVLVPFKITDLQWVISATTTGT
 ITHTYVPVPPKIKVKQEKILIYHPAFIKYVFDNWLQGHGRYPSTGILSVIFSIIHICDEVDLYGFGADSKGNWHHY
 15 WENNPSAGAFRKTGVHDGDFEYNVTTTAAINKIRIFKGR

SEQ ID NO: 54.

chimp Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1)
 MVTLRKRTLKVLTFVLVFIFLTSEFLNYSHTMVATTWFPKQMVLELSENKRLIKHRPCTCHCIGQRKLSAWFD
 20 ERFNQTVQPLLTAQNALLEDDTYRWWLRLQREKPNNLNDTIKELFRVVPGNVDPMLEKRSVGCRRCAVVGNSGN
 LRESSYSGPEIDSHDFVLRMNKAPTAGFEADVGTTHHLVYPESFRELGDNVSMILVPFKITDLEWVVSATTTGT
 ISHTYTPVLVKIRVKQDKILIYHPAFIKYVFDNWLQGHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHY
 WENNPSAGAFRKTGVHDADFESNVTATLAANKIRIFKGR

25 SEQ ID NO: 55.

Truncated pig Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1
 Δ45)
 ELSENFKKLMKYPYRPCTCTRCIEEQRVSAWFDERFNRSMPQLLTAKNAHLEEDTYKWWLRLQREKQPNNLNDTI
 RELFQVVPGNVDPLLEKRLVSCRRCAVVGNSGNLKESYYGPQIDSHDFVLRMNKAPTEGFEADVGSKTTHHFVYP
 30 ESFRELAQEVSMILVPFKTTDLEWVISATTTGTISHTYVPVPAKIKVKKEKILIYHPAFIKYVFDRLQGHGRYP
 STGILSVIFSLHICDEVDLYGFGADSKGNWHHYWENNPSAGAFRKTGVHDGDFESNVTTILASINKIRIFKGR

SEQ ID NO: 56.

35 Truncated pig Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (ST3Gal-1
 Δ56)
 YPYRPCTCTRCIEEQRVSAWFDERFNRSMPQLLTAKNAHLEEDTYKWWLRLQREKQPNNLNDTIRELFQVVPGNV
 DPLLEKRLVSCRRCAVVGNSGNLKESYYGPQIDSHDFVLRMNKAPTEGFEADVGSKTTHHFVYPESFRELAQEV
 MILVPFKTTDLEWVISATTTGTISHTYVPVPAKIKVKKEKILIYHPAFIKYVFDRLQGHGRYPSTGILSVIFSL
 HICDEVDLYGFGADSKGNWHHYWENNPSAGAFRKTGVHDGDFESNVTTILASINKIRIFKGR

40

SEQ ID NO: 57.

maltose binding protein-tagged truncated pig Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (MBP-ST3Gal-1 Δ45)

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
5 LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNKDLLPNPPKTWEEIIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKXKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSYYYEELAKDPRIAATMENAQKEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
NNNNNNLGI EGRISEFGSELSENFKKLMKYPYRPCTCTRCIEEQRVSAWFDERFNRSMQPLLTAHNAHLEEDTYK
10 WWLRLQREKQPNLNDTIRELFQVVPGNVDPLLEKRLVSCRRCAVVGNSGNLKESEYYPQIDSHDFVLRMNKAPT
EGFEADVGSKTTHHFVYPESFRELAQEVSMILVPFKTTDLEWVISATTTGRISHTYVVPVPAKIKVKKEKILYHP
AFIKYVFDRWLQGHGRYPSTGILSVIFSLHICDEVDLYGFGADSKGNWHHYWENNPSAGAFRKTGVHDGDFESNV
TTILASINKIRIFKGR

SEQ ID NO: 58.

maltose binding protein-starch binding domain-tagged truncated pig Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase 1 (MBP-SBD-ST3Gal-1 Δ45)

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNKDLLPNPPKTWEEIIPALDKELKAKGKSALMF
20 NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKXKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSYYYEELAKDPRIAATMENAQKEIMPNI PQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
NNNNNNLGI EGRISEFGSIVATGGTTTTATPTGSGSVTSTSKTTATASKTSTSTSSSTCTTPTAVAVTFDLTATT
TYGENIYLVGSISQLGDWETSDGIALSADKYTSSDPLWYVTVTL PAGESFEYKFIRIESDDSVESWESDPNREYTV
25 PQACGTSTATVTDTRWGSELSSENFKKLMKYPYRPCTCTRCIEEQRVSAWFDERFNRSMQPLLTAHNAHLEEDTYK
WWLRLQREKQPNLNDTIRELFQVVPGNVDPLLEKRLVSCRRCAVVGNSGNLKESEYYPQIDSHDFVLRMNKAPT
EGFEADVGSKTTHHFVYPESFRELAQEVSMILVPFKTTDLEWVISATTTGRISHTYVVPVPAKIKVKKEKILYHP
AFIKYVFDRWLQGHGRYPSTGILSVIFSLHICDEVDLYGFGADSKGNWHHYWENNPSAGAFRKTGVHDGDFESNV
TTILASINKIRIFKGR

SEQ ID NO: 59.

human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I (ST6GalNAc-1)

MRSCWLWRCRHLSQGVQWSLLLVFLFALPSFIKEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTI
35 YAEVPVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGR
TEAQSWKSQDTKTTQNGGQTRKLTASRTVSEKHQGAATTAKTLIPKSQHRMLAPTGA VSTRTRQKGVTTAVIP
PKEKKPQATPPPAPFQSPTTQRNQRLLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKFLPN
LTLFLDSRHFNQSEWDRLEHFAPFFGFMELNYSLVQKVVTREPPVPQQQLLLASLPAGSLRCITCAVVGNGGILN
NSHMGQEIDSHDYVFRLLSGALIKGYEQDVGTRTSFYGFTAFSLTQSLILGNRGFKNVPLGKDVRYLHFLEGRD
40 YEWLEALLMNQTVMSKNLFWFRHRPQEA FREALHMDRYLLHLPDFLRMKNRFLRSKTL DGAHWRIYRPTTGALL
LLTALQLCDQVSAYGFITEGHERFSDHYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRP GPGTAKAKN

SEQ ID NO: 60.

human N-acetylgalactosamine- α 2,6-sialyltransferase I (ST6GalNAcI), from pcDNA3.1(+) ST6GalNAcI-N1C1#1

5 MRSCLRWCRHLSQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTI
YAEFVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGR
TEAQSWKSQDQTKTKGNGGQTRKLTASRTVSEKHQGAATTAKTLIPKSQHRMLAPTGA VSTRTRQKGVTTAVIP
PKEKKPQATPPFPAPFQSPTTQRNQR LKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKLELPN
10 LTLFLDSRHFNQSEWDRLEHFAPPF GFMELNYSLVQKVVT RFPPVPQQQLLLASLPAGSLRCITCAVVGNGGILN
NSHIGQEIDSHDYVFR LSGALIKGYEQDVGTRTSFYGFTAFSLTQSLLILGNRGFKNVPLGKDVRYLHFLEGTRD
YEWLEALLMNQTVMSKNLFWFRHRPQEA FREALHMDRYLLLPDFLRYMKNRFLRSKTL DGAHWRIYRPTTGALL
LLTALQLCDQVSAYGFITEGHERFSDHYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN

SEQ ID NO: 61.

15 mouse α -N-acetylgalactosaminide α -2,6-sialyltransferase I
(ST6GalNAc-1)

MTRYCRGLSQRQAFLLLTVLALLFILLFVVKDPRAKDSRRQFILNNDSSAQEILQKAEPQGPIMTLSPRVHNKEA
TSVSSKDLKKQEREAVQGEQAEGKEKRKLETIRPAPENPQSKAEPAAKTPVSEHLDKVPRTPGALSTRKTPMATG
AVPAKKKVQATKSPASSPHPTTRRRQR LKASEFKSEPRWDFEEFYSLDMSSLQTNCSASVKIKASKSPWLQNI F
20 LPNITLFLDSGRFTQSEWNRLEHFAPPF GFMELNQSLVQKVVT RFPPVRQQQLLLASLPTGYSKCITCAVVGNGG
ILNDSRVGREIDSHDYVFR LSGAVIKGYEQDVGTRTSFYGFTAFSLTQSILILGRRGFQHVPLGKDVRYLHFLEG
TRNYEWLEAMFLNQTLAKTHLSWFRHRPQEA FRNALDLD RYLLLPDFLRYMKNRFLRSKTLDTAXWRIYRPTTG
ALLLLTALHLCDKVSAYGFITEGHERFSDHYDTSWKRLIFYINHDFRLERMVWKRLHDEGIIWLYQR PQSDKAK
N

25

SEQ ID NO: 62.

chicken α -N-acetylgalactosaminide α -2,6-sialyltransferase I
(ST6GalNAc-1)

MGFLIRRLPKDSRIFRWLLILT VFSFIITSFSALFGMEKSIFRQLKIYQSI AHMLQVDTQDQGSNYSANGRISK
30 VGLERDIAWLELNTAVSTPSGEGKEEQKKTVPVAKVEEAKEKVTVPFPPEVMGITNTTASTASVVERTKEKTTA
RPVPGVG EADGKRTTIALPSMKEDKEKATVKPSFGMKVAHANSTSKDKPKAEPPASVKAIRPVTQAATVTEKKK
LRAADFKE TPQWDFDDEYILDSSSPVSTCSSESVRAKAASDWLRDLFLPNITL FIDKSYFNVSEWDRLEHFAPPY
GFMELNYSLVEEVMSRLPPNPHQQLLLANSSSNVSTCISCAVVGNGGILNNSGMGQEIDSHDYVFRVSGAVIKGY
EKDVGT KTSFYGFTAYSLVSSLQNLGHKGFKKIPQGKHIRYIHFLEAVRDYEWL KALLLDKDIRKGFLNYYGRRP
35 RERFDEDF TMNKYLVAHPDFLRYLKNRFLKSKNLQKPYWRLYRPTTGALLLLTALHLCDRV SAYGYITEGHQKYS
DHYYDK EWKRLVFYVNHDFNLEKQVWKRLHDENIMKLYQRS

SEQ ID NO: 63.

rat alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I (ST6GalNAc-1)

5 MENRCRGLSQGQTFILLTGLMLLFILPSVVKEPSTRVSRQFIEDNESSLQGVQKPAPQGPVTLTPTVHNKKT
SVRTKWVELQKQDRATARGEREGVEKKLQAIRLAPENPKGKAEPEVKT PASKHLDKLPRATGALSTRKTQMATG
AAPAKKKVQPTPTPASFPHTTQRRQRLKASDFKSEPRWDFEEYSLDGGSLQTLPWFLKITVSHSPWVQNI
10 PNITLFLDSGRFNQSEWYRLEHFTPPFGFMELNQSLVQKVVSRRFPVPVQQQLLLASLPTRNLTCITCAVVGNGGI
LNNSRMGQEIDSHDYVFRSLSGTVIKGYEQDVGTRTSFYGFTAFLGQSILNLGSSRFSACAFGLKDVRYLHFLEG
TRDYEWLEAMFLNRTMANTKLYWFRHRPQEAFFREALDLDRYFLVHPDFLRVMKNRFLRSKTLDTAHWRLYRPTTG
ALLLLTALHLCDKVSAYGFITQGHERFSDHYDYTSWKRLIFYINHDFALERTVWKRLHDEGI IQLYQRP

SEQ ID NO: 64.

chimp alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
(ST6GalNAc-1)

15 MRSCLWRCRHL SQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSREPAPENNALNTQTQPKAHTAGDRGKEANQAP
PEEQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTOAQSWKSQDTKTTQGNNGQTRKLMASRTVSEK
HOGKAATTAKTLIPKSQHRMLAPTGA VSTRTRQKGVTTAVIPPEKEKKPQATPPPAPFQSPTTQRNQLKAA NFKS
EPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKLF LFNLTFLDSRHFNQSEWDRLEHFAPPF GFMELNYS
LVQKVVTREFPVPVQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGQEIDSHDYVFRSLSGVLIKGYEQDAVDRT
20 SFYGFTAFLTQSLILGNRGFKNVPLGKDVRYLHFLEGTRDYEWLEALLMNQTVMSKNLFWFRHRPQEAFFREAL
HMDRYLLLHPDFLRVMKNRFLRSKTLDGAWRIYRPTTGALLLLTALQLCDQVSAYGFITEGHERFSDHYDYTSW
KRLIFYTNHDFKLEREVWKRLHDEGIIRLYQRP GPGTAKAKN

SEQ ID NO: 65.

25 cow alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I (ST6GalNAc-1)

MATCPACGRLRSLLRPGSRKNEHQQNIKERSPELLQNATSQAPTPRRRATTHVESVQGTRTRDTHPKATTLTADQ
RRRVQTSKARAEPEGRVPTPIGKAAPQTQASKDTRADTLPPMAGGGGVASSRTEAPSLNSQNPRMTKGSGDRKAR
PTGPRAVPTKL RDS P SPATQRSQKLKATNFKSEPQWDFEEYSLLEVGLQTGTTRTRDTHPKATTLTADQRRRVQT
30 SKARAEPEGRVPTPIGKAAPQTQASKDTRADTLPPMAGGGGVASSRTEAPSLNSQNPRMTKGSGDRKDRAQPTRS
SAPLQSPATQRSQKLKATNFKSEPQWDFEDEYSLEVGLQTTCPDSVKIKASKSPWLRTFLFPNLTLFLDSGHFN
QSEWDRLEHFAPPF GFMELNFSLVQKVVARFPVPVQQQLLLASLPAGSSRCISCAVVGNGGILNNSHVGP EIDSH
DYVFRLSGAIKGYEHDVGTRTSFYGFTAFLTQSLILGSRGFPHAPLGQDVRYLHFLEGTRDYEWLEALLNR
TLTSRNLSWFRRRPQEAFFQEALQLDRYLLLHPDLLRYMKNRFLRSKTLNTAHWRIYRPTTGALLLLTALQLCDQV
35 SAYGFITEGHERFSDHYDYKSWKRTIFYTNHDFKLERALWKRLHDEGIIRLYQRPITSKPTI

SEQ ID NO: 66.

dog alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I (ST6GalNAc-1)

40 MKIRLCRLSHLDQTTQWFLLSAVLTFFVFILPSFVKEPNTKPFRRHQHIRNEERSPESLHEVTGQAPTAGDRTMAP
AVASAOEKS VWAPQALATAYAAQEDTRA EALEATSAQPTSMRGQERKDTMPDTLPRAQDKGVASDRTGLPSVMS

QDTRTIKGRRDQKEEPTATRKYVTPGPQSKARTTRTPVPESQAKALTPGAGPMGRTRKGATTAAAPHRDTARAT
 PPSTSSQGRTRRSPLRAANFKSEPRWDFEEQYSFDTEALQPTCPDSVKVKASKSPWLQNLFLANLTLFLDSSR
 FSQREWRLEHFAPPPGFMEELNQSLVQKGVTRFPVPQQQLLRASLPAGSSQCITCAVVGNGGILNDSRVGQEID
 GHDHVFRLSGAVTKGYEQDVGTRTSFYGFTAFSITQSLALGGRGFPHVPLGKDVRYLHFLEGTDRDYEWLEALLL
 5 NQTLVKSSLSWFRRRPQEAFRDTLQLDRYLLHPDFMRYMKNRFLRSKTLNNIHWRIYRPTTGALLLLTALHLCD
 QVSAYGFITEGHERFSDHYIDKSWKRTIFYTNHDFPLERTLWKRLHDEGIIRLYQRPIISKPKMVQTRPQAPPDS
 PGDTGTKEENPHCRRPLDLAIQRHPHFRAFDLSTPVLLSGSLFTQELWDSLSQHKAPYQWQGLSHQAINSTLSL
 LNGSESSQLFATSRKPLSSCIRCAVVGNGGILNGSRQGLNIDAHDYVFRQTLKTERGGRVPARRGPPGRMSPEVR
 RCGDGLDGDGSGRGQHSYFPDGNVSVQGHAAKIIVQPPPLTPALGLPPPSLSCLPDDLRYIFIPSDIRDYVM
 10 LRSAILGVPVPEGPDKGDRSPSHSPATSASYIGPVRGIPSYIAPYVPRFLKEPPFFQPPTVLVGGKPHVPTVLA
 ATFYTPFFRYAQPWPYRPHRKPSLSGARSSRGLVWVEPNLIWDNPYKGPGRSLDLELGLYPVSAYGFITSNYW
 KFSHYFERVKKPLIFYVNHDLSEASLWRDLHKAGILQLYQR

SEQ ID NO: 67.

15 zebrafish alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
 (ST6GalNAc-1)
 MDYFLWPSVLFLLAGIVAAFLVKNKPAHHQSSVVVEEEKKADAGVSSQEQHCEHWDESCQLKTVEKRLQTEDK
 TLEKAKDLNRDDLEEDALVQPTAREEEPGIRSALREYAFSPDAESKPLKYMAGMLRTCQLEKMMTKEELEEEQSK
 TIISSILSLPLGNTNLFKTQKHLNVERKVNITPIPVLYKKNFTKLPVWDFEDVYLDRDSNARKPTCPKSLHNTED
 20 PEFKESVLPDIQLWLYKGQLMSEWNRLAHFNNPFGFMEYNYNEIKRAVDLIPKRSSILLPVPKSGDKGCIRCA
 VVGAGGILNNSKMGREIDSHDYVFRVNGAVTKGYEEDVGNRTSVYVHTAFSLYATILTLKKYGFHNIPQDEGIKY
 VMIPEGLRDFEWLQGLLQGAANGSFKGVRPLNFFNGHFNFSRNVNLHPDFLRYIRNRFMPSKQMQGNYWAMYRP
 TNGAFALFLAIHTCDMVNAYGFITEDHHKYSNYYYEKFKKTSVIFYINHGYGLEIKTWKKLHDSGIIRLFQRH

25 SEQ ID NO: 68.

zebrafish 2 alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
 (ST6GalNAc-1)
 MFLLRIFLVTTFIASLPLFIFVTFYNGTSTLQIKWRQYYYWLNQGVAGSLLTIVEPEPTNSSALTDILANTTTM
 RLDPFETLSTPLPIMDKHKFTSLPHWKFDLLYRLDPHFKPSECATSLQNSSNPTFKKKFIPNIQLFLQSDHLNMS
 30 EWNRLYHFNPNFGYMGNYTAIKA AVETIHKPASSQLLQVHPGVKDGCCIRCAVVGASGILNGSKLGEIDSSDYV
 FRMNAAIAGHEEDVGKRTSVYVHTAHSIIQSLMIHKKRGFKQIPTDKDIKYVLIPEGPRDYNFLESIMKNRKIP
 SGAYRGRTPRKYYSGHFNESYIILHPDFLRYVRNRFLRAKQLKTKRWWVVRPTNGAFTLLAMHTCDIVRVYGF
 CTADYRKYPSSYYDTKHTKLVFYGNHDYRLEMKTWKKFHDDKLIWMYLGKSN

35 SEQ ID NO: 69.

xenopus alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
 (ST6GalNAc-1)
 MRPRWLRAVGITIGALTVCSLFYGHYYSSLVSVARSPPQGGVGYDLPSKGLEWIPAVPHGRERNRELAETLRVQRP
 SKVPGTCPTSLQLRVKNDYFKTIFNFEVPLLLWNSHLTENNDWTLNKRVPYGWKDLNREDAVASALKLLNDSAN
 40 KAMFERQGPQKCCIRCAVVGNGGILNGSKKGKEIDGHDYVFRNGAVIKGFEKDVGTSTFYGFTVNTMKNSLISY
 QEHGFTETPKGKDLHYIFIPSDLRDYVMLRSGILGVKVPSTGYDEGDKPSEYFGPKPSPKKFKMLHPDFLLYTRDR

FLKSDILNTEYASLYMPSTGGIMLLTALHSCDQVSAYGFITPDYNRYSDHYERQKVPLEVDANHDMLMEMQLWG
RLHDRGIKLYKR

SEQ ID NO: 70.

- 5 Takifugu alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
(ST6GalNAc-1)

SSGSLESMKLSDDQIKVDSRTTATPAYVKSSGPRQHTHRAREQEIQNVSRSTPRSRGANQTRIEVPVKKPQQPT
QETRRTDPPFIDGTYMSEEIPQTTCPDGIRTRRTNTEFSGNFLNNVPLQWARHATHEEYGRLSQHRGAHGWGG
VDYNTLVDAHSVLSANWQMLDDWKDRSNNSECIRCAVVGNGGILKDSKRGAEIDSHHYVFRTNGAVIKGFEKD
10 VGSRTTHYTFSTTLMNSIRAYAGVGFKAPPVSKETRYIFLPENDRDYLLVKAATHLVERGPQRNQKPAFFFG
KDVSAEKVKMYHPDFVRYLRNRFRLSATLTKYKDIYRSTGAVMLLVALHTCDKVSAYGFMTPDYMKYSDHYD
KKQKSVVFFKNHDLRMEMALWQRLHQAGLIQLYMRQ

SEQ ID NO: 71.

- 15 sea urchin alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
(ST6GalNAc-1)

MRLWRKLTSCKTRSRFVLLAFWYTMIALSVSLLYSLRGTHHSQDDSGTTSRRSRFYNSLFPFSRNDVADYATLTQ
HDVADDIVAPPHVGATGEIGPEKNVAIDDNKQIQHLIRRNSSGGIAGQSGNLIGDMLNDEEQRTFEVVEEEVAEQE
EIEVVNEFRDLNGKRPSGDDPRAPVTSTIDVTSIHFNATVKPTPPPEVQGYNVQTKDKLRMRCSCALVSSSG
20 HLVNTSAGAEIDSYPVLRMSAPVRGYEVDVGRRTTIRIMGHVNLKVLNASNELQDEILINSTTRAEKIIVPWL
YNVKVNQATDMYYKSARNLSSLYPHVEFYLLTPDKMKIAESLFQTETGLTRQETRTWLSTGWMNMLYAVDVCDKV
DIFGLVPENYCLKNPNSSVLYHYYESDGLKECDYITISEERLTSGHKFITEKAVFARWATKFNMVIPPTSLEPHG
RCPY

- 25 SEQ ID NO: 72.

Truncated human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase I
(ST6GalNAc-1 Δ35)

- KEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTIYAEPVPENNALNTQTQPKAHTTGDRGKEANQAPPE
EQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQNGGGQTRKLTASRTVSEKHQ
30 GKAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIPPKEKKPQATPPPAPFQSPTTQRNQLKAANFKSEP
RWDFEKYSFEIGGLQTTCPDSVKIKASKSLWLQKFLPNLTFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLV
QKVVTFRFPVPQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGEIDSHDYVFRLSGALIKGYEQDVGTRTSF
YGFTAFSLTQSLILGNRGFKNVPLGKDVRYLHFLEGRDYEWLEALLMNQTVMSKNLFWFRHRPQEAFFREALHM
DRYLLLHPDFLRYMKNRFLRSKTLTDGAHWRIYRPTTGALLLLTALQLCDQVSAYGFITEGHERFSDHYDTSWKR
35 LIFYNHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN

SEQ ID NO: 73.

maltose binding protein-tagged truncated human alpha-N-
acetylgalactosaminide alpha-2,6-sialyltransferase I (MBP-ST6GalNAc-1 Δ35)

- 40 MKIEEGKLVINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDREGGYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF

NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAM
 TINGPWAWSNIDTSKVNIGVTVLPFTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
 GAVALKSYYEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSSNNNN
 NNNNNNLGIEGRISEFGSKEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTIYAEPVPENNALNTQTQP
 5 KAHTTGDRGKEANQAPPEEQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGN
 GGQTRKLTASRTVSEKHQGAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIPPKEKKPQATPPPAPFQS
 PTTQRNQRLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKLEFLPNLTLFLDSRHFNQSEWDR
 LEHFAPPPFGFMELNYSVLQKVVTTRFPVPQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGQEIDSHDYVFR
 SGALIKGYEQDVGTRTSFYGFTAFSLTQSLILGNRGFKNVPLGKDVRYLHFLEGTRDYEWLEALLMNQTVMSKN
 10 LFWFRHRPQEAFFREALHMDRYLLHLPDFLRMKNRFLRSKTLDGAWRIYRPTTGALLLLTALQLCDQVSAYGFI
 TEGHERFSDHYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN

SEQ ID NO: 74.

Truncated chicken N-acetylgalactosamine-alpha2,6-sialyltransferase I
 15 (ST6GalNAc-1 Δ231)
 KTEPQWDFDDEYILDSSSPVSTCSESVRAKAASDWLRDLFLPNITLFDKSYFNVSEWDRLEHFAPPYGFMEIN
 YSLVEEVMSRLPPNPHQQQLLLANSSSNVSTCISCAVVGNGGILNNSGMGQEIDSHDYVFRVSGAVIKGYEKDVGT
 KTSFYGFTAYSLVSSLQNLGHKGFKKIPQKGHIRYIHFEAVRDYEWLKALLLDKDIRKGFLNYYGRRPRERFDE
 DFTMNKYLVAHPDFLRYLKNRFLKSKNLQKPYWRLYRPTTGALLLLTALHLCDRVSAYGYITEGHQKYS DHYYDK
 20 EWKRLVFYVNHDFNLEKQVWKRLHDENIMKLYQRS

SEQ ID NO: 75.

human Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase III (ST3Gal3)
 MGLLVFVRNLLALCLFLVLGFLYSAWKLHLLQWEEDSNSVVLSDSAGQTLGSEYDRLGFLNLD SKLPAELA
 25 TKYANFSEGACKPGYASALMTAIFPRFSKPAPMFLDDSFRRKWARIREFVPPFGIKGQDNLIKAILSVTKEYRLTP
 ALDSLRCRRCIIVGNGGVLANKSLGSRIDDYDIVRLNSAPVKGFEKDVGSKTTLRITYPEGAMQRPEQYERDSL
 FVLAGEFKWQDFKWLKYIVYKERSASDGFWKSVATRVKPEPPEIRILNPYFIQEAFTLIGLPFNNGLMGRGNIP
 TLGSAVVTMALHGCDEVAVAGFGYDMSTPNAPLHYYETVRMAAIKESWTHNIQREKEFLRKLVKARVITDLSSGI

30 SEQ ID NO: 76.

rat Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase III (ST3Gal3)
 MGLLVFVRNLLALCLFLVLGFLYSAWKLHLLQWEDSNSLILSLDSAGQTLGTEYDRLGFLNLD SKLPAELAT
 35 KYANFSEGACKPGYASAMMTAIFPRFSKPAPMFLDDSFRRKWARIREFVPPFGIKGQDNLIKAILSVTKEYRLTPA
 LDLSLHCRRCIIVGNGGVLANKSLGSRIDDYDIVIRLNSAPVKGFEKDVGSKTTLRITYPEGAMQRPEQYERDSL
 VLAGFKWQDFKWLKYIVYKERSASDGFWKSVATRVKPEPPEIRILNPYFIQEAFTLIGLPFNNGLMGRGNIPT
 LGSAVVTMALDGCDEVAVAGFGYDMSTPNAPLHYYETVRMAAIKESWTHNIQREKEFLRKLVKARVITDLSSGI

SEQ ID NO: 77.

Truncated rat Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase III (ST3Gal3 Δ72) V162I D311H mutant

5 LATKYANFSEGACKPGYASAMMTAIFPRFSKPAPMFLDDSFRRWARIREFVPPFGIKGQDNLIKAILSVTKEYRL
TPALDSLHCRRCCIIGNGGVLANKSLGSRIDDYDIVIRLNSAPVKGFEKDVGSKTTLRITYPEGAMQRPEQYERD
SLFVLGFKWQDFKWLKYIVYKERSASDGFWSVATRVKPEPPEIRILNPYFIQEAFTLIGLPFNGLMGRGN
IPTLGSVAVTMALHGCDEVAVAGFGYDMNTPNAPLHYETVRMAAIKESWTHNIQREKEFLRKLVKARVITDLSS
GI

10 SEQ ID NO: 78.

MBP-tagged truncated rat Gal beta-1,3-GalNAc alpha-2,3-sialyltransferase III (ST3Gal3 Δ72) V162I D311H mutant

15 MKIEEGKLVWINGDKGYNGLAEVGKKFEKDTGIKVTVHEHPDKLEEKFPQVAATGDGPDIIFWAHDREGGYAQSG
LLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMF
NLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVNAGAKAGLTFLVDLIKNNHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPL
GAVALKSYYEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNN
NNNNNNLGIEGRISEFGSLATKYANFSEGACKPGYASAMMTAIFPRFSKPAPMFLDDSFRRWARIREFVPPFGIK
20 GQDNLIKAILSVTKEYRLTPALDSLHCRRCCIIGNGGVLANKSLGSRIDDYDIVIRLNSAPVKGFEKDVGSKTTL
RITYPEGAMQRPEQYERDSLFVLGFKWQDFKWLKYIVYKERSASDGFWSVATRVKPEPPEIRILNPYFIQEA
AFTLIGLPFNGLMGRGNIPTLGSVAVTMALHGCDEVAVAGFGYDMNTPNAPLHYETVRMAAIKESWTHNIQRE
KEFLRKLVKARVITDLSSGI

SEQ ID NO: 79.

25 Rat Gale

MEEKVLVTGGAGYIGSHTVLELLEAGYSPVVIDNFHNSIRGEDSMPESLRRVQELTGRSVEFEEMDILDQAALQH
LFKKHNFKAVIDHFAGLKAVGESVQKPLDYRVNLTGTIQLLEIMRAHGVSLSVFSSSATVYGNPQYLPLDEAHPT
GGCTNPYGKSKFFIEEMIQDLCRADTAWNALLRYFNPIGAHASGRIGEDPQGIPNNLMPYVSQVAIGRREALNV
FGDDYATEDGTGVRDYIHVVDLAKGHIAALKKLKEQCGCRIYNLGTGTGYSVLQMVQAMEKASGKKIPYKVVAR
30 EGDVAACYANPSLAHEELGWTAAALGLDRMCEDLWRWQKQNPSTGFGAHG

SEQ ID NO: 80.

Campylobacter jejuni GNE

35 MKILISGGAGYIGSHTLRQFLKTDHEICVLDNLSKGSKIAIEDLQKTRAFKFFEQDLSDFQGVKALFEREKFDAI
VHFAASIEVFESMQNPLKYMNNTVNTNLIETCLQTGVNKFIFSSSTAATYGEPTPVVSETSP LAPINPYGRSK
LMSEEVLRDASMANPEFKHCILRYFNVAGACMDYTLGQRYPKATLLIKVAAECAAGKRDKLFI FGDDYDTKDGTC
IRDFIHVDDISSAHLAALDYLKENESNVFNVGYGHGFSVKEVIEAMKKVSGVDFKVELAPRRAGDPSVLISDASK
IRNLTSWQPKYDDLELICKSAFDWEKQC

SEQ ID NO:81.

Human G-CSF 174 aa form

MTPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLLGHSLGIPWAPLSSCPSQALQLAGC
LSQLHSGFLFYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAG
5 GVLVASHLQSFLEVSYRVLRHLAQP

WHAT IS CLAIMED IS:

- 1 1. A method of producing an O-glycosylated soluble therapeutic protein
2 in a prokaryotic microorganism,
3 the method comprising the steps of:
4 a) expressing a soluble therapeutic protein in the prokaryotic microorganism,
5 b) expressing a heterologous soluble active nucleotide sugar : polypeptide
6 glycosyltransferase protein in the prokaryotic microorganism, and
7 c) growing the microorganism under conditions that allow intracellular
8 transfer of a first sugar moiety from a first donor substrate to an amino acid acceptor substrate
9 on the therapeutic protein by the heterologous soluble active nucleotide sugar : polypeptide
10 glycosyltransferase protein, thereby producing the O-glycosylated soluble therapeutic protein.
- 1 2. The method of claim 1, wherein the prokaryotic microorganism has an
2 intracellular oxidizing environment.
- 1 3. The method of claim 2, wherein the prokaryotic microorganism is
2 genetically modified to have the intracellular oxidizing environment.
- 1 4. The method of claim 1, wherein the prokaryotic microorganism is an
2 *E. coli* or a *Pseudomonas* bacterium.
- 1 5. The method of claim 3, wherein the microorganism is *E. coli*.
- 1 6. The method of claim 5, wherein the prokaryotic microorganism has a
2 mutation in an endogenous reductase nucleic acid.
- 1 7. The method of claim 1, wherein the heterologous soluble active
2 nucleotide sugar : polypeptide glycosyltransferase is a soluble active eukaryotic GalNAcT
3 protein.
- 1 8. The method of claim 1, further comprising the step of
2 d) expressing a first heterologous soluble active glycosyltransferase in the
3 prokaryotic microorganism and allowing intracellular transfer of a second sugar moiety from
4 a second donor substrate to a first acceptor substrate on the therapeutic protein by the first
5 heterologous soluble active glycosyltransferase protein, thereby producing the O-
6 glycosylated soluble therapeutic protein.

1 9. The method of claim 8, wherein the first heterologous soluble active
2 glycosyltransferase protein is selected from a eukaryotic core I galactosyltransferase (Core 1
3 GalT1) protein and a ST6 GalNAc 1 protein.

1 10. The method of claim 9, further comprising the step of
2 e) expressing a second heterologous soluble active glycosyltransferase in the
3 prokaryotic microorganism, and allowing intracellular transfer of a third sugar moiety from a
4 third donor substrate to a second acceptor substrate on the therapeutic protein by the second
5 soluble active glycosyltransferase protein, thereby producing the O-glycosylated soluble
6 therapeutic protein.

1 11. The method of claim 10, wherein the second heterologous soluble
2 active glycosyltransferase protein is a sialyltransferase selected from the group consisting of a
3 eukaryotic $\alpha(2,3)$ sialyltransferase (ST3Gal1) protein, and a bacterial $\alpha(2,3)$ sialyltransferase
4 protein.

1 12. The method of claim 1, wherein the microorganism is grown in a
2 medium that comprises a precursor of the donor substrate.

1 13. The method of claim 1, wherein the microorganism is genetically
2 modified to enhance production of the donor substrate.

1 14. The method of claim 13, wherein the microorganism is *E. coli*.

1 15. The method of claim 1, wherein the prokaryotic microorganism is
2 grown at a temperature lower than an optimal growth temperature.

1 16. The method of claim 1, further comprising the step of isolating the O-
2 glycosylated soluble therapeutic protein from the prokaryotic microorganism.

1 17. The method of claim 1, wherein the O-glycosylated soluble therapeutic
2 protein is produced on a commercial scale.

1 18. The method of claim 1, wherein the prokaryotic microorganism
2 expresses an accessory enzyme, and wherein the accessory enzyme is selected from the group
3 consistin of a UDP-glucose 4' epimerase protein, a UDP-GlcNAc 4' epimerase protein or a
4 dual function UDP-glucose 4' epimerase protein/ UDP-GlcNAc 4' epimerase protein.

1 19. The method of claim 1, further comprising the step of modifying the
2 O-glycosylated therapeutic protein by addition of a PEG moiety.

1 20. A prokaryotic microorganism comprising an O-glycosylated soluble
2 therapeutic protein and a heterologous soluble active nucleotide sugar : polypeptide
3 glycosyltransferase protein,
4 wherein the heterologous soluble active nucleotide sugar : polypeptide
5 glycosyltransferase protein transfers a first sugar moiety from a first donor substrate to an
6 amino acid acceptor substrate on the O-glycosylated soluble therapeutic protein inside the
7 prokaryotic microorganism.

1 21. The prokaryotic microorganism of claim 20, wherein the prokaryotic
2 microorganism has an intracellular oxidizing environment.

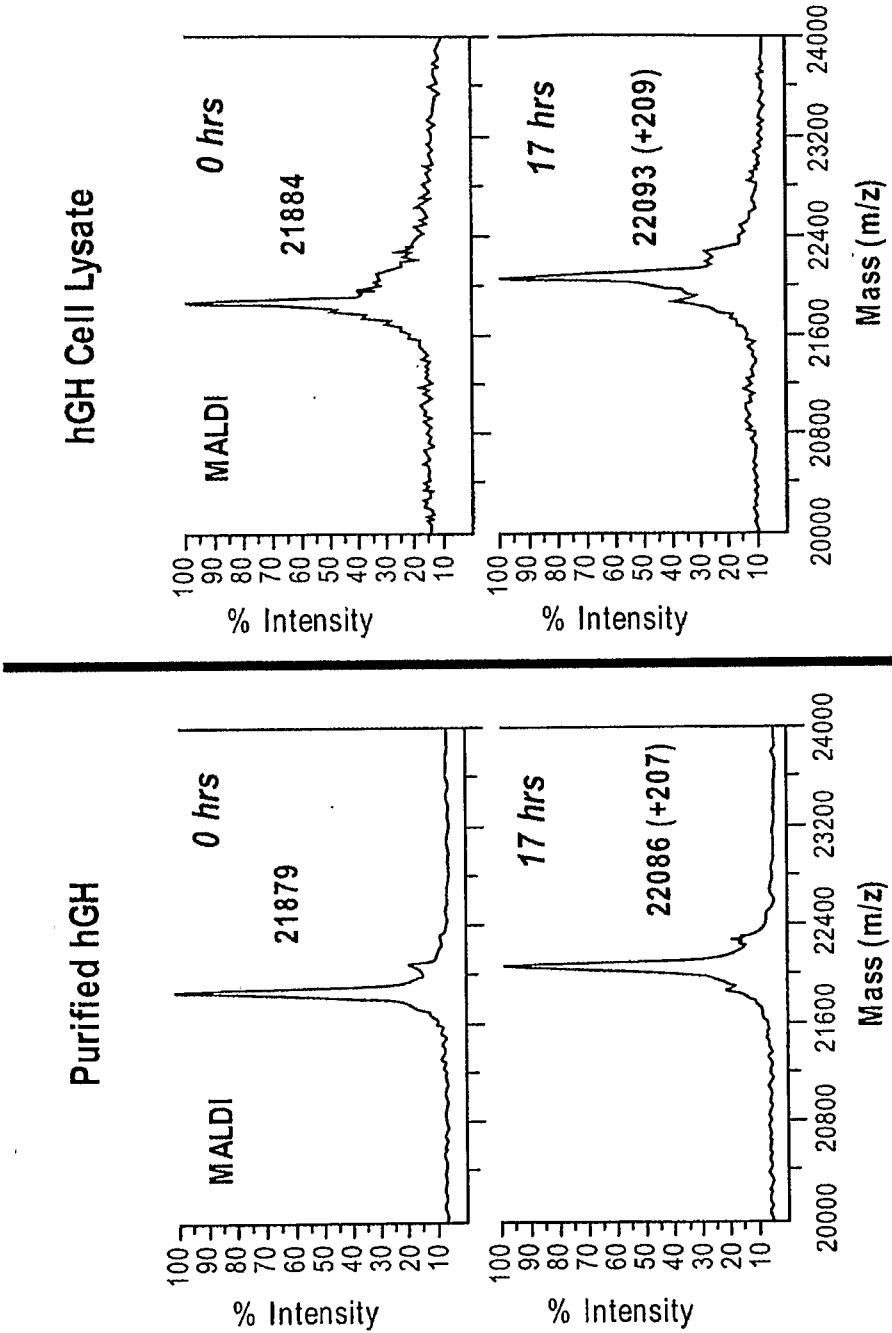


FIG. 1

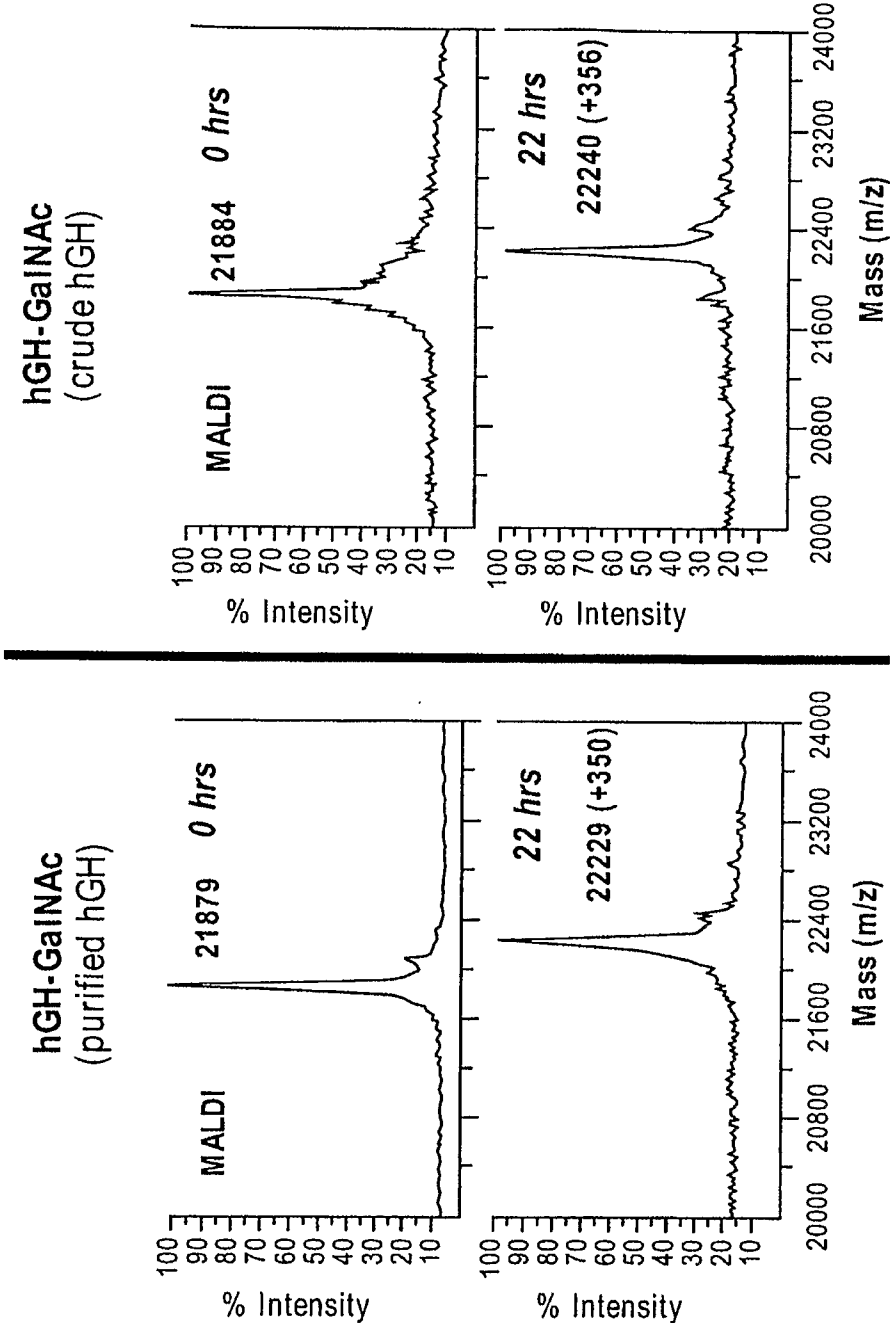


FIG. 2

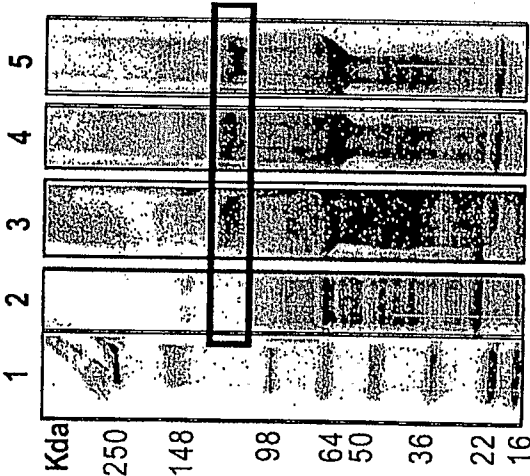


FIG. 3A

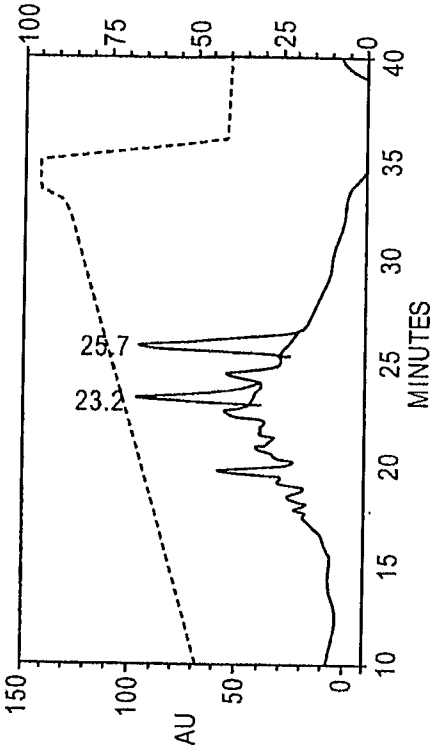


FIG. 3B

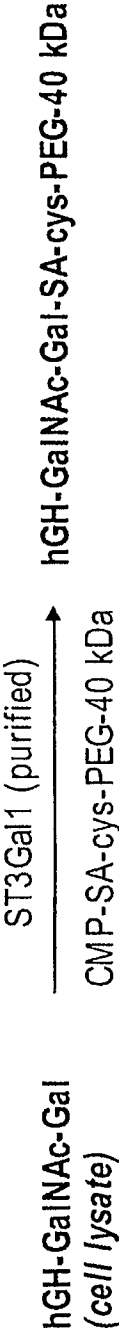


FIG. 4A

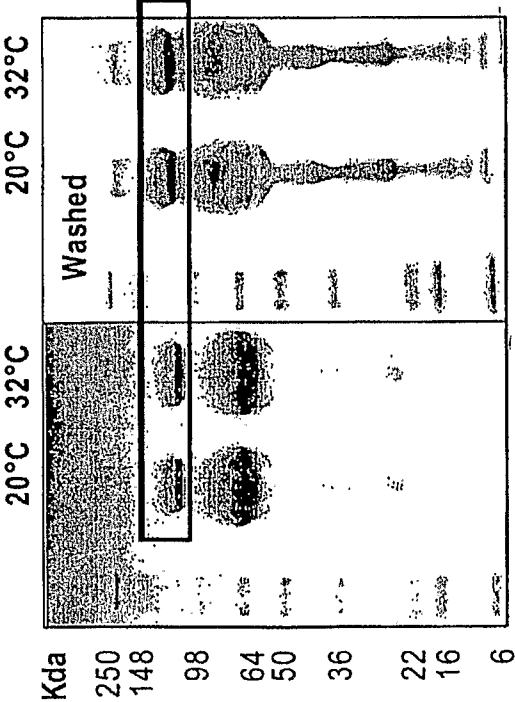


FIG. 4B

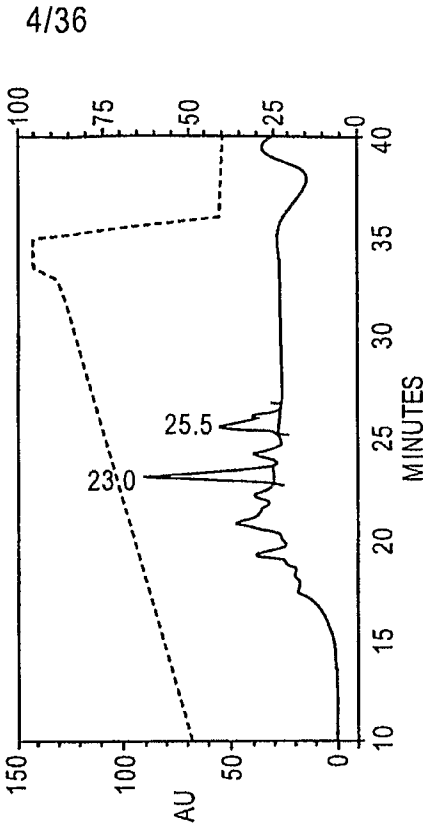
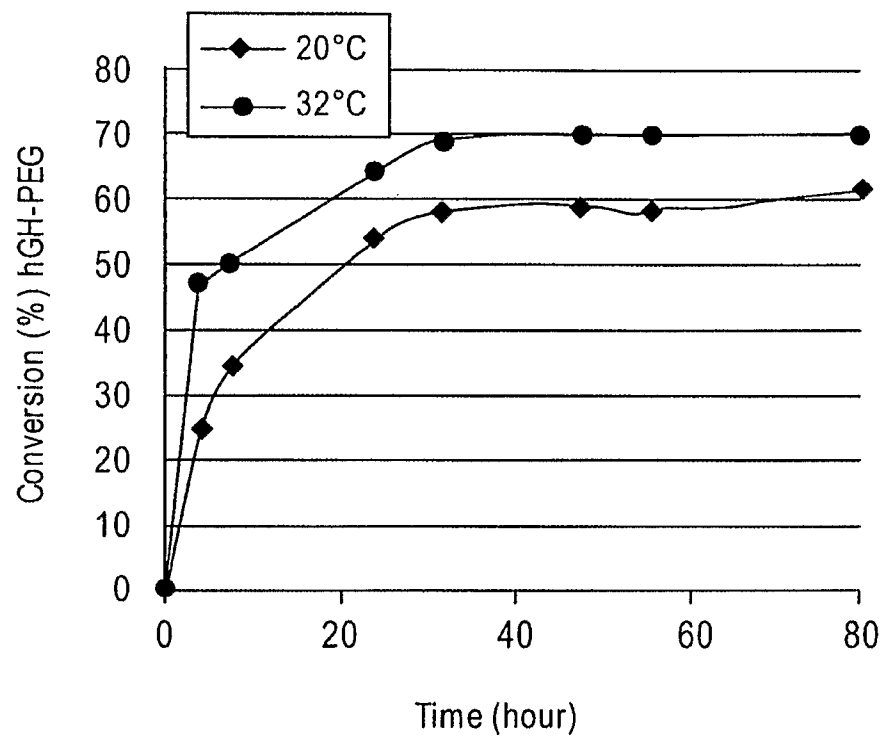


FIG. 4C

5/36

**FIG. 4D**

6/36

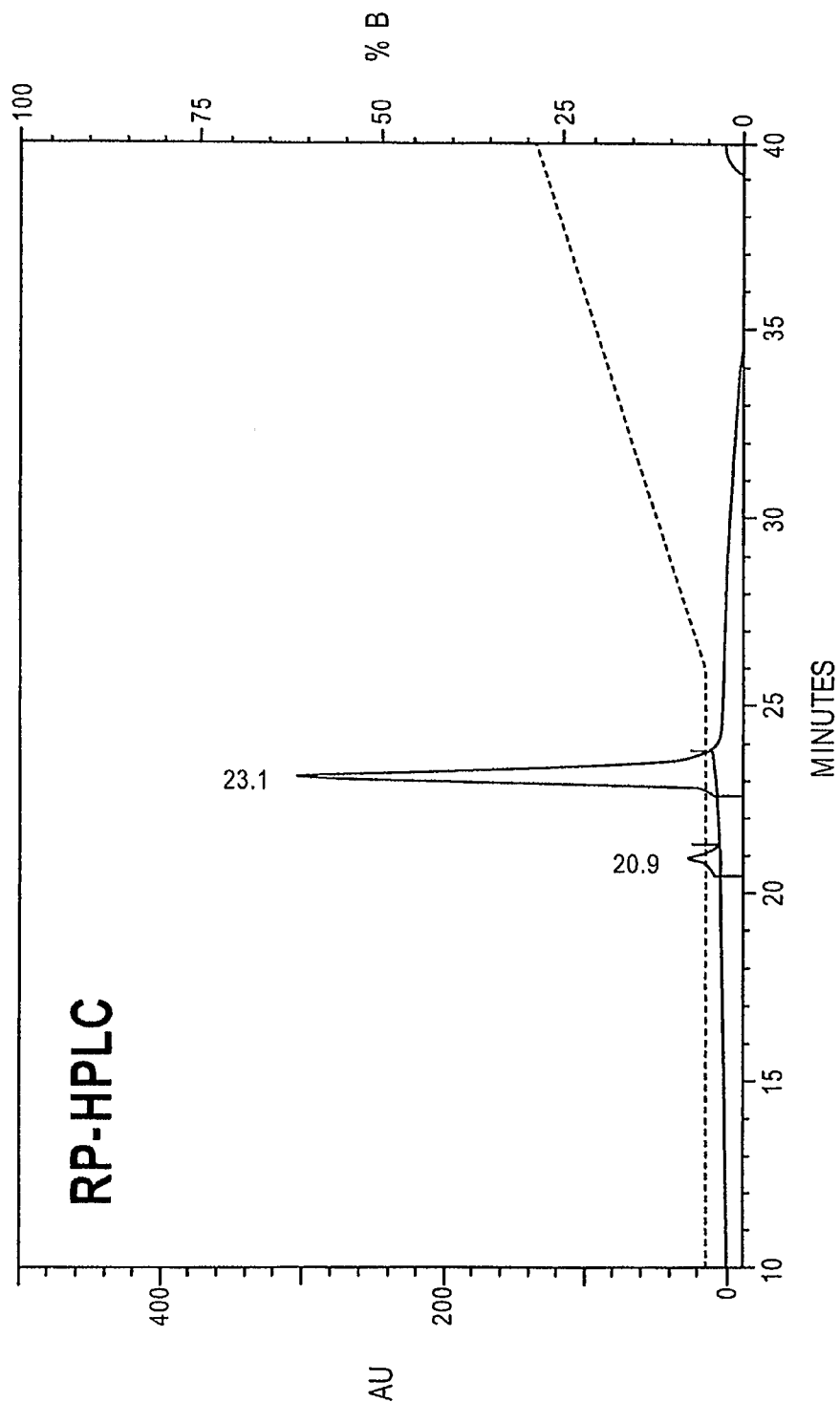


FIG. 5

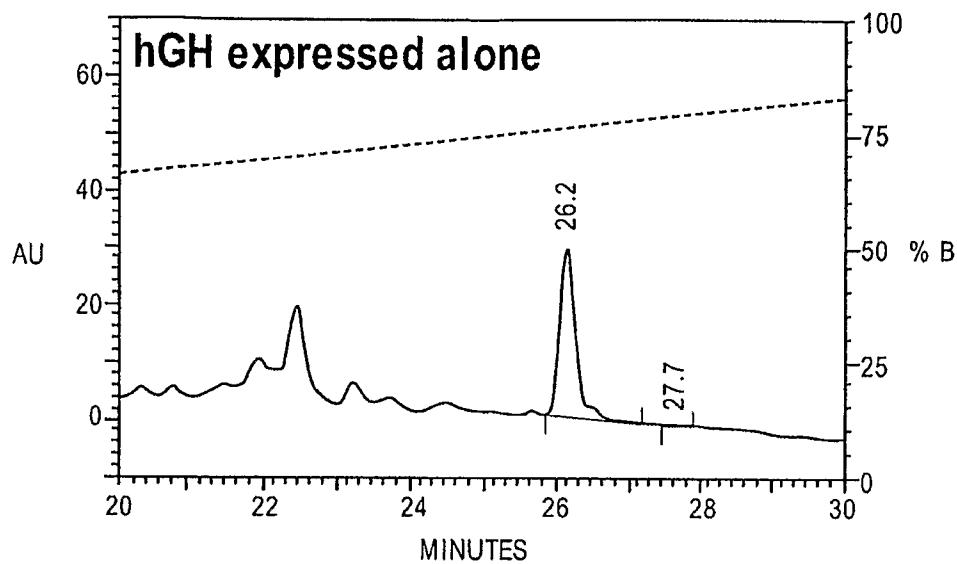
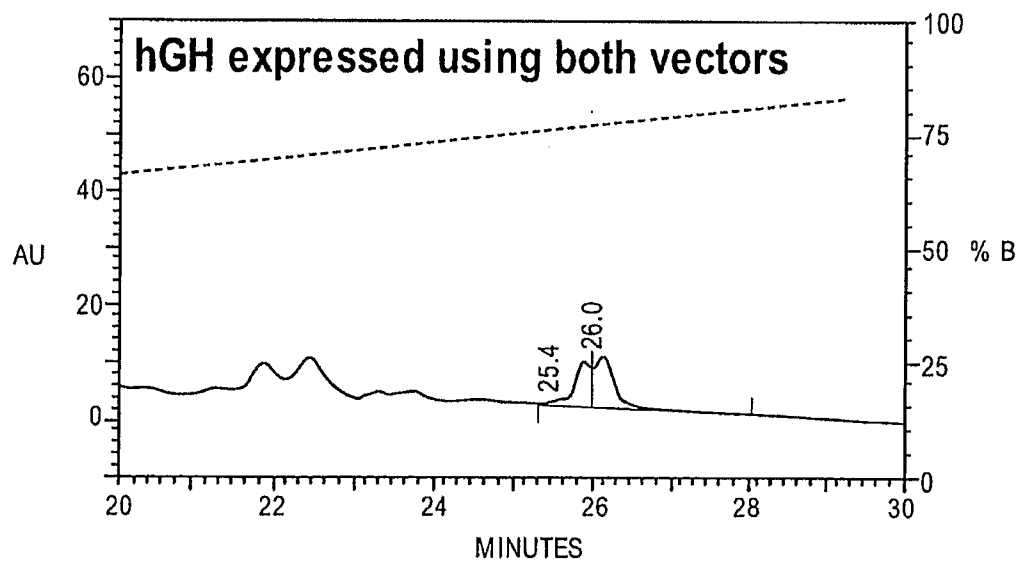
7/36

Combined Expression of Enzymes in a Single Cell

hGH (cell lysate)	trxB gor supp <i>E. coli</i>
GalNAc-T2 (cell lysate) GalE/GNE (cell lysate) Core-1-GalT1 (cell lysate) ST3Gal1 (cell lysate)	trxB gor supp <i>E. coli</i>
CMP-SA-Cys-PEG-40 kDa UDP-GlcNAc UDP-Glucose	

**hGH-GalNAc-Gal-SA-Cys-PEG-40 kDa****FIG. 6**

8/36

**FIG. 7A****FIG. 7B**

9/36

Epimerase	GalNAcT2	Total hGH	Conversion Yield	hGH-GalNAc
GalE	GalNAcT2 Δ51 Δ445	9.6 mg/L	38.5%	3.7 mg/L
		10.2 mg/L	41.2%	4.2 mg/L
GNE		13.5 mg/L	38.5%	5.2 mg/L
		9.3 mg/L	31.2%	2.9 mg/L

FIG. 8

10/36

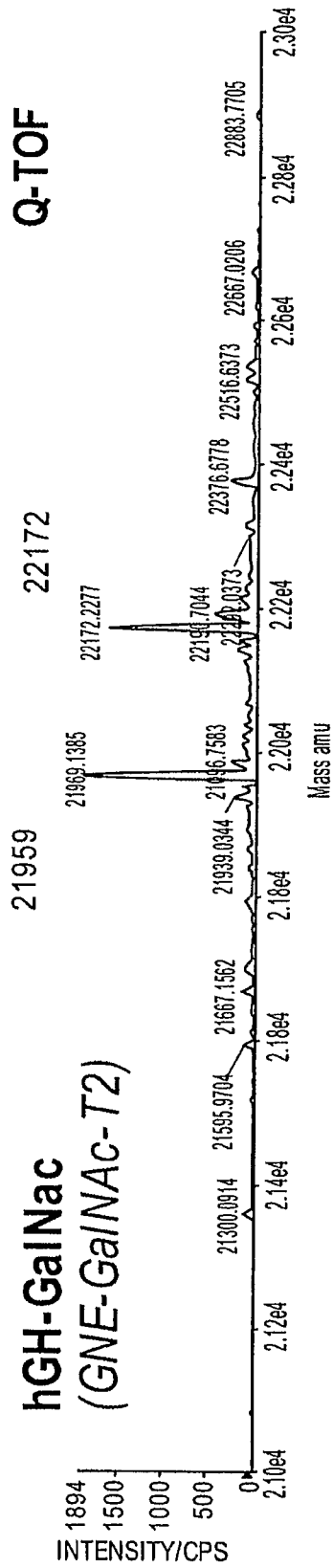


FIG. 9A

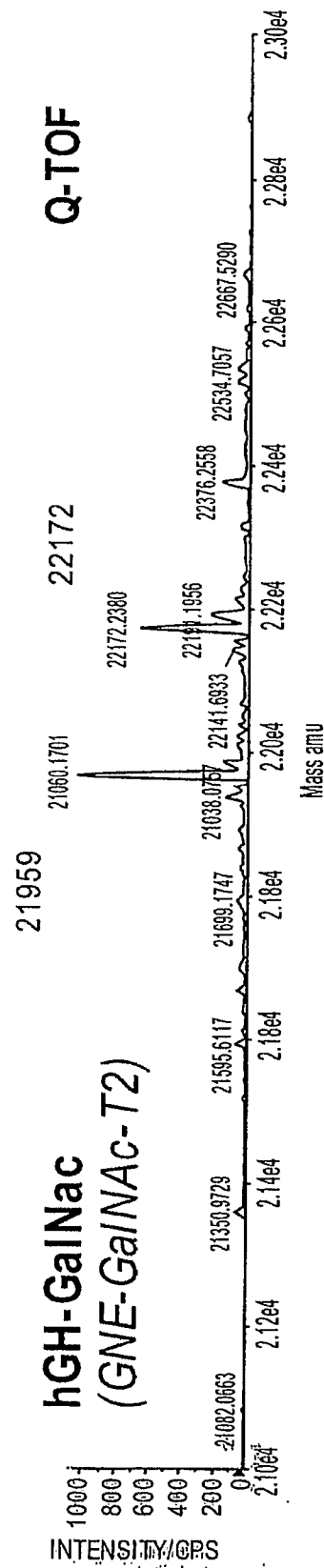
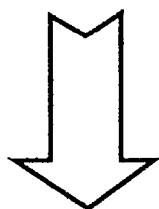
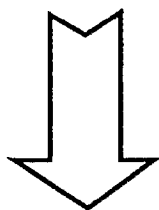
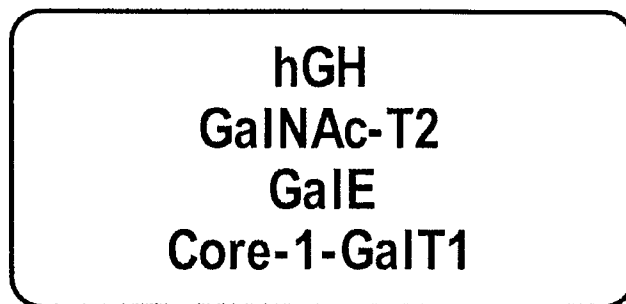


FIG. 9B

11/36

E. coli

MBP-ST3Gal1

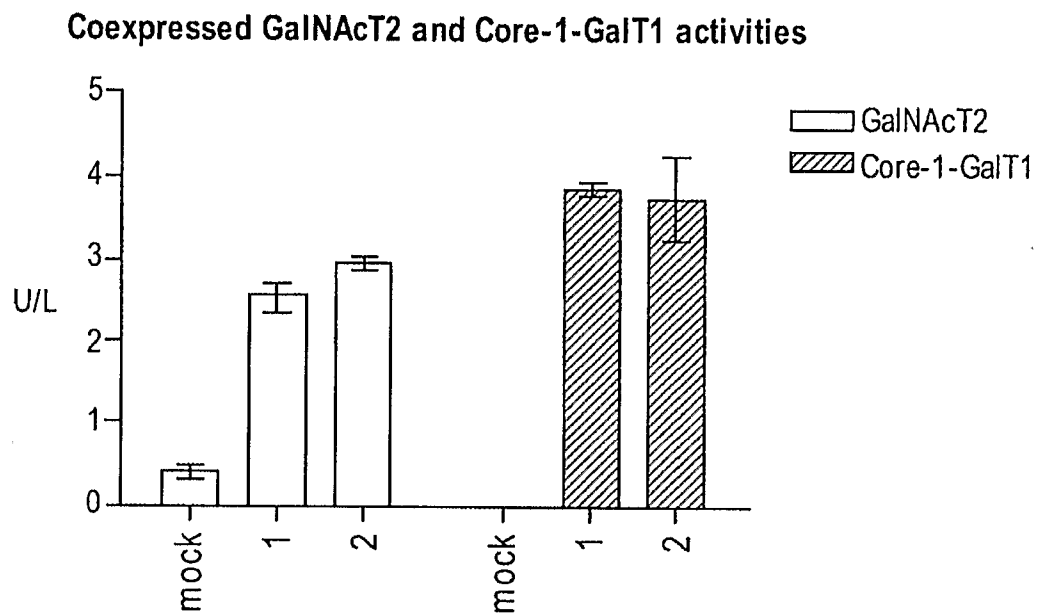
FIG. 10

12/36

	hGH mg/L	hGH-GalNAc	hGH-GalNAc mg/L
Cassette 1	59.2	30.4 %	18.0
Cassette 2	49.5	30.1 %	14.9

FIG. 11

13/36

**FIG. 12**

14/36

Increased yield of glycosylated hGH by addition of donor nucleotides

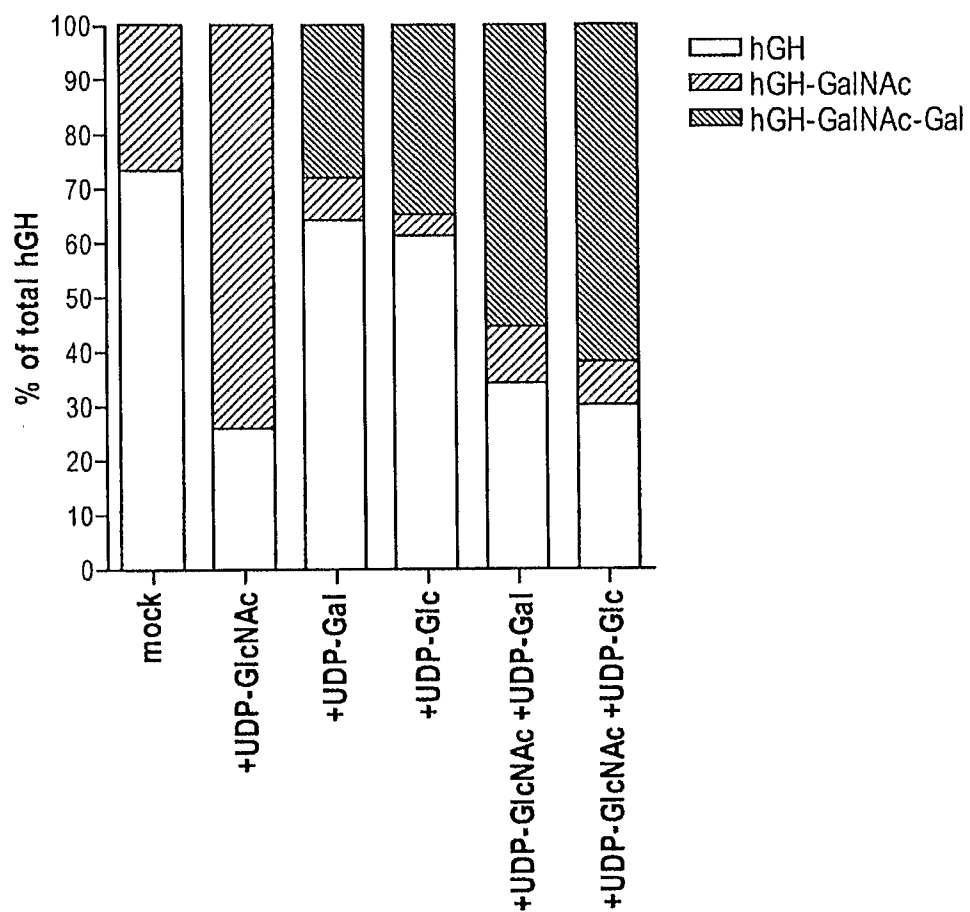


FIG. 13

15/36

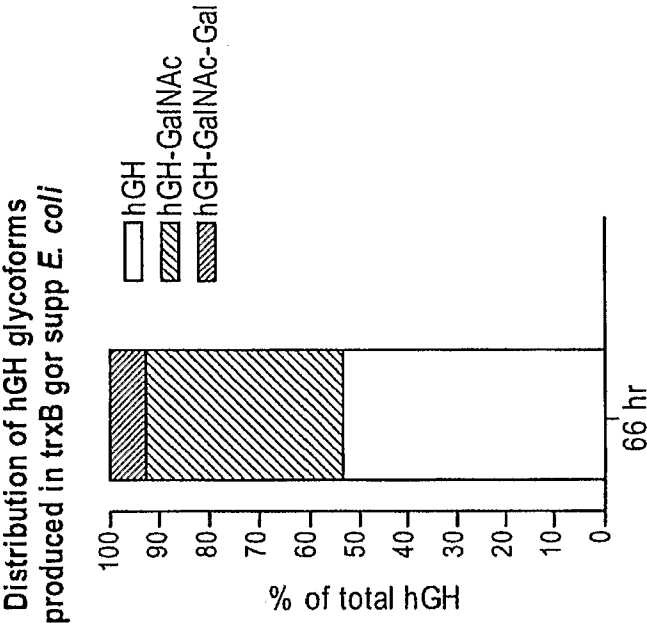


FIG. 14B

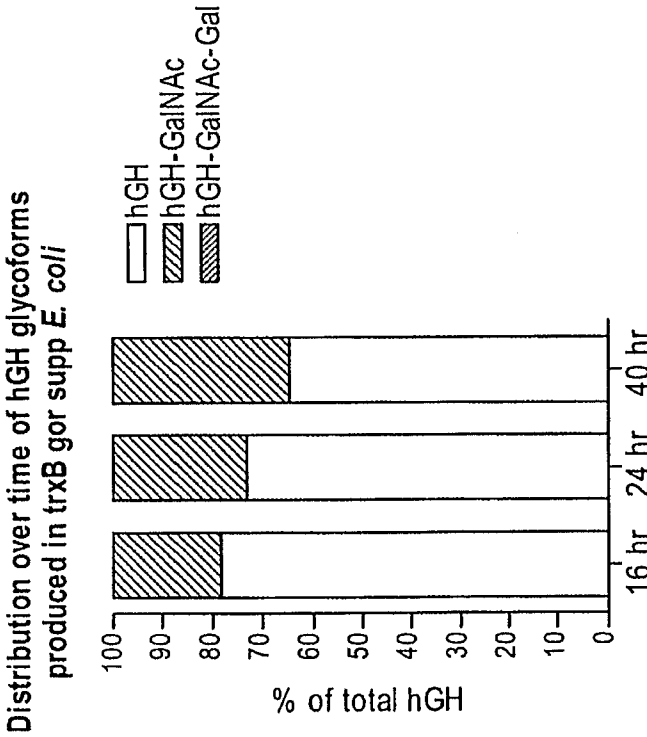
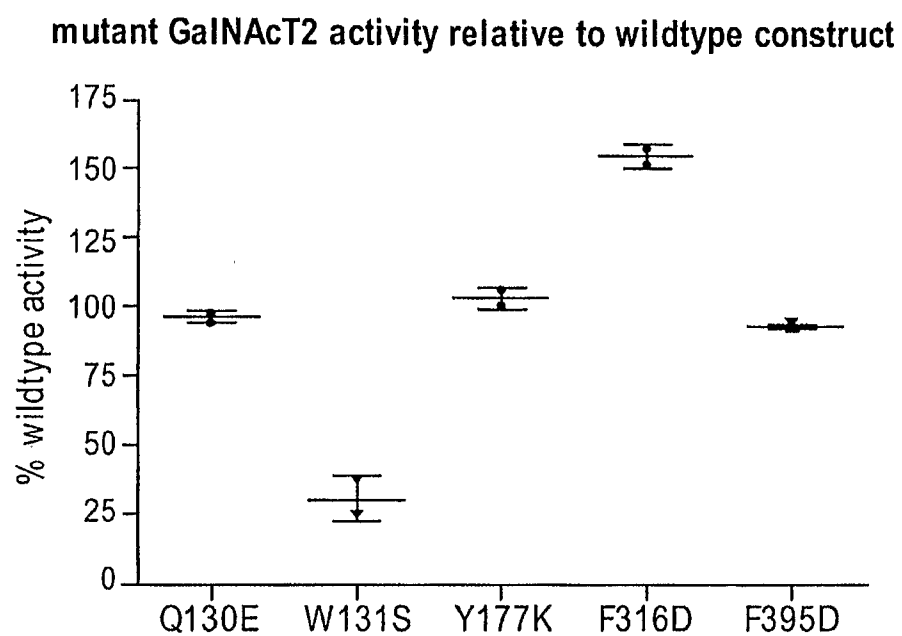
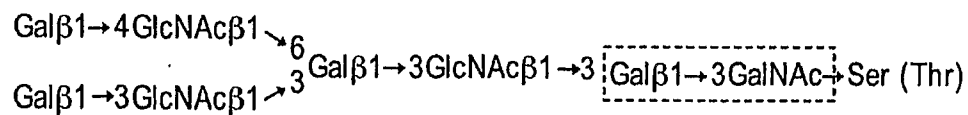
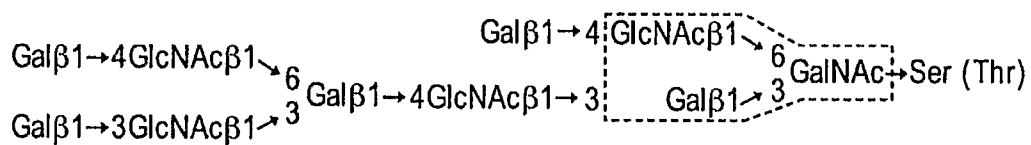
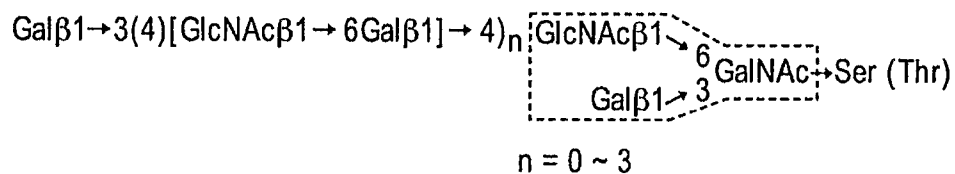
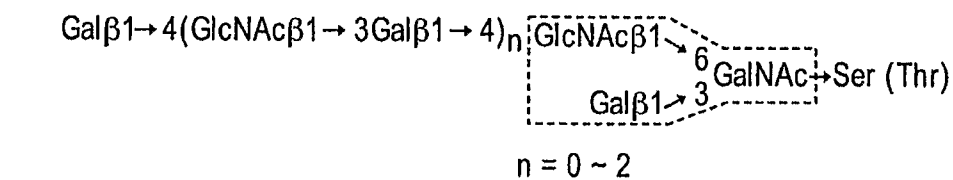
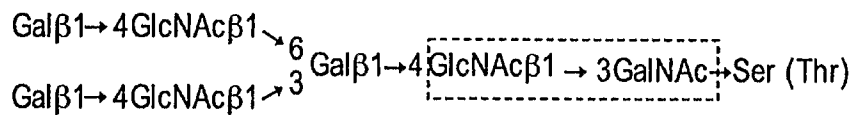
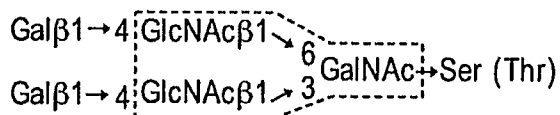


FIG. 14A

16/36

**FIG. 15**

17/36

Core 1*Core 2**Core 3**Core 4***FIG. 16**

18/36

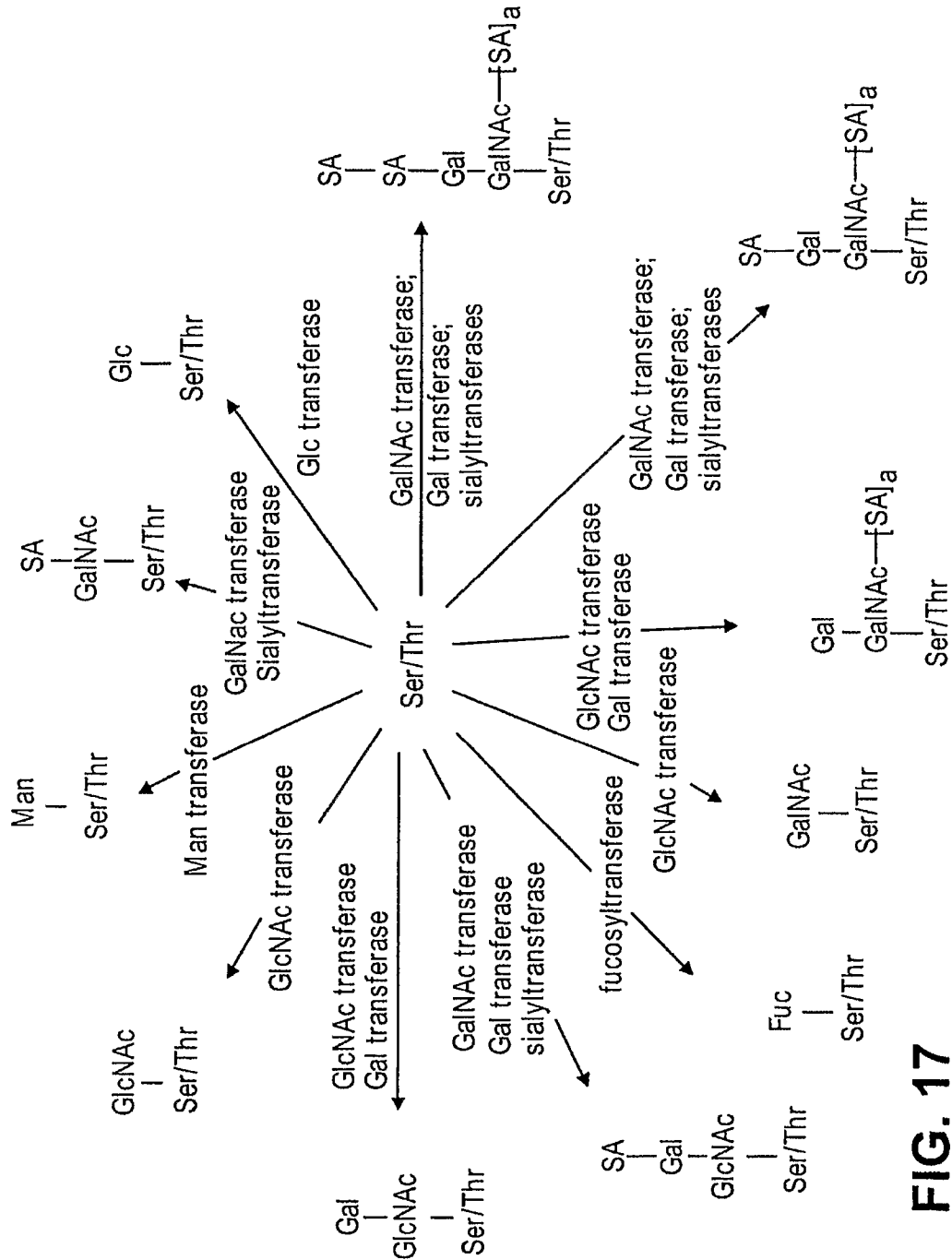


FIG. 17

Glycosyl Transferase Family 27				
BACTERIA	Protein	Organism	GenBank/GenPept	UniPort
CHY_2609		Carboxydothermus hydrogenoformans Z-2901	CP000141	ABB14482.1
CHY_2608		Carboxydothermus hydrogenoformans Z-2901	CP000141	ABB14167.1
EUKARYOTA	Unknown(protein for MGC:133911)	Bos taurus	BC111260	AAI11261.1
	GALNT6	Bos taurus	BT020728	AAI08745.1
	Polypeptide GalNAc transferase T1	Bos taurus	L07780	AAA30532.1
			L17437	AAA68489.1
				AAE24810
				AAE24816
				NP_803485.1
ZK688.8 (gly3)		Caenorhabditis elegans	NM_177519	NP_803485.1
			L16621	AAI28224.3
			AF031833	AAC13669.1
Y46H3A.6 (gly-7)		Caenorhabditis elegans	NM_066321	NP_498722.1
			AF031841	AAC13677.1
			AC006774	AAF60620.1
Y39E4B.12 (gly-5)		Caenorhabditis elegans	NM_071111	NP_503512.1
			AF031835	AAC13671.1
			AF031836	AAC13672.1
			AF031837	AAC13673.1
			AL110487	CAB54435.1
			AL110487	CAC42368.1
			AL110487	CAC42369.1
			NM_067320	NP_499721.3
			NM_067321	NP_499722.2
			NM_171894	NP_741277.1
Y75B8A.9		Caenorhabditis elegans	AL033514	CAI22098.1
			AL033514	CAE47472.1
			NM_067181	NP_499582.1
Y66A7A.6 (gly-8)		Caenorhabditis elegans	AF031842	AAC13678.1
			AL032622	CAAI1500.1
			AL590342	CAC35860.1
Y45F10D.3		Caenorhabditis elegans	NM_067103	NP_499504.1
			AL021492	CAI16378.1

FIG. 18

20/36

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
H38K22.5 (gly-6)	Caenorhabditis elegans	AL021492	CAJ76949.1	
		NM_070255	NP_502656.1	
		AF031838	AAC13674.1	O61394
		AF031839	AAC13675.1	O61395
		AF031840	AAC13676.1	O61396
		AL024499	CAA19707.1	
		AL024499	CAC42317.1	
		AL024499	CAC42318.1	
		NM_065469	NP_497870.2	
		NM_065470	NP_497871.2	
Y116F11B.12 (gly-4)	Caenorhabditis elegans	NM_065471	NP_497872.2	
		AF031834	AAC13670.1	O61390
		AL132943	CAB81983.1	Q81136
		AL132943	CAB81985.3	
		AL132943	CAC14394.1	
		NM_075449	NP_507850.1	
		AF031843	AAC13679.1	O61398
		AL117202	CAB57897.1	Q9U2C4
		AL117202	CAI46621.1	
		NM_067050	NP_499451.1	Q7YYJ5
GLY9 (Y47D3A.23)	Cryptosporidium parvum	AY466374	AAE23115.1	
		BX538351	CAD98487.1	
		AY424893	AAQ95174.1	
		AY424894	AAQ95175.1	
		AY424895	AAQ95176.1	
		BC095642	AAH95642.1	
		CR626885	CAI11605.1	
		BX323558	CAI11790.1	
		BC090692	AAH90692.1	
		BC067340	AAH67340.1	
Polypeptide GalNAc transferase T2 (ppGalNAc-T2; 1MB.187)	Danio rerio	CR626885	CAI20718.1	
		CH211-147C14.1-001		
		CH211-214J24.2-001 (fragment)		
		Unknown (protein for MGC:113219)		
		MGC:77836		
		CH211-147C1.1-001		
		Unknown (protein for MGC:112011)		
		CH211-147C14.1-001		
		CH211-214J24.2-001 (fragment)		
		Unknown (protein for MGC:113219)		
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T1	Cryptosporidium parvum	AY466374	AAE23115.1	
		BX538351	CAD98487.1	
		AY424893	AAQ95174.1	
		AY424894	AAQ95175.1	
		AY424895	AAQ95176.1	
		BC095642	AAH95642.1	
		CR626885	CAI11605.1	
		BX323558	CAI11790.1	
		BC090692	AAH90692.1	
		BC067340	AAH67340.1	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T3	Cryptosporidium parvum	AY466374	AAE23115.1	
		BX538351	CAD98487.1	
		AY424893	AAQ95174.1	
		AY424894	AAQ95175.1	
		AY424895	AAQ95176.1	
		BC095642	AAH95642.1	
		CR626885	CAI11605.1	
		BX323558	CAI11790.1	
		BC090692	AAH90692.1	
		BC067340	AAH67340.1	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T4 (fragment)	Cryptosporidium parvum	AY466374	AAE23115.1	
		BX538351	CAD98487.1	
		AY424893	AAQ95174.1	
		AY424894	AAQ95175.1	
		AY424895	AAQ95176.1	
		BC095642	AAH95642.1	
		CR626885	CAI11605.1	
		BX323558	CAI11790.1	
		BC090692	AAH90692.1	
		BC067340	AAH67340.1	

FIG. 18 (cont.)

21/36

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
CH211-214J24.3-001 (fragment) CG7579	Danio rerio Drosophila melanogaster	BX323558 AE003530 AF326979 NM_140542 AF145655 AE003842 NM_136412 AE003806 AE003806 AY121661 NM_166188 NM_166189	Q8VUT5 Q8IA41	
polypeptide N-acetylgalactosaminyltransferase T3 (CG4445)	Drosophila melanogaster	AF145655 AE003842	Q9Y117	
CG30463	Drosophila melanogaster	NM_136412 AE003806 AE003806 AY121661 NM_166188 NM_166189	Q9V7T0 Q8MRC9	
polypeptide N-acetylgalactosaminyltransferase T6 (CG2103)	Drosophila melanogaster	AE003476 AE003476 AY061629 AY268067 NM_139492 NM_167966	Q95R40 Q9VZX5	
polypeptide N-acetylgalactosaminyltransferase T5 (CG31651)	Drosophila melanogaster	AE003608 AY060338 AY268066 NM_135062 AE003579 AF324752 AY268065 NM_164539	Q9VMU3 Q95T43	
polypeptide N-acetylgalactosaminyltransferase T4 (CG31956)	Drosophila melanogaster	AE003579 AF324752 AY268065 NM_164539 AE003530 AF324750 AE003530 NM_140541	Q8IQ11 Q8IA42	
CG7304	Drosophila melanogaster	AE003530 AF324750 AE003530 NM_140541	Q9VUT4 Q8IA44	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
polypeptide N-acetylgalactosaminyltransferase T8 (CG7297)	Drosophila melanogaster	AE003530	Q9VUT6	AAF49587.1
		AY070966		AAL48588.1
		NM_140543		NP_648800.1
polypeptide N-acetylgalactosaminyltransferase T1 (CG8182)	Drosophila melanogaster	AE003810	Q9V7C8	AAF58130.1
		AF218236		AAG13184.1
		AY113411		AAM29416.1
		AE003810		AAM70974.1
		AY268063		AAQ56699.1
		NM_137199		NP_611043.1
		NM_166099		NP_725472.1
		AE003509		AAF48851.1
		AF493067		AAM62412.1
		AE003509		AAN09470.1
polypeptide N-acetylgalactosaminyltransferase T7 (CG6394)	Drosophila melanogaster	AY268068	Q9VWT6 Q8MV48	AAQ56704.1
		BT016123		AAV37008.1
		NM_133073		NP_573301.2
		NM_167623		NP_728178.1
		AE003580		AAF51113.2
		AE003580		AAF51113.3
		BT010030		AAQ22499.1
polypeptide N-acetylgalactosaminyltransferase T2 (CG3254)	Drosophila melanogaster	AY268064	Q9VQQ2 Q7YU21	AAQ56700.1
		AE003580		AAQ564620.1
		NM_134929		NP_608773.1
		AE003766		AAF56810.1
		AE003766		AAF56810.2
		AY113484		AAM29489.1
CG10000	Drosophila melanogaster		Q9VAT9 Q8MYV6	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
polypeptide N-acetylgalactosaminyltransferase T35A (CGT480)	Drosophila melanogaster	NM_143373		
		AE003409		
		AE003643		
		AF158747		
		AF478697		
		AF478698		
		AF478699		
CG31776	Drosophila melanogaster	NM_143812		
		AE003579		
		AE003579		
		AF324751		
		BT022827		
		NM_164538		
		AY353720		
N-acetylgalactosaminyltransferase (ppGalNAc-T1) RCJMB04_17f16 RCJMB04_1b1 GALNT16	Echinococcus granulosus Gallus gallus Gallus gallus Homo sapiens	AJ720423		
		AJ851373		
		BC036812		
		BC098578		
		AB032956		
		AB078143		
		AJ505951		
		CQ816250		
		XM_031104		
		BC006269		
		BC006269		
		BC010659		
FLJ12691	Homo sapiens	AK022753		
		AK024039		
		AX262414		
		AX882259		
		NM_024572		
		NP_078848.1		
		NP_078848.1		
polypeptide N-acetylgalactosaminyltransferase T35A (CGT480)	Homo sapiens	NP_651630.2		
		AAF44871.1		
		AAF53391.1		
		AAK66862.1		
		AAM62405.1		
		AAM62406.1		
		AAM62407.1		
		NP_652069.2		
		AAF51101.2		
		AAF51101.3		
		AAN75750.1		
		AAV55243.1		
CG31776	Drosophila melanogaster	NP_722909.1		
		AAQ56593.1		
polypeptide N-acetylgalactosaminyltransferase T35A (CGT480)	Homo sapiens	CAG32082.1		
		CAH65007.1		
		AAH36812.1		
		AAH98578.1		
		BAA85444.1		
		BAD93178.1		
		CAD44534.1		
		CAG32966.1		
		XP_031104.6		
		AAH06269.1		
		AAH06269.2		
		AAH10659.1		
FLJ12691	Homo sapiens	BAB14227.1		
		BAB14795.1		
FLJ12691	Homo sapiens	CAD10224.1		
		CAE91506.1		
FLJ12691	Homo sapiens	NP_078848.1		
		NP_078848.1		

FIG. 18 (cont.)

24/36

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase (GalNAc-T10)	Homo sapiens	Y09324	Q8IV14	cryst
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase GalNT11 (FLJ21634)	Homo sapiens	AC06017	Q8NCW6	
		BC059377	Q9UDR8	
		AK025287	Q9H6Z5	
		AK026056	Q9H6C2	
		Y12434		
		AX461608		
		NM_022087		
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T6 (GalNAc-T6)	Homo sapiens	BC035822	Q9UIV5	
		BC114505	Q9H6G2	
		AK025961	Q8NCL4	
		AK074658		
		Y08565		
		CQ782628		
		NM_007210		
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 5 (GALNT5)	Homo sapiens	AF154107	Q9UGK7	
		AY277591	Q9UHL6	
		AJ245539	Q7Z7M9	
		AJ505956		
		NM_014568		
		XM_050509		
Unknown (protein for MGC:26636)	Homo sapiens	BC022021	Q7Z4T8	
		AF440400	Q8WW05	
		AF440404	Q8NCV4	
		AC099345		
		AC074257		

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
polypeptide GalNAc transferase 8 (GalNAc-T8)	Homo sapiens			
UDP-GalNAc polypeptide N-acetylgalactosaminyltransferase 12	Homo sapiens			
UDP-GalNAc polypeptide N-acetylgalactosaminyltransferase 12 (GalNAc-T2;GalNT2)	Homo sapiens			
UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNT7)	Homo sapiens			

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
pp-GalNAc-transferase 17 (GALNT17)	Homo sapiens	.	ABA31473.1	
		.	ABA36473.1	
		.	ABA38055.1	
		.	ABA68732.1	
		.	ABA69735.1	
		.	ABA70117.1	
		.	ABA70240.1	
		.	ABC04061.1	
		.	ABC16096.1	
		.	ABE23551.1	
		.	ABE26114.1	
		.	ABE26387.1	
		.	BAD29961.1	
		AB078149	CAD38585.1	
		AL831925	CAG32963.1	
		CQ816244	NP_473451.2	
GalNAc-transferase 18 (GALNT18) aka GalNAc-T15	Homo sapiens	NM_054110	AAH44793.1	
		BC044793	AAH47551.1	
		BC047551	CAF25036.1	
		AJ626725	AAH37341.3	
		BC037341	AAH60864.1	
		BC080864	ABE14321.1	
		.	BAD93179.1	
		AB078147	CAF25035.1	
		AJ626724	NP_940918.1	
		NM_198516	Q9H0E1	
polypeptide N-acetylgalactosaminyltransferase 10 (GalNT10)	Homo sapiens	BC007224	AAH07224.2	Q8TEJ2
		BC050333	AAH50333.1	Q96IV2
		BC072450	AAH72450.1	Q86VP8
		BC098444	AAH98444.1	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
polypeptide N-acetylgalactosaminyltransferase 9 (GalNAc-T9)	Homo sapiens		Q86SR1	
			Q9Y4M4	
			Q8IXJ2	
			AAP65370.1	
			AAP65371.1	
			AAP65372.1	
			AAP65373.1	
			AAP65374.1	
			AAP65376.1	
			AAP65377.1	
			AAP65378.1	
			AAP65379.1	
			ABA33146.1	
			ABA33147.1	
			ABA33148.1	
			ABA33149.1	
			ABA33150.1	
			ABA33151.1	
			ABA33152.1	
			ABA33153.1	
			ABA33154.1	
			ABA33155.1	
		AK023782	BAB14676.1	
		AK074132	BAB84958.1	
		AB078145	BAC56890.1	
		AL096739	CAB46378.1	
		AJ505950	CAD44532.1	
		AX880860	CAE91109.1	
		NM_017540	NP_060010.2	
		NM_198321	NP_938080.1	
		BC069354	AAH69354.1	Q8NFR1
		BC093817	AAH93817.1	Q9HCQ5
		BC093819	AAH93819.1	
		AF458594	AAM49722.1	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
GlaNAc-T15 UDP-GalNAc:polypeptide N-acetyl/galactosaminyltransferase 7	Homo sapiens Homo sapiens	.	Q9UJ28 Q7Z5W7 Q86SF2	28/36
		.		
		AB040672		
		AJ505960		
		NM_021808		
		CQ816248		
		BC035303		
		BC046129		
		BC047468		
		AC105285		
		AA18857.1		
		ABE14200.1		
		BAB13699.2		
		CAD44540.1		
		NP_068580.1		
		CAG32965.1		
		AAH35303.1		
		AAH46129.1		
		AAH47468.1		
		AAH41020.1		

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
polypeptide N-acetylgalactosaminyltransferase 14 (GalNAc-T14)	Homo sapiens	AJ002744		CAB60270.1
		AX136283		CAC39783.1
		AJ505958		CAD44538.1
		CS051319		CAI72154.1
		NM_017423		NP_059119.1
		AY358758		AAQ89118.1
		AAU97249		AAU97249.1
		AC009301		AAK88899.1
		-		ABA66763.1
		AK091313		BAC03634.1
		AB078144		BAC56889.1
		AX405663		CAD34770.1
		AJ505966		CAD44543.1
		AX881147		CAE91230.1
		BC056246	Q14435	AAH56246.1
UDP-GalNAc:polypeptide-N-acetylgalactosaminyltransferase T3 (GALNAc-T3;GALNT3)	Homo sapiens	AC009495		AAV14678.1
		X92689		CAA63371.1
		AJ505954		CAD44536.1
		NM_004482		NP_004473.1
		BC067524		AAH67524.1
pp-GalNAc-transferase 20 (GALNT20)	Homo sapiens	BC067525		AAH67525.1
		BC069624		AAH69624.1
		BC069628		AAH69628.1
		BC069636		AAH69636.1
		BC069645		AAH69645.1
		BC069997		AAH69997.1
		AF410457		AAM62306.1
		-		ABA66799.1
		AB078148		BAD93180.1

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase T1 (GalNAc-T1;GALNT1)	Homo sapiens	AJ626726	CAF25037.1	
		NM_022479	NP_071924.1	
		U41514	AAC50327.1	Q10472
				Q86TJ7
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase T4 (GalNAc-T4;GALNT4)	Homo sapiens	BC047746	AAH47746.1	
		X85018	CAA59380.1	
		AJ505952	CAD44535.1	
		NM_020474	NP_065207.2	
		BC036390	AAH36390.1	Q00208
Polypeptide GalNAc transferase T13 (contains KIAA1918)	Homo sapiens		AAN94203.1	Q8N4A0
			AAN94205.1	
		Y08564	CAA69875.1	
		NM_003774	NP_003765.1	
ORF (fragment) -unnamed protein product (fragment) UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase ORF	Macaca fascicularis	BC101031	AAI01032.1	
		BC101033	AAI01034.1	
		AB078142	BAC54545.1	
		AJ505991	CAD44533.2	
ORF (fragment) ORF (fragment) UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 16	Macaca fascicularis	Q0816246	CAG32964.1	
		NM_052917	NP_443149.1	Q9GLZ7
		AB050513	BAB17281.1	
		AB172088	BAE89150.1	
ORF (fragment) ORF (fragment) UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 16	Macaca fascicularis	AB050509	BAB17277.1	Q9GM01
		AB069985	BAB62930.1	Q95JX4
		AB070020	BAB62965.1	Q95K42
		AB070053	BAB62998.1	Q95K07
ORF (fragment) ORF (fragment) UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 16	Mus musculus	AB168494	BAE00614.1	
		AB169420	BAE01503.1	
		AB045325	BAA97985.1	Q9JU61

FIG. 18 (cont.)

31/36

(ppGalNAc-T16)	Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 14 (pp-GalNAc-T14)		Mus musculus	AK173105	BAD32383.1	
			AB175683	BAD52071.1	
			AK138883	BAE23810.1	
			AK002747	BAB22325.1	Q8BTI6
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 7		Mus musculus	AK078292	BAC37208.1	Q9DCJ2
			AK090146	BAC41113.1	Q8BVG5
			AB175681	BAD52069.1	
			BC007484	AAH07484.1	Q8BZW0
			BC049907	AAH49907.1	Q91VW6
			BC052461	AAH52461.1	Q80VA0
			AF349573	AAK37549.1	Q99MD7
			AK033427	BAC28284.1	Q8BZ70
			AK036523	BAC29461.1	
			AK151424	BAE30388.1	
			AK151753	BAE30662.1	
			AK152205	BAE31034.1	
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 12 (pp-GalNAc-T12)		Mus musculus	AK153143	BAE31754.1	
			NM_144731	NP_653332.2	
			BC056425	AAH56425.1	Q88GT9
Polypeptide GalNAc transferase T3 (GALNT3)		Mus musculus	AK033638	BAC28401.1	
			AK042133	BAC31179.1	
			AB175680	BAD52068.1	
			AK156751	BAE33836.1	
			NM_172693	NP_766281.1	
			U70538	AAB09579.1	P70419
			BC043331	AAH43331.1	Q80V55

FIG. 18 (cont.)

32/36

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide-N-acetylgalactosaminyltransferase t1 (GalNAc-T1,ppGalNAcTase-1;Galnt1)	Mus musculus	AK135489	Q08912	1XHB
		NM_015736		
		U73820		
		BC056215		
pp-GalNAc-T17	Mus musculus	BC090962	Q7TND1	
		AK148550		
		NM_013814		
		BC052469		
		AF467979		
		AK035817		
		AK048758		
		AK048758		
		AK051281		
		AK051281		
		AK081B48		
		AK087726		
		AB175684		
		NM_145218		
		U73819		
		BC057882		
Polypeptide GalNAc transferase T4	Mus musculus	AK033494	Q08832	
		AK148036		
		AK153253		
		NM_015737		
ppGalNAcTase-T5	Mus musculus	AK029323	Q8C102	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
ORF	Mus musculus	NM_172855	Q8K1B9	NP_766443.1
		BC024988		AAH24988.1
		AB175685		BAD91811.1
		NM_173739		NP_776100.2
Unknown (protein for MGC:63144) (fragment)	Mus musculus	BC054061	Q7TQL0	AAH54061.1
		AK171529		BAE42509.1
		NM_198306		NP_938048.1
		BC016585		AAH16585.1
GALNT10;GALNT9	Mus musculus	BC060617	Q91YJ6 Q8BZU8	AAH60617.1
		AK033515		BAC28334.1
		AK131155		BAD21405.1
		AK159088		BAE34806.1
		AK159362		BAE35020.1
		AK172609		BAE43090.1
		AL662868		CAI24136.1
		AL662868		CAI24137.1
		AL662868		CAI24811.1
		NM_134189		NP_598950.1
		AK080614		BAC37957.1
		NM_175032		NP_778197.1
		AK019470		BAB31741.1
		AB175682		BAD52070.1
		BC007172		AAH07172.1
		BC053063		AAH53063.1
UDP-GalNAc:polypeptide -N-acetylgalactosaminyltransferase 15 (pp-GalNAc-T15)	Mus musculus	BC059818	Q922K5 Q7TSI5 Q99ME1 Q8BL27	AAH59818.1
		AF348988		AAK37548.1
		AK046567		BAC32790.1
		AK145122		BAE26246.1
		AK151536		BAE30483.1
		NM_139272		NP_844678.1
polypeptide GalNAc transferase-T2	Mus musculus			

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide-GalNAc-transferase (Gal.nt13)	Mus musculus	AK038387	Q8BYT3 Q8BLE4	BAC29981.1 BAC32353.1 BAC54546.1 NP_766618.1 BAC25984.1
		AK045417		
		AB082928		
		NM_173030		
		AK028506		
Polypeptide GalNAc transferase T6	Mus musculus	AK049222	Q8CED2 Q8C7U7 Q9QZ16	BAC33618.1 BAE32989.1 BAE35316.1 CAB55352.1 NP_766039.1
		AK155008		
		AK159721		
		AJ133523		
		NM_172451		
1700021B12Rik	Mus musculus	BC049554	Q810S4 Q9CW06 Q9D4M9	AAH49554.1 AAI00448.1 BAB24147.1 BAB30220.1 NP_080725
		BC100447		
		AK005605		
		AK016415		
		NM_026449		
A630038F14 UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase GalNT11	Mus musculus	AK041791	Q8BY62 Q921L8 Q8K032 Q8BU26	BAC31068.1 AAH11428.1 AAH21504.1 AAH34184.1 AAH34185.1
		BC011428		
		BC021504		
		BC034184		
		BC034185		
DKFZp469B1122 DKFZp469L0611 Polypeptide GalNAc transferase T6	Pongo pygmaeus Pongo pygmaeus Rattus norvegicus	BC036143	Q9R0C5	AAH36143.1 AAH36145.1 AAQ06688.1 BAC40096.1 BAE36442.1 CAC79626.1 CAH89461.1 CAH89838.1 AAC399426.1
		BC036145		
		AF495528		
		AK088014		
		AK161524		
		Y12435		
		CR857160		
		CR857560		
		AF076167		

FIG. 18 (cont.)

35/36

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
N-acetylgalactosaminyltransferase T5 (GalN5)	Rattus norvegicus	NM_022926 AF049344 NM_031796 BC101886	O88422	
UDP-GalNAc polypeptide N-acetylgalactosaminyltransferase-3 (GalNAc-T3)	Rattus norvegicus			
Polypeptide GalNAc transferase T1	Rattus norvegicus	AB040674 U35890 NM_024373 AY371923 NM_199106	Q10473	
Polypeptide GalNAc transferase T13	Rattus norvegicus	BC078995 BC091351 AF241241 NM_130742 BC105818		
Unknown (protein for MGC:93899)	Rattus norvegicus			
Unknown (protein for IMAGE:7317097) (fragment)	Rattus norvegicus			
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T9	Rattus norvegicus		Q925R7	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase-4 (GalNAc-T4)	Rattus norvegicus			
Unknown (protein for MGC:94136)	Rattus norvegicus	AB040675 BC079128 AB040671 BC062004 NM_199393 AY814498 AY811519 AY810748 D85389 AF234624		
p1-GalNAc-T ORF	Rattus norvegicus			
ORF	Schistosoma japonicum			
SJCHGC07375 (fragment)	Schistosoma japonicum			
SJCHGC03147 (fragment)	Schistosoma japonicum			
Polypeptide GalNAc transferase T1	Sus scrofa			
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T1 (ppGalNAc-T1)	Toxoplasma gondii RH		Q29121 Q8MM26	

FIG. 18 (cont.)

Protein	Organism	GenBank/GenPept	UniPort	PDB/3D
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T4 (ppGalNAc-T4) (fragment)	Toxoplasma gondii RH	AF321881 AY424889	AAM81087.1 AAQ95170.1	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T3 (ppGalNAc-T3)	Toxoplasma gondii RH	AY160970	AAO20278.1	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T5 (ppGalNAc-T5) (fragment)	Toxoplasma gondii RH	AY424892 AYA424890	AAQ95173.1 AAQ95171.1	
UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T2 (ppGalNAc-T2)	Toxoplasma gondii RH	AY136814	AAM95450.1	
Unknown (protein for MGC:81150)	Xenopus laevis	AY424891	AAQ95172.1	
Unknown (protein for MGC:78803)	Xenopus laevis	BC071009	AAH71009.1	
Unknown (protein for MGC:81846)	Xenopus laevis	BC070527	AAH70527.1	
ORF	Xenopus laevis	BC080006	AAH80006.1	
Unknown (protein for MGC:83963)	Xenopus laevis	BC060419	AAH60419.1	
Unknown (protein for MGC:130697)	Xenopus laevis	BC074234	AAH74234.1	
MGC:76103	Xenopus laevis	BC110706	AA110707.1	
TGas006c10.1-001 (probable fragment)	Xenopus tropicalis	BC067317	AAH67317.1	
Unknown (protein for MGC:79650)	Xenopus tropicalis	CR855751	CAJ81616.1	
Unknown (protein for MGC:69370)	Xenopus tropicalis	BC075106	AAH75106.1	
TGas071o13.1-001	Xenopus tropicalis	BC090583 CR942331	AAH90583.1 CAJ83416.1	
UNCLASSIFIED				
Sequence 7 from patent US 6939698 (fragment)	Unknown.		AT18860.1 ABA38429.1	

FIG. 18 (cont.)

6075235 v1