

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 August 2008 (14.08.2008)

PCT

(10) International Publication Number
WO 2008/097829 A2

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

(21) International Application Number:
PCT/US2008/052766

(22) International Filing Date: 1 February 2008 (01.02.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/888,028 2 February 2007 (02.02.2007) US
60/889,706 13 February 2007 (13.02.2007) US
60/893,589 7 March 2007 (07.03.2007) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JOHNSON, Karl, F.** [US/US]; 5230 Ivystream Road, Hatboro, Pennsylvania 10040 (US). **SCHWARTZ, Marc, F.** [US/US]; 324 Oak Lane, West Windsor, New Jersey 08550 (US). **WU, Bingyuan** [CN/US]; 23 Pebble Drive, Horsham, Pennsylvania 19044 (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, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, 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, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

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

— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: LARGE SCALE PRODUCTION OF EUKARYOTIC N-ACETYLGLUCOSAMINYLTRANSFERASE I IN BACTERIA

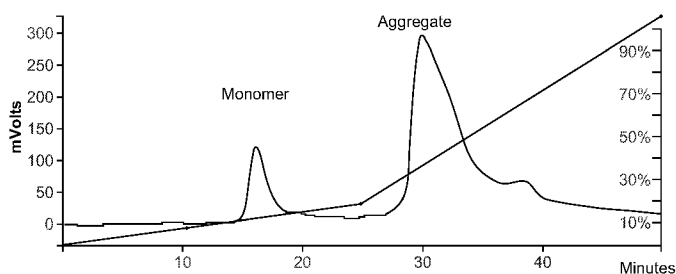


FIG. 2A

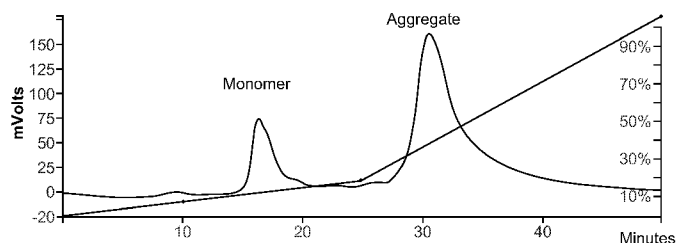


FIG. 2B

(57) Abstract: The present invention provides large scale methods of producing eukaryotic N-acetylglucosaminyltransferase I in bacterial cells.



WO 2008/097829 A2

LARGE SCALE PRODUCTION OF EUKARYOTIC N-ACETYLGLUCOSAMINYLTRANSFERASE I IN BACTERIA

CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 [0001] This application claims the benefit of US Provisional Application No. 60/888,028, filed February 2, 2007; US Provisional Application 60/889,706, filed February 13, 2007; and US Provisional Application 60/893,589, filed March 7, 2007; each of which is herein incorporated by reference for all purposes.

FIELD OF INVENTION

- 10 [0002] The present invention provides large scale methods of producing eukaryotic N-acetylglucosaminyltransferase I in bacterial cells.

BACKGROUND OF THE INVENTION

- [0003] Eukaryotic organisms synthesize oligosaccharide structures or glycoconjugates, such as glycolipids or glycoproteins, that are commercially and therapeutically useful. *In*
15 *vitro* synthesis of oligosaccharides or glycoconjugates can be carried out using recombinant eukaryotic glycosyltransferases. The most efficient method to produce recombinant eukaryotic glycosyltransferases for oligosaccharide synthesis is to express the protein in bacteria. However, in bacteria, many eukaryotic glycosyltransferases are expressed as insoluble proteins in bacterial inclusion bodies, and yields of active protein from the inclusion
20 bodies can be very low. Thus, there is a need for improved methods to produce eukaryotic glycosyltransferases in bacteria. The present invention solves this and other needs.

BRIEF SUMMARY OF THE INVENTION

- [0004] In one aspect, the invention provides a large scale method of producing an active
25 mammalian N-acetylglucosaminyltransferase I (GnT1) protein in bacteria, by expressing the mammalian GnT1 protein in bacteria as an insoluble protein, solubilizing the insoluble mammalian GnT1 protein after harvest from the bacterial cells; and refolding the soluble mammalian GnT1 protein in a buffer comprising a redox couple. In some embodiments, the soluble GnT1 protein will be subjected to additional purification steps.

[0005] In another aspect, the invention provides an expression plasmid for expression of a eukaryotic GnT1 protein in bacteria. The nucleic acid that encodes the GnT1 protein terminates translation of the GnT1 protein with two contiguous stop codons. In one embodiment, the GnT1 protein is a human GnT1 protein. In one embodiment, the expression plasmid has the sequence shown in SEQ ID NO:12. In another embodiment, the invention provides a host cell transformed the GnT1 expression plasmid. In a further embodiment, the invention provides a method of making a GnT1 protein, by growing the host cell transformed the GnT1 expression plasmid under conditions suitable for expression of the GnT1 protein. The invention also includes methods of purifying the expressed GnT1 protein from the host cell, including steps of solubilizing and refolding the GnT1 protein.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0006] Figure 1 provides the complete peptide sequence of MBP-GNT-1 double STOP, including the two stop codons.
- [0007] Figure 2 shows a comparison of the chromatographic separation by anionic exchange of refolded MBP-GnT1 with either a double STOP (Figure 2A) codon or single STOP codon (Figure 2B). The monomer and aggregate peaks, as determined by SDS-PAGE, analysis, are indicated.
- [0008] Figure 3 shows the SDS-PAGE analysis of the MBP-GnT1 monomer pools from the anionic exchange column chromatography for the single and double STOP codons (Figure 2). Following electrophoresis, the gel was silver stained. Lane 1 is the monomer pool from the Double STOP construct; lane 2 is the monomer pool from the Single STOP codon construct. Full length MBP-GnT1 is indicated with a box.
- [0009] Figure 4 shows the mass yield and activity yield from the monomer pools from the anionic exchange column chromatography for single and double STOP forms of MBP-GnT1. The monomer pools were assayed for protein concentration and GnT1 activity. Panel A shows the mass yield in mg MBP-GnT1 per gram wet weight of inclusion bodies. Panel B shows the activity yield in Units GnT1 activity per gram wet weight of inclusion bodies. DS represents the double STOP construct and SS represents the single STOP construct.
- [0010] Figure 5 shows the activity yield for both the single and double STOP forms of MBP-GnT1. Inclusion bodies from both MBP-GnT1 constructs were refolded using optimized conditions. Following buffer exchange, the samples were assayed for GnT1

activity. Panel A shows the activity yield in units of GnT1 per liter of refold reaction. Panel B shows the activity yield in Units of GnT1 activity per gram wet weight of inclusion bodies. DS represents the double STOP construct and SS represents the single STOP construct.

[0011] Figure 6 provides a restriction map for the GnT1ds expression plasmid.

5 [0012] Figure 7 provides results of initial experiments comparing HA Type I, HA Type II, and fluoroapatite resins.

[0013] Figure 8 provides a comparison of the pooled fractions from the hydroxyapatite and fluoroapatite chromatography. A 4-12% NuPAGE was used. Thirty microliters of protein solution were mixed with 10 μ l 4X loading buffer (reducing samples plus 4 μ l of 1 M DTT stock solution) and denatured at 85 °C for 5 min. Thirty microliters of denatured mixture were loaded each well. The gel was run at 150 V/RT for 1 hr and stained with Silver stain II kit [Wako 291-50301]. Lane 1: Pooled Source 30Q fractions, non-reducing; Lane 2: HAI/pH8, non-reducing; Lane 3: HAI/pH9, non-reducing; Lane 4: HAI/pH8, non-reducing; Lane 5: FAII/pH8, non-reducing; Lane 6: MW marker; Lane 7: Pooled Source 30Q
10 fractions, reducing; Lane 8: HAI/pH8, reducing; Lane 9: HAI/pH9, reducing; Lane 10: HAI/pH8, reducing; and Lane 11: FAII/pH8, reducing.

[0014] Figure 9 provides the results of purification of MBP-GNT-1 by hydroxyapatite type I from three separate fifteen liter production runs.

[0015] Figure 10 provides a GNT1 purification process flow diagram.

20 [0016] Figure 11 provides SDS-PAGE analysis of the Purified MGnT-1 using G25 Chromatography. MW = Molecular weight standards (See Blue Plus2, Invitrogen) **Lane 1.** MGnT-1 HA pool (2 μ g), **Lane 2.** MGnT-1 G25 pool (1.5 μ g). **Panel A.** Coomassie stained non-reducing gel (Simply Blue Safe Stain, Invitrogen), **Panel B.** Silver stained nonreducing gel (Wako Silver Stain II kit). A 4-20 % tris-glycine gel (Invitrogen) was used. Samples (50
25 μ L) were mixed with 4X gel loading dye and heated at 80°C for 4 min. Aliquots (20 μ L) aliquots were loaded onto two separate 4-20 % tris-glycine gels. One gel was stained with Simply Blue Safe stain, the other with Silver stain according to manufacturers' protocols.

[0017] Figure 12 provides MGnT-1 enzyme activity and mass yields after G25 sephadex chromatography. A total of 88 % of the HA chromatography pool was loaded onto the G25
30 column.

DEFINITIONS

[0018] The recombinant glycosyltransferase proteins 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*

[0019] 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).

[0020] 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.

[0021] All oligosaccharides described herein are described with the name or abbreviation for the non-reducing saccharide (*e.g.*, Gal), followed by the configuration of the glycosidic bond (α or β), the ring bond, the ring position of the reducing saccharide involved in the bond, and then the name or abbreviation of the reducing saccharide (*e.g.*, GlcNAc). The

linkage between two sugars may be expressed, for example, as 2,3, 2→3, or (2,3). Each saccharide is a pyranose or furanose.

[0022] 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-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.

[0023] An “acceptor substrate” for a glycosyltransferase is an oligosaccharide moiety that 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 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, 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.*, peptides, proteins, glycopeptides, and glycoproteins.

[0024] 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.

[0025] 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.

[0026] A "eukaryotic glycosyltransferase" as used herein refers to an enzyme that is derived from a eukaryotic organism and that catalyzes transfer of a sugar residue from a donor substrate, *i.e.*, 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 eukaryotic glycosyltransferase transfers a sugar from a donor substrate to a peptide, a protein, a glycopeptide, or a glycoprotein. In another preferred embodiment, a eukaryotic glycosyltransferase is a type II transmembrane glycosyltransferase. A eukaryotic glycosyltransferase can be derived from an eukaryotic organism, *e.g.*, a 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.

[0027] 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 GnT1 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. Exemplary GnT1 proteins are disclosed at, *e.g.*, SEQ ID NOs:1-11. Other eukaryotic *N*-acetylglucosaminyltransferase proteins that can be used in the present invention are include, *e.g.*, BGnT-1, GnT-II, GnT-III, GnT-IV (*e.g.*, GnT-IVa and GnT-IVb), GnT-V, GnT-VI, and GnT-IVH, which are disclosed in Schwartz and Soliman, WO/2006/102652.

[0028] 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 bind with another cysteine residue.

[0029] An "insoluble glycosyltransferase" refers to a glycosyltransferase that is expressed in bacterial inclusion bodies. Insoluble glycosyltransferases are typically solubilized or denatured using *e.g.*, detergents or chaotropic agents or some combination. "Refolding" refers to a process of restoring the structure of a biologically active glycosyltransferase to a glycosyltransferase that has been solubilized or denatured. Thus, a refolding buffer, refers to a buffer that enhances or accelerates refolding of a glycosyltransferase.

[0030] A "redox couple" refers to mixtures of reduced and oxidized thiol reagents and include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteine/cysteamine, cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g.*, Clark, *Cur. Op. Biotech.* 12:202-207 (2001)).

[0031] The term "contacting" is used herein interchangeably with the following: combined with, added to, mixed with, passed over, incubated with, flowed over, etc.

[0032] 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 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.

5 [0033] The term “specific activity” as used herein refers to the catalytic activity of an enzyme, *e.g.*, a recombinant glycosyltransferase fusion protein of the present invention, and may be expressed in activity units. As used herein, one activity unit catalyzes the formation of 1 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
10 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.

[0034] “N-linked” oligosaccharides are those oligosaccharides that are linked to a peptide backbone through asparagine, by way of an asparagine-N-acetylglucosamine linkage. N-linked oligosaccharides are also called “N-glycans.” All N-linked oligosaccharides have a
15 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, this structure may also contain a core fucose molecule and/or a xylose molecule.

[0035] “O-linked” oligosaccharides are those oligosaccharides that are linked to a peptide
20 backbone through threonine, serine, hydroxyproline, tyrosine, or other hydroxy-containing amino acids.

[0036] 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.*,
25 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.

[0037] The term “biological activity” refers to an enzymatic activity of a protein. For
30 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 GnT1

protein refers to the activity of transferring a *N*-acetylglucosamine moiety from a donor molecule to an acceptor molecule.

[0038] "Commercial scale" refers to gram scale production of a product saccharide in a single reaction. In preferred embodiments, commercial scale refers to production of greater than about 50, 75, 80, 90 or 100, 125, 150, 175, or 200 grams.

[0039] 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.

[0040] The term "amino acid" refers to naturally occurring and 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.

[0041] "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).

[0042] 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 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. In preferred embodiments, a recombinant eukaryotic glycosyltransferase is produced by a recombinant bacterial cell.

[0043] 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.

[0044] 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 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. The recombinant fusion protein 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 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 suitable tags are known to those of skill in the art, and include, for example, an affinity tag such as a hexahistidine peptide, 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.). Maltose binding domains are preferably used for their ability to enhance refolding of insoluble eukaryotic glycosyltransferases, but can also be used to assist in purification of a fusion protein. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains 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.

[0045] 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, and signal-anchor domain.

[0046] 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.

[0047] 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.

[0048] 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 trans-membrane 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 is 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 1GalT1, amino acid residues from about 36 to about 102 (see *e.g.*, the *Drosophila* enzyme); the stem region of mammalian Core 1 GalT1, 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 ST6GalNAcI, 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 GalNAcT2, amino acid residues from about 71 to about 129 (see *e.g.*, the rat enzyme).

[0049] 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 ST3GalIII, amino acid residues from about 85 to about 374 (see, *e.g.*, the rat enzyme). Catalytic domains and truncation mutants of GalNAcT2 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.

[0050] 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.

[0051] 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 GalNAcT2 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 ST6GalNAcI enzymes, and truncated GalNAcT2 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.

[0052] 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.

[0053] 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.

[0054] 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 for modifying a heterologous sequence.

[0055] 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 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 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.

[0056] 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 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.

[0057] 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 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.

[0058] 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.

[0059] 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.

[0060] 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)).

[0061] 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 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 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 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)).

[0062] 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 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.

[0063] A further indication that two nucleic acid sequences or proteins are substantially 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] “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.

[0068] 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.

[0069] 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. 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 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 also “conservatively modified variations”.

[0070] 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 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
5 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*)

[0071] Examples of protocols sufficient to direct persons of skill through *in vitro*
10 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-
15 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
20 described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0072] The present invention provides conditions for large scale production of mammalian
25 GnT1 enzymes that are expressed as insoluble proteins in bacterial inclusion bodies. Refolding buffers comprising redox couples are used to enhance refolding of insoluble GnT1 proteins. Refolding can also be enhanced by fusing a maltose binding domain to the insoluble GnT1 protein. For some insoluble GnT1 proteins, refolding can also be enhanced by site directed mutagenesis to remove unpaired cysteines. Additional refolding
30 enhancement can be provided by truncating a GnT1 protein to remove, *e.g.*, a signal-anchor domain, a transmembrane domain, and/or all or a portion of a stem region of the protein. The refolded GnT1 proteins can be used to produce or to remodel polysaccharides,

oligosaccharides, glycolipids, proteins, peptides, glycopeptides, and glycoproteins. The refolded GnT1 protein can also be used to glycoPEGylate proteins, peptides, glycopeptides, or glycoproteins as described in PCT/US02/32263, which is herein incorporated by reference for all purposes.

- 5 [0073] The invention also provides a unique system for enhanced expression of the GnT1 protein in bacteria. The GnT1 protein is expressed using an expression vector that terminates GnT1 translation with two contiguous stop codons.

II. Eukaryotic *N*-Acetylglucosaminyltransferase I proteins

- 10 [0074] The GnT1 proteins of use in practicing the present invention are preferably mammalian GnT1 proteins. Gn1 proteins typically include structural domains common to many eukaryotic glycosyltransferases.

Eukaryotic glycosyltransferases

- 15 [0075] Some eukaryotic glycosyltransferases have topological domains at their amino terminus that are not required for catalytic activity (*see*, US Patent No. 5, 032,519). Of the glycosyltransferases characterized to date, the “cytoplasmic domain,” is most commonly between about 1 and about 10 amino acids in length, and is the most amino-terminal domain; the adjacent domain, termed the “signal-anchor domain,” is generally between about 10-26 amino acids in length; adjacent to the signal-anchor domain is a “stem region,” which is generally between about 20 and about 60 amino acids in length, and known to function as a retention signal to maintain the glycosyltransferase in the Golgi apparatus; and at the 20 carboxyl side of the stem region is the catalytic domain.

- 25 [0076] Many mammalian glycosyltransferases have been cloned and expressed and the recombinant proteins have been characterized in terms of donor and acceptor substrate specificity and they have also been investigated through site directed mutagenesis in attempts to define residues or domains involved in either donor or acceptor substrate specificity (Aoki *et al.* (1990) *EMBO. J.* 9: 3171-3178; Harduin-Lepers *et al.* (1995) *Glycobiology* 5(8): 741-758; Natsuka and Lowe (1994) *Current Opinion in Structural Biology* 4: 683-691; Zu *et al.* (1995) *Biochem. Biophys. Res. Comm.* 206(1): 362-369; Seto *et al.* (1995) *Eur. J. Biochem.* 234: 323-328; Seto *et al.* (1997) *J. Biol. Chem.* 272: 14133-14138).

- 30 [0077] In one group of embodiments, a functional domain of the recombinant GnT1 protein of the present inventions is obtained from known GnT1 proteins. Exemplary GnT1 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; and mouse, accession number P27808.

III. Refolding insoluble mammalian GnT1 proteins

5 [0078] Many recombinant proteins expressed in bacteria are expressed as insoluble aggregates in bacterial inclusion bodies. Inclusion bodies are protein deposits found in both the cytoplasmic and periplasmic space of bacteria. (*See, e.g., Clark, Cur. Op. Biotech.* 12:202-207 (2001)). Eukaryotic glycosyltransferases, including GnT1 proteins, are frequently expressed in bacterial inclusion bodies. Some eukaryotic glycosyltransferases are
10 soluble in bacteria, *i.e.*, not produced in inclusion bodies, when only the catalytic domain of the protein is expressed. However, many eukaryotic glycosyltransferases remain insoluble and are expressed in bacterial inclusion bodies, even if only the catalytic domain is expressed, and methods for refolding these proteins to produce active glycosyltransferases are provided herein.

15 A. Conditions for refolding active glycosyltransferases

[0079] To produce active mammalian GnT1 proteins in bacterial cells, mammalian GnT1 proteins are expressed in bacterial inclusion bodies, the bacteria are harvested, disrupted and the inclusion bodies are isolated and washed. The proteins within the inclusion bodies are then solubilized. Solubilization can be performed using denaturants, *e.g.*, guanidinium
20 chloride or urea; extremes of pH, such as acidic or alkaline conditions; or detergents.

[0080] After solubilization, denaturants are removed from the GnT1 mixture. Denaturant removal can be done by a variety of methods, including dilution into a refolding buffer or buffer exchange methods. Buffer exchange methods include dialysis, diafiltration, gel filtration, and immobilization of the protein onto a solid support. (*See, e.g., Clark, Cur. Op. Biotech.* 12:202-207 (2001)). Any of the above methods can be combined to remove
25 denaturants.

[0081] Disulfide bond formation in the mammalian GnT1 protein is promoted by addition of a refolding buffer comprising a redox couple. Redox couples include reduced and oxidized glutathione (GSH/GSSG), cysteine/cystine, cysteine/ cystamine,
30 cysteamine/cystamine, DTT/GSSG, and DTE/GSSG. (*See, e.g., Clark, Cur. Op. Biotech.* 12:202-207 (2001), which is herein incorporated by reference for all purposes). In some

embodiments, redox couples are added at an particular ratio of reduced to oxidized component, *e.g.*, 1/20, 20/1, 1/4, 4/1, 1/10, 10/1, 1/2, 2/1, 1/5, 5/1, or 5/5.

[0082] Refolding can be performed in buffers at pH's ranging from, for example, 6.0 to 10.0. Refolding buffers can include other additives to enhance refolding, *e.g.*, L-arginine (0.4-1M); PEG; low concentrations of denaturants, such as urea (1-2M) and guanidinium chloride (0.5-1.5 M); and detergents (*e.g.*, Chaps, SDS, CTAB, lauryl maltoside, and Triton X-100). In one embodiment, refolding is performed at a pH of about 8.2.

[0083] Refolding can be over a given period of time, *e.g.*, for 1-48 hours, or overnight. Longer refolding periods can also be used, *e.g.*, 50, 60, 70, 80, 90, or 100 hours. Refolding can be done from about 4°C to about 40°C, including ambient temperatures.

[0084] A mammalian GnT1 protein comprising a catalytic domain is expressed in bacterial inclusion bodies and then refolded using the above methods. Mammalian GnT1 proteins that comprise all or a portion of a stem region and a catalytic domain can also be used in the methods described herein, as can mammalian GnT1 proteins comprising a catalytic domain fused to an MBP protein.

[0085] Those of skill will recognize that a mammalian GnT1 protein has been refolded correctly when the refolded protein has detectable biological activity. For a mammalian GnT1 protein, biological activity is the ability to catalyze transfer of a donor substrate to an acceptor substrate. Biological activity includes *e.g.*, specific activities of at least 1, 2, 5, 7, or 10 units of activity. Unit is defined as follows: one activity unit catalyzes the formation of 1 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.

[0086] In one embodiment, eukaryotic GnT1 is expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

B. *Fusion of mammalian GnT1 proteins to maltose binding protein domains to enhance refolding*

[0087] Maltose binding protein (MBP) domains are typically fused to proteins to enhance solubility of a the protein with a cell. *See, e.g.,* Kapust and Waugh *Pro. Sci.* 8:1668-1674

(1999). However, many eukaryotic glycosyltransferases, including truncated eukaryotic glycosyltransferases, remain insoluble when expressed in bacteria, even after fusion to a MBP domain. However, this application discloses that MBP domains can enhance refolding of insoluble eukaryotic glycosyltransferases after solubilization of the proteins from *e.g.,* an inclusion body. MBP domains from a variety of bacterial sources can be used in the invention, for example *Yersinia E. coli*, *Pyrococcus furiosus*, *Thermococcus litoralis*, *Thermatoga maritime*, and *Vibrio cholerae*. In a preferred embodiment an *E. coli* MBP protein is fused to a mammalian GnT1 protein. Amino acid linkers can be placed between the MBP domain and the mammalian GnT1 protein. In another preferred embodiment, the MBP domain is fused to the amino terminus of the mammalian GnT1 protein.

[0088] In one embodiment, a eukaryotic GnT1 protein is fused to an MBP domain, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.,* GSH/GSSG or cystamine/cysteine.

[0089] Additional amino acid tags can be added to an MBP-mammalian GnT1 fusion. For example, purification tags can be added to enhance purification of the refolded protein.

Purification tags include, *e.g.,* a polyhistidine tag, a glutathione S transferase (GST), a starch binding protein (SBP), an *E. coli* thioredoxin domain, a carboxy-terminal half of the SUMO protein, a FLAG epitope, and a myc epitope. Refolded glycosyltransferases can be further purified using a binding partner that binds to the purification tag. In a preferred embodiment, an MBP tag is fused to the mammalian GnT1 protein to enhance refolding. Additional purification tags can be fused to MBP mammalian GnT1 fusion protein.

[0090] In another embodiment, addition of an MBP domain to a mammalian GnT1 protein can increase the expression of the protein.

[0091] In another embodiment a self-cleaving protein tag, such as an intein, is included between the MPB domain and the mammalian GnT1 protein to facilitate removal of the MBP domain after the fusion protein has been refolded. Inteins and kits for their use are commercially available, *e.g.,* from New England Biolabs.

C. *Mutagenesis of mammalian GnT1 protein to enhance refolding*

[0092] Refolding of glycosyltransferases can also be enhanced by mutagenesis of the glycosyltransferase amino acid sequence. In one embodiment an unpaired cysteine residue is identified and mutated to enhance refolding of a glycosyltransferase. In another
5 embodiment, the amino terminus of the glycosyltransferase is truncated to remove a transmembrane domain, or to remove a transmembrane domain and all or a portion of the stem region of the protein. In a further embodiment, a glycosyltransferase is mutated to remove at least one unpaired cysteine residue and to truncate the amino terminus of the protein, *e.g.*, to remove a transmembrane domain, or to remove a transmembrane domain and
10 all or a portion of the stem region of the protein. Once a glycosyltransferase nucleic acid sequence has been isolated, standard molecular biology methods can be used to change the nucleic acid sequence and thus the encoded amino acid sequence in a manner described herein.

1. Mutagenesis of unpaired cysteines in glycosyltransferases to enhance
15 refolding

[0093] As refolding occurs, cysteine residues in a denatured protein form disulfide bonds that help to reproduce the structure of the active protein. Incorrect pairing of cysteine residues can lead to protein misfolding. Proteins with unpaired cysteine residues are susceptible to misfolding because a normally unpaired cysteine can form a disulfide bond
20 with normally paired cysteine making correct cysteine pairing and protein refolding impossible. Thus, one method to enhance refolding of a particular glycosyltransferase is to identify unpaired cysteine residues and remove them.

[0094] Unpaired cysteine residues can be identified by determining the structure of the glycosyltransferase of interest. Protein structure can be determined based on actual data for
25 the glycosyltransferase of interest, *e.g.*, circular dichroism, NMR, and X-ray crystallography. Protein structure can also be determined using computer modeling. Computer modeling is a technique that can be used to model related structures based on known three-dimensional structures of homologous molecules. Standard software is commercially available. (*See e.g.*, www.accelrys.com for the multitude of software available to do computer modeling.) Once
30 an unpaired cysteine residue is identified, the DNA encoding the glycosyltransferase of interest can be mutated using standard molecular biology techniques to remove the unpaired cysteine, by deletion or by substitution with another amino acid residue. Computer modeling

is used again to select an amino acid of appropriate size, shape, and charge for substitution. Unpaired cysteines can also be determined by peptide mapping. Once the glycosyltransferase of interest is mutated, the protein is expressed in bacterial inclusion bodies and refolding ability is determined. A correctly refolded glycosyltransferase will have biological activity.

5 [0095] In preferred embodiments, the following amino acid residues are substituted for an unpaired cysteine residue in a eukaryotic glycosyltransferase to enhance refolding: Ala, Ser, Thr, Asp, Ile, or Val. Gly can also be used if the unpaired cysteine is not in a helical structure.

[0096] GnT1 proteins can exhibit enhanced refolding on mutation of an unpaired cysteines
10 in the GnT1 sequence. The crystal structure of a truncated form of rabbit GnTI (105 amino terminal amino acids deleted) shows an unpaired cysteine residue (CYS123) near the active site. *See, e.g., Unligil et al., EMBO J. 19:5269-5280 (2000).* To improve refolding of the human GnT1 protein, the corresponding unpaired cysteine in the human GnTI was identified as CYS121 and was replaced with a series of amino acids that are similar in size and
15 chemical characteristics. *See, e.g., Saribas et al. WO/2005/089102.* The amino acids used include serine (Ser), threonine (Thr), alanine (Ala) and aspartic acid (Asp). In addition, a double mutant, ARG120ALA, CYS121HIS, was also made. The mutant GnTI/MBP fusion proteins were expressed in *E. coli*, refolded and assayed for GnTI activity towards glycoproteins.

20 [0097] In one embodiment, a GnT1 protein is mutated to remove an unpaired cysteine residue, *e.g.*, CYS121SER, expressed in bacterial inclusion bodies, solubilized, and refolded in a buffer comprising a redox couple, *e.g.*, GSH/GSSG or cystamine/cysteine.

2. Truncation of glycosyltransferases to enhance refolding

[0098] Eukaryotic glycosyltransferases generally include the following domains: a
25 catalytic domain, a stem region, a transmembrane domain, and a signal-anchor domain. When expressed in bacteria, the signal anchor domain, and transmembrane domains are typically deleted. Mammalian GnT1 protein used in the methods of the invention can include all or a portion of the stem region and the catalytic domain. In some embodiments, the mammalian GnT1 proteins comprise only the catalytic domain.

30 [0099] Glycosyltransferase domains can be identified for deletion mutagenesis. For example, those of skill in the art can identify a stem region in a eukaryotic

glycosyltransferase and delete stem region amino acids one by one to identify truncated eukaryotic glycosyltransferase proteins with high activity on refolding.

[0100] The deletion mutants in this application are referenced in two ways: Δ or D followed by the number of residues deleted from the amino terminus of the native full length amino acid sequence, or by the symbol and residue number of the first amino acid residue translated from the native full length amino acid sequence.

[0101] Deletion mutations can also be made in a GnT1 protein. For example, the human GnT1 protein includes a stem region from about amino acid residues 31-112. Thus, a truncated human GnT1 protein can have deletions at the amino terminus of about *e.g.*, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, or 111 residues.

IV. Nucleic acids

[0102] Nucleic acids that encode mammalian GnT1 proteins, and methods of obtaining such nucleic acids, are known to those of skill in the art. Suitable nucleic acids (*e.g.*, cDNA, genomic, or subsequences (probes)) can be cloned, or amplified by *in vitro* methods such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), or the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies 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* 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook *et al.*); *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., (1994 Supplement) (Ausubel); Cashion *et al.*, U.S. patent number 5,017,478; and Carr, European Patent No. 0,246,864.

[0103] A DNA that encodes a mammalian GnT1 protein, or a subsequences thereof, can be prepared by any suitable method described above, including, for example, cloning and

restriction of appropriate sequences with restriction enzymes. In one preferred embodiment, nucleic acids encoding glycosyltransferases are isolated by routine cloning methods. A nucleotide sequence of a mammalian GnT1 protein as provided in, for example, GenBank or other sequence database (see above) can be used to provide probes that specifically hybridize to a glycosyltransferase gene in a genomic DNA sample, or to an mRNA, encoding a glycosyltransferase, in a total RNA sample (e.g., in a Southern or Northern blot). Once the target nucleic acid encoding a mammalian GnT1 protein is identified, it can be isolated according to standard methods known to those of skill in the art (see, e.g., Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory; Berger and Kimmel (1987) *Methods in Enzymology*, Vol. 152: *Guide to Molecular Cloning Techniques*, San Diego: Academic Press, Inc.; or Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York). Further, the isolated nucleic acids can be cleaved with restriction enzymes to create nucleic acids encoding the full-length mammalian GnT1 protein, or subsequences thereof, e.g., containing subsequences encoding at least a subsequence of a stem region or catalytic domain of a mammalian GnT1 protein. These restriction enzyme fragments, encoding a mammalian GnT1 protein or subsequences thereof, may then be ligated, for example, to produce a nucleic acid encoding a recombinant mammalian GnT1 protein or fusion protein.

[0104] A nucleic acid encoding a mammalian GnT1 protein, or a subsequence thereof, can be characterized by assaying for the expressed product. Assays based on the detection of the physical, chemical, or immunological properties of the expressed protein can be used. For example, one can identify a cloned mammalian GnT1 protein, including a mammalian GnT1 fusion protein, by the ability of a protein encoded by the nucleic acid to catalyze the transfer of a saccharide from a donor substrate to an acceptor substrate. In a preferred method, capillary electrophoresis is employed to detect the reaction products. This highly sensitive assay involves using either saccharide or disaccharide aminophenyl derivatives which are labeled with fluorescein as described in Wakarchuk *et al.* (1996) *J. Biol. Chem.* 271 (45): 28271-276. For example, to assay for a *Neisseria lgtC* enzyme, either FCHASE-AP-Lac or FCHASE-AP-Gal can be used, whereas for the *Neisseria lgtB* enzyme an appropriate reagent is FCHASE-AP-GlcNAc (*Id.*).

[0105] Also, a nucleic acid encoding a mammalian GnT1 protein, or a subsequence thereof, can be chemically synthesized. Suitable methods include the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.*

(1979) *Meth. Enzymol.* 68: 109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill recognizes that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences.

[0106] Nucleic acids encoding mammalian GnT1 proteins, or subsequences thereof, can be cloned using DNA amplification methods such as polymerase chain reaction (PCR). Thus, for example, the nucleic acid sequence or subsequence is PCR amplified, using a sense primer containing one restriction enzyme site (*e.g.*, *NdeI*) and an antisense primer containing another restriction enzyme site (*e.g.*, *HindIII*). This will produce a nucleic acid encoding the desired glycosyltransferase or subsequence and having terminal restriction enzyme sites. This nucleic acid can then be easily ligated into a vector containing a nucleic acid encoding the second molecule and having the appropriate corresponding restriction enzyme sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided in GenBank or other sources. Appropriate restriction enzyme sites can also be added to the nucleic acid encoding the mammalian GnT1 protein or protein subsequence by site-directed mutagenesis. The plasmid containing the mammalian GnT1 protein-encoding nucleotide sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods 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.

[0107] Other physical properties of a cloned mammalian GnT1 protein, including a mammalian GnT1 fusion protein, expressed from a particular nucleic acid, can be compared to properties of known mammalian GnT1 proteins to provide another method of identifying

suitable sequences or domains of the mammalian GnT1 protein that are determinants of acceptor substrate specificity and/or catalytic activity. Alternatively, a putative mammalian GnT1 gene or recombinant gene can be mutated, and its role as GnT1 enzyme, its ability to be refolded, or the role of particular sequences or domains established by detecting a variation in the structure of a carbohydrate normally produced by the unmutated, naturally-occurring, or control mammalian GnT1 protein.

[0108] Functional domains of cloned mammalian GnT1 protein can be identified by using standard methods for mutating or modifying the mammalian GnT1 protein and testing the modified or mutated proteins for activities such as acceptor substrate activity and/or catalytic activity, as described herein. The functional domains of the various mammalian GnT1 protein can be used to construct nucleic acids encoding recombinant mammalian GnT1 fusion proteins comprising the functional domains of one or more mammalian GnT1 protein. These fusion proteins can then be tested for the desired acceptor substrate or catalytic activity.

[0109] In an exemplary approach to cloning recombinant mammalian GnT1 fusion proteins, the known nucleic acid or amino acid sequences of cloned mammalian GnT1 proteins are aligned and compared to determine the amount of sequence identity between various mammalian GnT1 proteins. This information can be used to identify and select protein domains that confer or modulate GnT1 activities, *e.g.*, acceptor substrate activity and/or catalytic activity based on the amount of sequence identity between the mammalian GnT1 proteins of interest. For example, domains having sequence identity between the mammalian GnT1 protein of interest, and that are associated with a known activity, can be used to construct recombinant mammalian GnT1 fusion proteins containing that domain, and having the activity associated with that domain (*e.g.*, acceptor substrate specificity and/or catalytic activity).

V. Expression of recombinant mammalian GnT1 proteins

[0110] In preferred embodiments, the GnT1 polypeptides of the invention are expressed in *E. coli* host cells. In a further preferred embodiment, *E. coli* strains JM109 or BNN93 are used as host cells.

[0111] In another preferred embodiment, the pCWin2 expression vector is used to express the GnT1 protein in an *E. coli* host cell. The pCWin2 vector is known and includes versions that express MBP fusion proteins. See, e.g., WO/2005/067601 (2005).

[0112] Recombinant mammalian GnT1 proteins can be expressed in a variety of host cells, including *E. coli*, other bacterial hosts. The host cells are preferably bacterial cells. Examples of suitable host cells include, for example, *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. Examples of useful bacteria include, but are not limited to, *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Klebsiella*.

[0113] Typically, the polynucleotide that encodes the mammalian GnT1 protein is placed under the control of a promoter that is functional in the desired host cell. An extremely wide variety of promoters is 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 host cell.

[0114] 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.

[0115] For expression of recombinant eukaryotic glycosyltransferases 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*.

[0116] 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).

[0117] 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 mammalian GnlT1 protein is induced. High level expression of heterologous proteins slows cell growth in some situations. 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, provides a level of expression that is greater than that provided by either promoter alone.

[0118] Inducible promoters for other organisms are also well known to those of skill in the art. These include, for example, the arabinose promoter, the *lacZ* promoter, the metallothionein promoter, and the heat shock promoter, as well as many others.

[0119] 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.

[0120] It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems include the *lac*, *tac*, and *trp* operator systems.

[0121] 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 *Streptomyces* or *Bacillus* is also possible.

[0122] 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*. A preferred selectable marker for use in bacterial cells is a kanamycin resistance marker (Vieira and Messing, *Gene* 19: 259 (1982)). Use of kanamycin selection is advantageous over, for example, ampicillin selection because ampicillin is quickly degraded by β -lactamase in culture medium, thus removing selective pressure and allowing the culture to become overgrown with cells that do not contain the vector.

[0123] 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 plasmids can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequencing 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 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).

[0124] 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.

[0125] The methods for introducing the expression vectors into a chosen host cell 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 or electroporation. Other transformation methods are also suitable.

5 [0126] Translational coupling may be used to enhance expression. The strategy uses a 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
10 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.

[0127] The recombinant eukaryotic glycosyltransferases of the invention can also be further linked to other bacterial proteins. This approach often results in high yields, because
15 normal prokaryotic control sequences direct transcription and translation. In *E. coli*, *lacZ* fusions are often used to express heterologous proteins. 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 and/or accessory enzyme amino acids from the fusion protein after purification. This can be
20 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., 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.

25 [0128] 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-
30 terminal residue.

[0129] Expression of mammalian GnT1 proteins in bacteria can also be enhanced by use of particular stop codons in bacteria, as disclosed in the Example section herein.

Proteins and protein purification

[0130] The recombinant mammalian GnT1 proteins 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)). In preferred embodiments, purification of the recombinant eukaryotic glycosyltransferase proteins occurs after refolding of the protein. Substantially pure compositions of at least about 70 to 90%, homogeneity are preferred; more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, or 97%; and 98 to 99% or more homogeneity are most preferred. The purified proteins may also be used, *e.g.*, as immunogens for antibody production.

[0131] To facilitate purification of the recombinant mammalian GnT1 proteins of the invention, the nucleic acids that encode the recombinant eukaryotic glycosyltransferase 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 fusion proteins 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)).

[0132] Purification tags also include maltose binding domains and starch binding domains. Purification of maltose binding domain proteins is known to those of skill in the art. Starch binding domains are described in WO 99/15636, herein incorporated by reference. Affinity purification of a fusion protein comprising a starch binding domain using a betacylodextrin

(BCD)-derivatized resin is described in USSN 60/468,374, filed May 5, 2003, herein incorporated by reference in its entirety.

[0133] 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.

[0134] One of skill would recognize that modifications can be made to the GnT1 catalytic or functional domains without diminishing their biological activity. 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.

VI. Uses of refolded mammalian GnT1 protein

[0135] The invention provides recombinant mammalian GnT1 proteins and methods of using the recombinant mammalian GnT1 proteins to enzymatically synthesize glycoproteins, glycolipids, and oligosaccharide moieties, and to glycoPEGylate glycoproteins. The GnT1 reactions of the invention take place in a reaction medium comprising at least one mammalian GnT1 protein, acceptor substrate, and donor substrate, and typically a soluble divalent metal cation. The recombinant eukaryotic mammalian GnT1 proteins and methods of the present invention rely on the use the recombinant mammalian GnT1 proteins to catalyze the addition of a saccharide to an acceptor substrate.

[0136] A number of methods of using glycosyltransferases to synthesize glycoproteins and glycolipids having desired oligosaccharide moieties are known. Exemplary methods are

described, for instance, WO 96/32491, Ito *et al.* (1993) *Pure Appl. Chem.* 65: 753, and US Patents 5,352,670, 5,374,541, and 5,545,553.

[0137] The recombinant mammalian GnT1 proteins prepared as described herein can be used in combination with additional glycosyltransferases, that may or may not have required refolding for activity. For example, one can use a combination of refolded recombinant mammalian GnT1 protein and a bacterial glycosyltransferase, which may or may not have been refolded after isolation from a host cell. Similarly, the recombinant mammalian GnT1 protein can be used with recombinant accessory enzymes.

[0138] The products produced by the above processes can be used without purification. In some embodiments, oligosaccharides are produced. Standard, well known techniques, for example, thin or thick layer chromatography, ion exchange chromatography, or membrane filtration can be used for recovery of glycosylated saccharides. Also, for example, membrane filtration, utilizing a nanofiltration or reverse osmotic membrane as described in commonly assigned AU Patent No. 735695 may be used. As a further example, membrane filtration wherein the membranes have a molecular weight cutoff of about 1000 to about 10,000 can be used to remove proteins. As another example, nanofiltration or reverse osmosis can then be used to remove salts. Nanofilter membranes are a class of reverse osmosis membranes which pass monovalent salts but retain polyvalent salts and uncharged solutes larger than about 200 to about 1000 Daltons, depending upon the membrane used. Thus, for example, the oligosaccharides produced by the compositions and methods of the present invention can be retained in the membrane and contaminating salts will pass through.

[0139] 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 nucleic acid" includes a plurality of such nucleic acids and reference to "the polypeptide" includes reference to one or more polypeptides and equivalents thereof known to those skilled in the art, and so forth.

[0140] 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. All citations are incorporated herein by reference.

EXAMPLES

Example 1: Construction of GnT1 expression vector

[0141] A nucleic acid encoding the human β 1,3-*N*-acetylglucosaminyltransferase I (GnT1) protein was inserted into plasmid pCWin2-MBP and fused in frame to nucleic acid sequence encoding the maltose binding protein (MBP). The plasmid is named pCWin2-kan^r-MBP-GNT-1. The nucleic acid sequence of the pCWin2-kan^r-MBP-GNT-1 plasmid is shown in SEQ ID NO:12. The GnT1 protein had a truncation at the N-terminus and included amino acids 104 (Ala) to 445 (Asp) of the full length protein. An unpaired cysteine at position 121 was changed to a serine residue. The maltose binding protein sequence was obtained from plasmid pMal-c2X. The maltose binding protein and several amino acids from the polylinker including a Factor Xa cleavage site occur between the MBP and GnT-1 proteins. Translation of the GnT1 polypeptide is terminated by the presence of two stop codons. The translated GnT1 polypeptide is shown in Figure 1, including the two stop codons. SEQ ID NO:13 shows the same construct with the exception that only a single STOP codon is present at the end of the MBP-GNT1 coding sequence.

[0142] The pCWin2-kan^r-MBP-GNT-1 plasmid was transformed into the *E. coli* strain BNN93 for expression of the GnT1 protein. Expression of GnT1 from the double STOP version of the pCWin2-kan^r-MBP-GNT-1 plasmid resulted in 2-3 times the activity on a per gram of wet inclusion body weight as compared to an identical expression vector with a single STOP codon. See, below.

Example 2: Optimization of double stop MBP-GnT1 expression and refolding

[0143] To determine whether protein stabilizers such as glycerol or sucrose would improve MBP-GnT1 double stop refolding yields, 10% (w/v) glycerol and 0.4 M maltose was included in the refolding buffer, respectively. The refolding buffers used are presented in Table 1. Refolds were carried out at 2 mL scale, 4°C for approximately five hours. The samples were buffer exchanged to 50 mM MOPS pH 7.0, 50 mM NaCl by dialysis, and assayed for activity. As shown in Table 1, the presence of glycerol increased GnT1 activity by approximately 30%, while maltose had no effect.

	Control	Glycerol	Maltose
Refolding Buffer	50 mM Tris pH 8.56 550 mM Arginine 10 mM NaCl 1 mM MnCl ₂ 4 mM Cysteine 1 mM Cystamine 0.055% PEG3350	50 mM Tris pH 8.56 550 mM Arginine 10 mM NaCl 1 mM MnCl ₂ 4 mM Cysteine 1 mM Cystamine 0.055% PEG3350 10% (w/v) Glycerol	50 mM Tris pH 8.56 550 mM Arginine 10 mM NaCl 1 mM MnCl ₂ 4 mM Cysteine 1 mM Cystamine 0.055% PEG3350 0.4 M Maltose
Enzyme Activity (U/L)	9.9	12.7	9.5
Percentage	100%	128%	96%

Table 1: Refolding in the Presence of Maltose or Glycerol

[0144] Two rounds of refold screening and optimization were carried out at microplate scale for double stop MBP-GnT1. Inclusion bodies (IBs) were solubilized at 20 mg wet weight per mL solubilization buffer. IBs were solubilized in 8M urea, 50 mM Tris pH 8.5, 5 mM EDTA, 100 mM NaCl, and 10 mM DTT for 30 min RT on the rotator. Solubilized IBs were clarified by centrifugation at 17,900 x g, RT, for 2 minutes. Refolding was initiated by dilution of 50 μ L solubilized IB into 950 μ L refold buffer prepared in a 2 mL square deep well microplate. Refolding was carried out at 10°C overnight in a microplate shaker at approximately 350 rpm. After the refold incubation, 100 μ L samples or each refold were buffer exchanged using spin desalting columns or plates into GnT1 assay dilution buffer (0.1M MES pH 6, 0.1 mg/mL BSA), and assayed for activity. Refold buffer conditions specific to each run are given in Table 2.

Run:	Factor	Range tested	Fixed buffer
1. MBP-GnT1 double stop Screening (+/-) design. Full factorial. No replicates, 20 samples total.	Arginine and/or Proline	None, 750 mM either, 750 mM both	50 mM tris pH 8.6
	Glycerol or Sorbitol or Trehalose or Sucrose	None or 1M None or 1M None or 250 mM None or 250 mM	10 mM NaCl 1 mM cystamine 4 mM cysteine 1 mM MnCl ₂
2. MBP-GnT1 double stop RSM (optimization) design. Three points within each range Conditions in triplicate. 30 samples total.	Arginine	500 – 1250 mM	50 mM tris pH 8.6
	Glycerol	0 – 3 M (0 – 27.6 % w/v) (0 – 21.9 % v/v)	10 mM NaCl 1 mM cystamine 4 mM cysteine 1 mM MnCl ₂

Table 2: Specific Conditions for each Microplate Refold Screen/Optimization

[0145] Analysis of the recovered activities from the first round of microplate optimization found that all refolds lacking arginine were inactive, indicating that arginine was important for GnT1 refolding ($p < 0.01$). By contrast, proline, either alone or in conjunction with arginine, did not have a significant effect. Of the stabilizers, sorbitol demonstrated a clear negative effect on refold yields ($p < 0.1$), reducing recovered activity by half versus similar refolds without sorbitol. Sucrose did not have a significant effect, while glycerol and trehalose may have had a mild positive impact on refold yields, albeit at levels that were not statistically significant ($p > 0.2$).

[0146] The final microplate optimization experiment refined factor concentrations.

Arginine concentrations from 500 – 1250 mM were examined, and glycerol concentrations from 0 – 3M were also tested. The factors were combined in a three-level, full factorial RSM design with three additional center points, for a total of thirty runs. Analysis of the activities at the different conditions determined that a maximal response was observed. The optimal concentrations of the tested factors were calculated to maximize activity yield at approximately 875 mM arginine, and 1.65 M glycerol. Compared to a GnT1 refold reaction using buffer similar to the control buffer in Table 1, the optimal refold buffer more than doubled the GnT1 refold activity yield. Scale up of the refolding procedure to, *e.g.*, 10L or 35L, resulted in similar purity and activity yields after purification.

Example 3: Comparison of MBP-GnT1 single stop to double stop after 30Q chromatography

[0147] Refolding of MBP-GnT1 inclusion bodies (IBs): The IBs were solubilized at 20 mg/ml in 8 M urea, 50 mM Tris HCl (pH 8.5), 5 mM EDTA, 100 mM NaCl, and 10 mM DTT at room temperature with constant stirring until no significant pellets were observed. The suspension was then centrifuged at 27,000 x g, 4°C, for 20 minutes. The supernatant was further clarified by filtration through a 0.45 µm membrane. The refolding reaction was initiated by diluting the solubilized IBs 1:20 dropwise into a partially optimized refolding buffer containing 850 mM Arginine, 50 mM Tris HCl, 10% glycerol, 10 mM NaCl, pH 8.56, 4 mM cysteine, 1 mM cystamine, and 1 mM MnCl₂. The 200 mL refolding reaction was carried out at 10°C for 18-24 hours with gentle stirring or shaking. The refolded protein was buffer exchanged into 20 mM Tris HCl (pH 8.0) 5 mM NaCl by tangential flow filtration (TFF). The TFF was performed a total of five times by diluting the refolding solution with an equal volume of the desired buffer followed by concentrating to the original volume. The TFF buffer and the protein solution were kept cold on ice.

[0148] The diafiltered and concentrated refold reactions were separated by anion exchange column chromatography on Source 30Q column (~9.42 mL, 1 x 12 cm) at 4 ml/min. A gradient was generated between Buffer A (20 mM Tris-HCl, pH 8.0) and Buffer B (20 mM Tris-HCl, pH 8.0, 1 M NaCl). The protein was eluted in 75 min at 4 ml/min using linear
5 gradient of 0-500 mM NaCl in the equilibration buffer. The elution profiles, shown in Figure 2, were generally similar. However, there was a notable difference in the symmetry of the monomer peak. Specifically, the monomer peak for the double stop construct (Figure 2A) was more symmetric, whereas the monomer peak for the single stop construct (Figure 2B) was shouldered, indicating multiple populations. Further, analysis of the monomer peak by
10 SDS-PAGE and silver stain showed that the GnT1 protein expressed from the double stop construct ran as a single protein band. By contrast, the GnT1 protein expressed from the single stop construct had two major protein bands.

[0149] The monomer pools from the double and single STOP constructs were assayed for protein concentration and GnT1 activity. The yield of refolded double STOP GnT1 as mg
15 per g IB wet weight (Figure 4A) was significantly improved versus single STOP construct. Notably, the yield of active GnT1 as Units per g IB wet weight (Figure 4B) was more than doubled with the double STOP construct by comparison to the single STOP construct.

Example 4: Comparison of GnT1 activity of single and double STOP constructs after refolding using optimized conditions.

[0150] Refolding of MBP-GnT1 inclusion bodies (IBs): The IBs were solubilized 20
20 mg/ml in 8 M urea, 50 mM Tris HCl (pH 8.5), 5 mM EDTA, 100 mM NaCl, and 10 mM DTT at room temperature with constant stirring until no significant pellets were observed. The suspension was then centrifuged at 27,000 x g, 4°C, for 20 minutes. The supernatant was further clarified by filtration through a 0.45 µm membrane. The refolding reaction was
25 initiated by diluting solubilized IBs 1:20 dropwise into an optimized refolding buffer containing 875 mM Arginine, 50 mM Tris HCl, 1.65 M glycerol, 10 mM NaCl, pH 8.56, 4 mM cysteine, 1 mM cystamine, and 1 mM MnCl₂. The 200 mL refolding reaction was carried out at 10°C for 18-24 hours with gentle stirring or shaking.

[0151] An aliquot of the refolding reactions were applied to a PD-10 desalting column,
30 previously equilibrated with 25 ml buffer containing 20 mM Tris HCl (pH 8.0) and 10 mM NaCl. Two and half milliliters of filtered refolding solution were loaded onto the column and eluted with 3.5 ml of the equilibration buffer. The eluted material was then assayed for GnT1

activity. The yields of active GnT1 protein expressed as Units per liter of refold (Figure 5A), or Units per g IB wet weight (Figure 5B) were increased more than two fold with the double stop construct versus the single stop construct.

Example 4: Purification of the GnT1 protein using additional chromatography steps.

5 **Preparation of the 0.5 M and 10 mM Phosphate Buffers.**

[0152] Sixty-nine grams of $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ [MW 137.99] was dissolved in 800 mL MilliQ H_2O . The pH of the solution was adjusted to 8.0 with 50% NaOH. The volume of the solution was then adjusted to 1000 mL in a graduated cylinder with MilliQ H_2O . This yielded a 0.5 M phosphate buffer (pH 8.0). The preparation of 10 mM phosphate buffer (pH 8.0) was carried
10 out by adding 20 mL of the 0.5 M phosphate solution prepared above to 980 mL of MilliQ H_2O . No further pH adjustment was performed after dilution. Both solutions were filtered through a 0.22 μm membrane filter and stored at 22 °C (room temperature).

Packing of the Hydroxyapatite and Fluoroapatite Columns.

[0153] The desired resin was weighed out according to 0.6 g resin powder per mL of bed
15 volume. The resin was then gently mixed 1:1 (w/v) with 0.5 M phosphate buffer (pH 8.0). Caution was taken to avoid any vigorous mixing that could cause the disruption of resin beads. The suspended resin was poured slowly into a clean empty column along the inner wall. The resin settled while washing the column with 0.5 M phosphate buffer (pH 8.0) at 10 mL/min. When the bed height became constant, the column was equilibrated with 10 column
20 volumes of 10 mM phosphate buffer (pH 8.0).

Purification of MBP-GnT-1 using Different Hydroxyapatite and Fluoroapatite Resins.

[0154] The detailed purification procedures for each of the chromatography resins are below. As an example of the procedure, ten milliliters (0.097 mg protein/mL) of pooled Source 30Q fractions prepared from a 10-L refold reaction that had been stored at -80 °C in
25 50% glycerol, was thawed and diluted to 50 mL with 10 mM sodium phosphate to a conductivity of less than 2 mS/cm. The solution was then filtered through a 0.45 μm syringe filter and loaded onto a column packed with the desired resin. The column was washed with 2 CV's of 10 mM sodium phosphate (pH 8.0). The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate in 20 min at a flow rate of 2 mL/min (20 CV's).
30 The elution of protein was monitored by $\text{OD}_{280\text{nm}}$. The eluted fractions were analyzed by SDS-PAGE. The fractions containing predominantly MBP-GnT-1 were pooled and submitted for activity assay and RP-HPLC analysis.

MBP-GnT-1 Enzyme Activity Assay.

[0155] Two methods were used to analyze the activity of MBP-GnT-1, a radioactive assay and a continuous spectrophotometric assay. The spectrophotometric assay provided more consistent activity data so it was developed to replace the radioactive assay.

- 5 [0156] Radioactive assay: Briefly, this assay measures the transfer of [^3H]N-acetylglucosamine (GlcNAc) from UDP-[^3H]GlcNAc to a synthetic acceptor N-octyl 3,6-Di-O-(α -mannopyranosyl) β -D-mannopyranoside (OM3), a trimannosyl core with an octyl tail. Enzymatic activity is terminated by diluting the reaction mixture with water, and the extent of reaction is determined by separating the radioactive donor from the acceptor and product on a
10 reversed phase resin, which binds the hydrophobic OM3 and [^3H]GlcNAc-OM3. The water soluble UDP-[^3H]GlcNAc and [^3H]GlcNAc impurity are washed through the resin using aqueous buffer and the radiolabeled product is eluted with methanol. The radioactive product eluted with MeOH is measured on a scintillation counter.

- [0157] A continuous coupled spectrophotometric assay: In this assay, a donor substrate
15 (UDP-GlcNAc) provides GlcNAc to be transferred by MBP-GnT-1 to an acceptor substrate, N-Octyl-trimannopyranoside (OM3). The released uridine diphosphate (UDP) is measured by two coupled enzymes, pyruvate kinase (PK) and lactate dehydrogenase (LDH) using phosphoenolpyruvate (PEP) and NADH as substrates. The oxidation of NADH, is directly proportional to the UDP concentration and thus to the activity of GnT-1 present, and is
20 monitored by absorbance at 340 nm. The activity is reported in Units/Liter of enzyme solution where 1 Unit is defined as the amount of GnT-1 required to transfer 1 μmol of GlcNAc from donor UDP-GlcNAc to acceptor OM3 per minute under the conditions of this assay.

Analysis of the Purity and Quantity of MBP-GnT-1 by RP-HPLC.

- 25 [0158] The MBP-GnT-1 sample was fractionated on a Vydac 214TP54 C4 (250 X 4.6 mm) column by a linear gradient of 35~70% phase B in 20 min with a flow rate of 1.0 mL/min at 35 $^{\circ}\text{C}$. The mobile phase A was 0.1% TFA/MilliQ water and Mobile phase B was 0.1% TFA/95% Acetonitrile. The MBP-GnT-1 monomer was eluted at about 52.5% phase B. The relative concentration of MBP-GnT-1 was estimated using a standard curve prepared with
30 BSA.

SDS-PAGE Analysis.

[0159] NuPAGE precast gels (4-12%, 1 mm) from Invitrogen were used. Three volumes of protein solution were mixed with 1 volume of NuPAGE LPS sample buffer (4X, Invitrogen

NP0007) and denatured at 85 °C for 5 min. When reduction of the protein is required, 10% (v/v) of DTT stock solution (1 M) was added to the mixture and heated as mentioned above. Twenty to thirty microliters of the denatured mixture were loaded in each well. The gel was run with 150 volts at room temperature in MES buffer for 1 hr and stained with SimplyBlue or Silver stain II kit according to the manufacturers' instructions.

Column Capacity Test.

[0160] A 1-mL column (5.0 mm X 5.5 cm) was packed with Hydroxyapatite Type I (HA Type I). Both 1 mg and 5 mg quantities of total protein from the pooled Source 30Q fractions (Batch 061208) were each diluted 5 fold with 10 mM sodium phosphate to a conductivity of less than 2 mS/cm and loaded onto separate columns. The column was washed with 5 CV's of 10 mM sodium phosphate (pH 8.0). The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate in 20 min at a flow rate of 1 mL/min. The elution of protein was monitored by A_{280nm}. The fractions were analyzed by SDS-PAGE.

[0161] Results of initial experiments comparing HA Type I, HA Type II, and fluoroapatite resins. Detailed methods of running the columns are provided below.

[0162] Hydroxyapatite Type I was packed into a 6.6 mm X 5.5 cm, CV ~3.08 ml. A 10 ml Source 30Q pool (stored at -80 °C containing 50% glycerol) was diluted with 40 ml of Buffer A (10 mM sodium phosphate, pH 8.0) and loaded at 2 ml/min onto the column pre-equilibrated with the same buffer. Protein was eluted over 20 minutes (flow rate of 2 ml/min) with a 0-100% gradient of Buffer B (200 mM sodium phosphate, pH 8.0). Elution of protein was monitored by measuring UV absorbance at 280 nm.

[0163] Hydroxyapatite Type II was packed into a 6.6 mm X 7 cm, CV ~2.39 ml. A 10 ml Source 30Q pool (stored at -80 °C containing 50% glycerol) was diluted with 40 ml of Buffer A (10 mM sodium phosphate, pH 8.0) and loaded at 2 ml/min onto the column pre-equilibrated with the same buffer. Protein was eluted over 20 minutes (flow rate of 2 ml/min) with a 0-100% gradient of Buffer B (200 mM sodium phosphate, pH 8.0). Elution of protein was monitored by measuring UV absorbance at 280 nm.

[0164] Fluoroapatite Type II was packed into a 6.6 mm X 5.5 cm, CV ~1.88 ml. A 10 ml Source 30Q pool (stored at -80 °C containing 50% glycerol) was diluted with 40 ml of Buffer A (10 mM sodium phosphate, pH 8.0) and loaded at 2.5 ml/min onto the column pre-equilibrated with the same buffer. Protein was eluted over 20 minutes (flow rate of 2 ml/min)

with a 0-100% gradient of Buffer B (200 mM sodium phosphate, pH 8.0). Elution of protein was monitored by measuring UV absorbance at 280 nm.

[0165] A comparison of the results of the three columns is provided in Figure 7.

Comparisons of the eluted proteins are shown in Figure 8. All the resins tested improved the purity of the MBP-GnT1. However, the hydroxyapatite Type I chromatography resin provided the best recovery results based on protein mass and activity.

Hydroxyapatite TYPE I Purification of MBP-GnT-1 Produced from a 15 L Fermentation.

Batch 1.

[0166] A column was prepared using Hydroxyapatite Type I, 20 μ m (16 mm X 18 cm, CV ~36 mL). The pooled Source 30Q fractions were diluted with 10 mM sodium phosphate (pH 8.0), filtered through 0.22 μ m membrane and loaded onto the column. The pooled Source 30Q fractions had 45.97 mg of total protein in 285 mL of buffer and was stored in 50% glycerol at -80°C. Batch 061208 was diluted to 1425 mL with 10 mM sodium phosphate (pH 8.0) prior to loading. The conductivity of the load solution was 2 mS/cm. The flow rate during sample loading was set at 10 mL/min. The column was washed with 2 CV's of 10 mM sodium phosphate (pH 8.0). The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate over 20 CV's at a flow rate of 10 mL/min. The elution of protein was monitored by A_{280nm} . The fractions were analyzed by SDS-PAGE. The fractions containing predominantly MBP-GnT-1 were pooled. The pooling started at 17.9 % Buffer B where A_{280} was 10% of that for the peak maximal (A_{280max}) and ended at 23.8% Buffer B where A_{280} was 7.8% of A_{280max} . The pooled fractions were submitted for activity assay, RP-HPLC analysis and endotoxin analysis. (Figure 9).

Batch 2

[0167] A column was prepared using Hydroxyapatite Type I, 20 μ m (16 mm X 18 cm, CV ~36 mL). The pooled Source 30Q fractions were diluted with 10 mM sodium phosphate (pH 8.0), filtered through 0.22 μ m membrane and loaded onto the column. The pooled Source 30Q fraction had 39.03 mg of total protein in a 320 mL buffer volume was stored in 50% glycerol at -80°C. The solution was diluted to 1200 mL with 10 mM sodium phosphate (pH 8.0) prior to loading. The conductivity of the load solution was 2 mS/cm. The flow rate during sample loading was set at 10 mL/min. The column was washed with 2 CV's of 10 mM sodium phosphate (pH 8.0). The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate over 20 bed volumes at a flow rate of 10 mL/min. The elution of

protein was monitored by $A_{280\text{nm}}$. The fractions were analyzed by SDS-PAGE. The major peak fractions containing predominantly MBP-GnT-1 were pooled. The pooling started at 18.5 % Buffer B where A_{280} was 19.7% of that for the peak maximal ($A_{280\text{max}}$) and ended at 24.2% Buffer B where A_{280} was 7.9% of $A_{280\text{max}}$. The pooled fractions were submitted for activity assay, RP-HPLC analysis and endotoxin analysis. (Figure 9).

Batch 3

[0168] This batch started with freshly pooled Source 30Q fractions and didn't contain any glycerol while Batch 1 and 2 started with a protein solution in 50% glycerol. A column was prepared using Hydroxyapatite Type I, 20 μm (16 mm X 18 cm, CV ~36 mL). The pooled Source 30Q fractions were diluted with 10 mM sodium phosphate (pH 8.0), filtered through 0.22 μm membrane and loaded onto the column. Batch 070215 contained 22.44 mg of total protein in 235 mL of buffer but was not formulated with glycerol. The enzyme from the Q-sepharose chromatography step was used directly for this step. The enzyme solution was diluted to 1175 mL with 10 mM sodium phosphate (pH 8.0) and added directly to the column. The conductivity of the load solution was 3.4 mS/cm. The flow rate during sample loading was set at 10 mL/min. The column was washed with 2 CV's of 10 mM sodium phosphate (pH 8.0). The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate over 20 bed volumes at a flow rate of 10 mL/min. The elution of protein was monitored by $A_{280\text{nm}}$. The fractions were analyzed by SDS-PAGE. The major peak fractions containing predominantly MBP-GnT1 were pooled. The pooling started at 17.6 % Buffer B where A_{280} was 7% of that for the peak maximal ($A_{280\text{max}}$) and ended at 23.2% Buffer B where A_{280} was 4% of $A_{280\text{max}}$. The pooled fractions and submitted for activity assay, RP-HPLC analysis and endotoxin analysis. (Figure 9).

[0169] A comparison of the results of the three batches is provided in Figure 9. The enzyme from the Source 30Q chromatography step of three separate process runs all provided similar purification results as demonstrated in Figure 9. Using this method, a majority of the smaller molecular weight impurities were removed. The purity of the MBP-GnT-1 was significantly increased and the method provided good recovery of both protein mass and enzyme activity. The endotoxin level decreased approximately 20 fold, from 611 EU to 31 EU/mg total protein.

Recommended Protocol for Hydroxyapatite (Type I) Purification of MBP-GnT-1 produced from a 15 L fermentation

[0170] A column was packed using Hydroxyapatite Type I (20 μ m). The diameter of the column was 16 mm and the bed height was 18 cm. The column volume was approximately 36 mL. The column was cleaned with 1 M NaOH at 2 mL/min for 1 hr followed by washing with Buffer A (10 mM sodium phosphate, pH 8.0) at 10 mL/min until pH 8.0. The column was then cleaned with 6 M Guanidine HCl at 2 mL/min for 30 min. The column was flushed with Buffer A at 10 mL/min for 20 min and regenerated with Buffer B (0.5 M sodium phosphate, pH 8.0) at 10 mL/min for 10 min. The column was re-equilibrated with Buffer A at 10 mL/min for 36 min. Pooled Source 30Q fractions were diluted 5 fold with Buffer A to reach a conductivity of less than 3.4 mS/cm, filtered through 0.45 μ m filter and loaded at 10 mL/min onto the column. The column was washed with 2 CV's of Buffer A at 10 mL/min. The MBP-GnT-1 was eluted using a linear gradient of 10-200 mM sodium phosphate (0~40% Buffer B) over 20 CV's at a flow rate of 10 mL/min. The elution of protein was monitored by OD_{280nm} and fractions collected (8 mL/fraction). The fractions were analyzed by SDS-PAGE. The fractions containing predominantly MBP-GnT-1 were pooled. The pooling typically started at 17.9 % Buffer B where the absorbance was 51.1 mAU and ended at 23.8% Buffer B where the absorbance was 39.7 mAU. The pooled fractions were submitted for activity assay, SEC, RP-HPLC analysis and endotoxin analysis. The pooled fractions were stored at 4 °C until next step of process.

Example 5: Buffer exchange of purified the GnT1 protein.

[0171] A G25 chromatography process was used to buffer exchange the MGnT-1 HA pool into 50 mM Tris, 500 mM NaCl, pH 8.2 as the final formulation buffer.

METHODS

[0172] A G25 gel chromatography step was developed as the last step of the MGnT1 purification process. The process steps are illustrated in Figure 10. MGnT1 was expressed in BNN93 cells using E 1.0 media. Cells were lysed and inclusion bodies were prepared. The MGnT1 inclusion bodies (IBs) were solubilized in a urea buffer and refolded in a complex refold buffer at pH 8.6. Refolded material was concentrated 10-fold and buffer exchanged by TFF. The TFF retentate was purified using Source 30Q chromatography step followed by a hydroxyapatite type I chromatography step. A G25 column was used to buffer exchange the enzyme into the final formulation buffer of 50mM Tris/ 500mM NaCl pH 8.2 and the material was sterile filtered.

Packing of the G25 Chromatography Column.

[0173] Dry powder G25 resin (medium) was allowed to swell in RO water overnight and the supernatant was decanted from the settled resin to remove fines. The resin was re-slurried in 20% ethanol (70 % resin, v/v) and packed in an XK 50 column to a bed height of 29.5 cm (CV= 579 mL).

G25 Chromatography Process.

[0174] An AKTA purifier system was used to perform the G-25 chromatography. The column, packed in 20% ethanol/water, was washed with 0.5 M NaOH (1 CV) and RO water (1 CV). Then it was equilibrated using a buffer composed of 50 mM Tris pH 8.2, 500 mM NaCl (2 CVs).

[0175] The enzyme (13.4 mg) from the HA chromatography step that contained 100 mL of buffer (102 mM NaPO₄, pH 8.0) was loaded onto the column at a flow rate of 20 mL/min. The injection volume of this step should be kept at ≤20% of the G25 column volume. After the sample was loaded onto the column, the product was eluted using 50 mM Tris pH 8.2, 500 mM NaCl at a flow rate of 20mL/min. When the absorbance of the column eluant reached 0.8 mAU at 280 nm, the flow rate was reduced to 5 mL/min and 8 mL fractions were collected. The MGnT-1 product eluted after 0.11 CV's as a broad peak of 0.24 CV. The MGnT-1 peak fractions were pooled, beginning from the first fraction of the leading edge of the peak starting at 0.8 mAU and ending when the absorbance fell below 25 % of the peak max at the tailing edge of the peak, providing a total volume of 121 mL. The buffer components, including sodium phosphate originating from the HA chromatography step, were observed to elute at 0.7 CV as indicated by the change in conductivity. The entire process was complete within 60 minutes after injection of the sample.

Sterile Filtration of the MGnT-1.

[0176] The G25 chromatography fraction pool was sterile filtered using a Millex GV (0.22 µm, SLGVR25KS) filter unit under aseptic conditions. Aliquots of 8 mLs were placed in 15 mL sterile Fisher brand polypropylene culture tubes. The aliquots were frozen and stored at -70 °C. The purified enzyme was analyzed by SDS-PAGE (Figure 11) and found to be predominantly one band. The purity by RP-HPLC was >98% and the overall step yield based on enzyme activity was 87% (Figure 12). The overall process results are summarized in Figure 12.

A G25 Sephadex chromatography step was used to buffer exchange the MGnT-1 that originated from the HA purified enzyme into 50 mM TrisHCl, 500 mM NaCl pH 8.2. After

sterile filtration, this produced an enzyme concentration of 0.098 mg/mL. A step yield based on enzyme activity was 87% and the step yield based on protein was 88%. A purity of >98% was obtained as determined by RP-HPLC. After sterile filtration, the enzyme was stored at -70°C.

- 5 [0177] 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
- 10 purposes.

WHAT IS CLAIMED IS:

1. A large scale method of producing an active mammalian *N*-acetylglucosaminyltransferase I (GnT1) protein in bacteria, the methods comprising the steps of

5 a) expressing the mammalian GnT1 protein in bacteria as an insoluble protein,
b) solubilizing the insoluble mammalian GnT1 protein;
c) refolding the soluble mammalian GnT1 protein; and
d) purifying the soluble mammalian GnT1 protein, thereby producing the active mammalian GnT1 protein in bacteria.

10 2. The method of claim 1, wherein the soluble mammalian GnT1 protein is purified using at least one step selected from the group consisting of anion exchange chromatography, hydroxyapatite chromatography, and gel filtration chromatography.

15 3. An expression plasmid for expression of a eukaryotic *N*-acetylglucosaminyltransferase I (GnT1) protein in bacteria, wherein a nucleic acid that encodes the GnT1 protein terminates translation of the GnT1 protein with two contiguous stop codons.

4. The expression plasmid of claim 3, wherein the GnT1 protein is a human GnT1 protein.

20 5. The expression plasmid of claim 3, wherein the expression plasmid has the sequence of SEQ ID NO:12.

6. A host cell that comprises the expression plasmid of claim 3.

7. A method of making a GnT1 protein, the method comprising growing the host cell of claim 6 under conditions suitable for expression of the GnT1 protein.

25 8. The method of claim 7, further comprising the steps of refolding and solubilizing the GnT1 protein.

9. The method of claim 7, further comprising the step of purifying the GnT1 protein.

10. The method of claim 9, wherein the GnT1 protein is purified using at least one step selected from the group consisting of anion exchange chromatography, hydroxyapatite chromatography, and gel filtration chromatography.

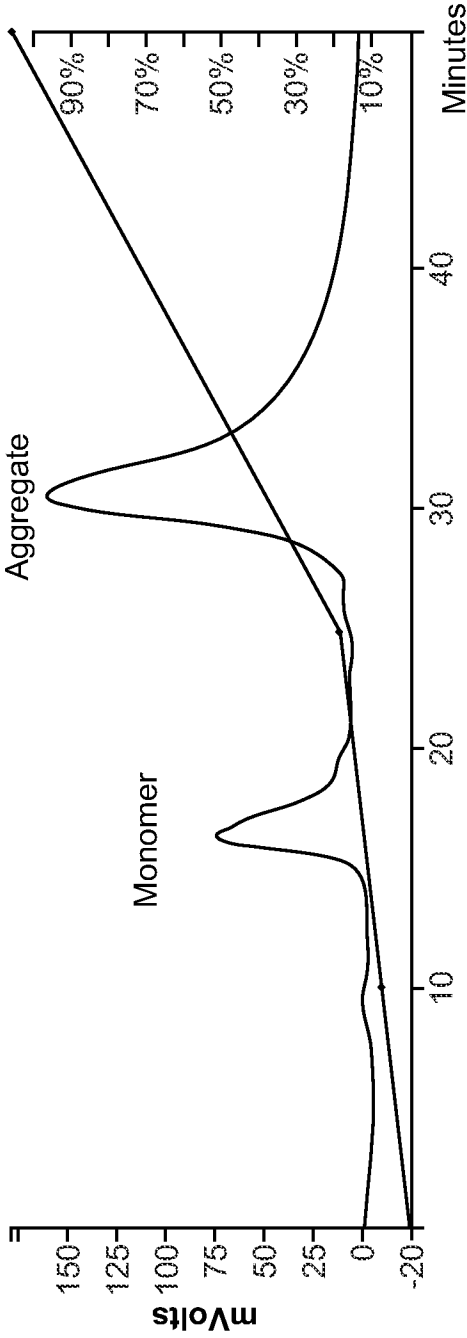
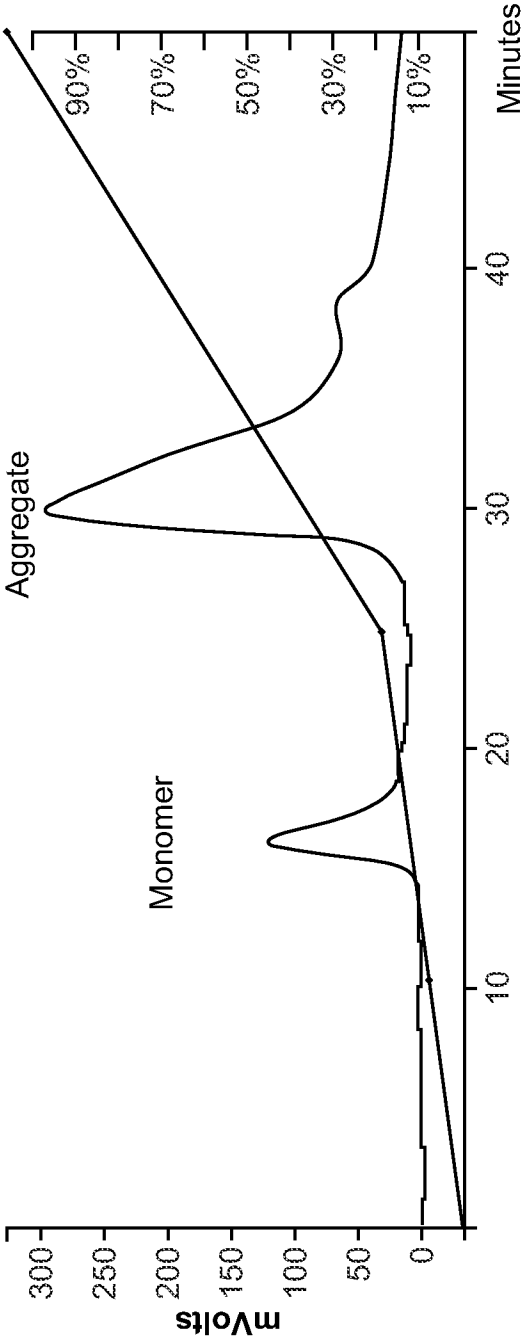
1 / 12

Complete peptide sequence of double STOP MBP-GNT-1:

mkieegklviwingdkgynglaevgkkfekdtgikvtvehpdkleekfpqvaatgdgpdiiifwahdrfggyaqsgll
 aeitpdkafqdklypftwdavryngkliaypiavealsliynkdllpnppktweeipaldkelkakgksalmfnlqepy
 ftwpliaadggyafkyengkydikdvgnagakagltflvdliknkhmnadt dysiaaeaafnkgetamtingpwa
 wsnidtskvnvgvtvlptfkqgqpskpfvgvlsaginaaspnkelakeflenylltdegleavnkdkplgavalksyeee
 lakdpriaatmenaqqgeimpnipqmsafwyavrtavinaasgrqtvdealkdaqtnssnnnnnnnnnlgiegri
 sefgsaaaavipilviacdrtvrrslkllhyrpsaelfpilvsqdcgheetaqaiasygsavthirqpdlssiaavppdhrk
 fqgyykiarhyrwalgqvfrqfrfpaavvveddleapdffeyfratypllkadpslwcvsawndngkeqmvdsarp
 ellyrtdffpglgwlllaelwaelepkwpkafwddwmrrpeqrqgracirpeisrtmtfgrkgvshgqffdkhlfikl
 nqgfvhftqldlsylqreaydrdflarvygapqlqvekvrtndrkelgevrvqytgrdsfkafakalgvmddllksgvpr
 agyrgivtfqfpgrrrvhlappptwegydpwn**

**= Double STOP codons

FIG. 1



3 / 12

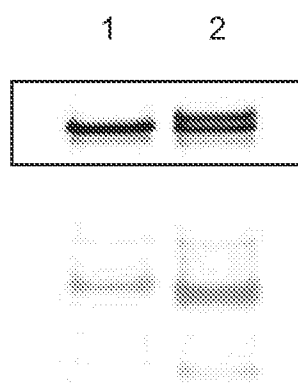


FIG. 3

4 / 12

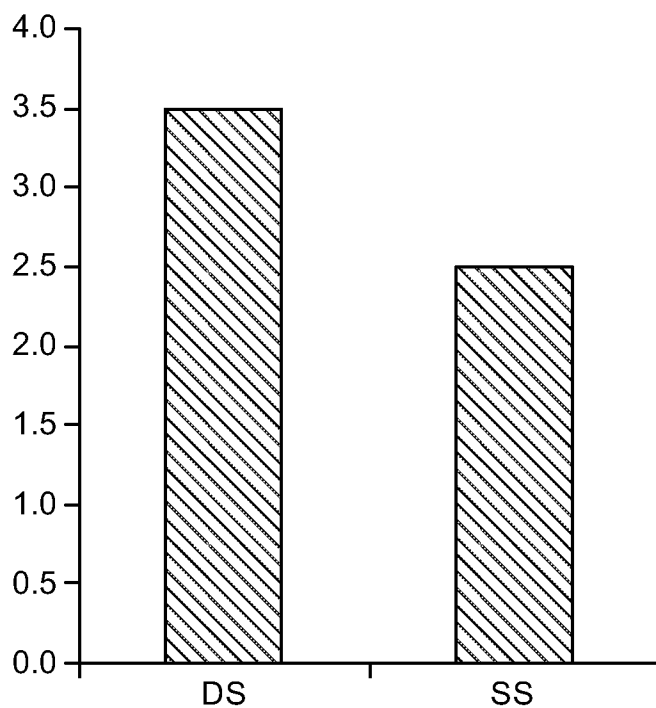


FIG. 4A

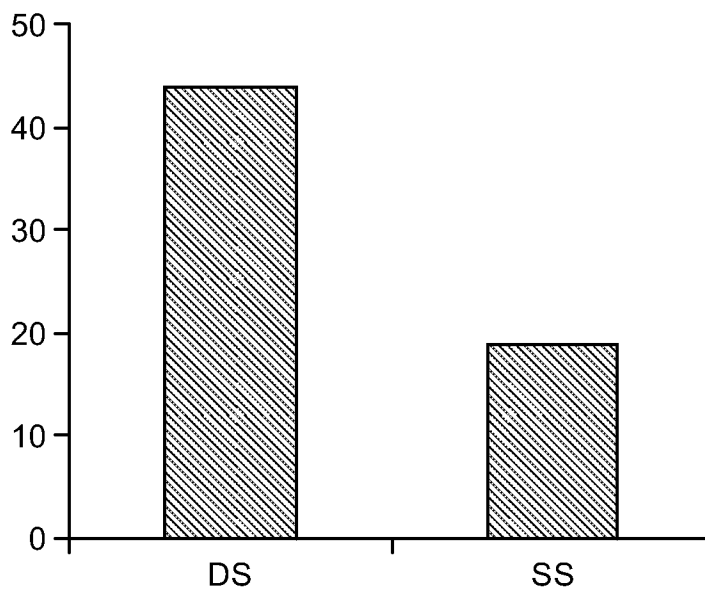


FIG. 4B

5 / 12

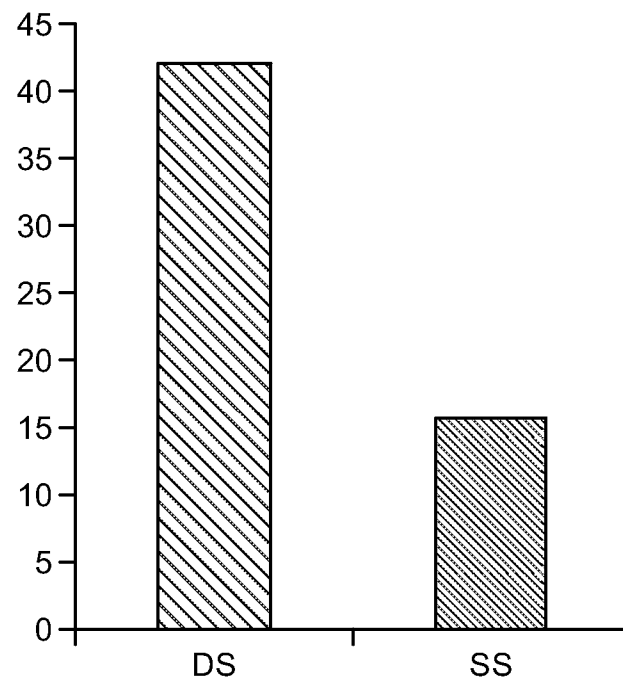


FIG. 5A

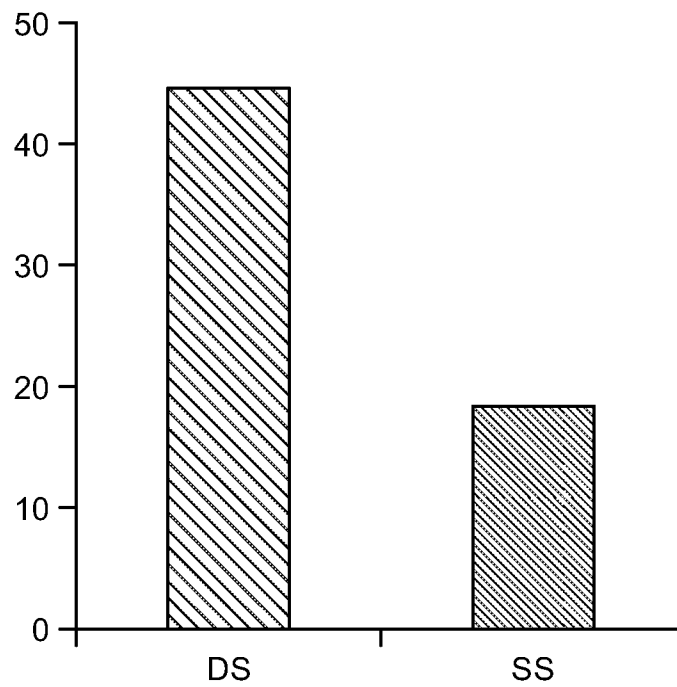


FIG. 5B

6 / 12

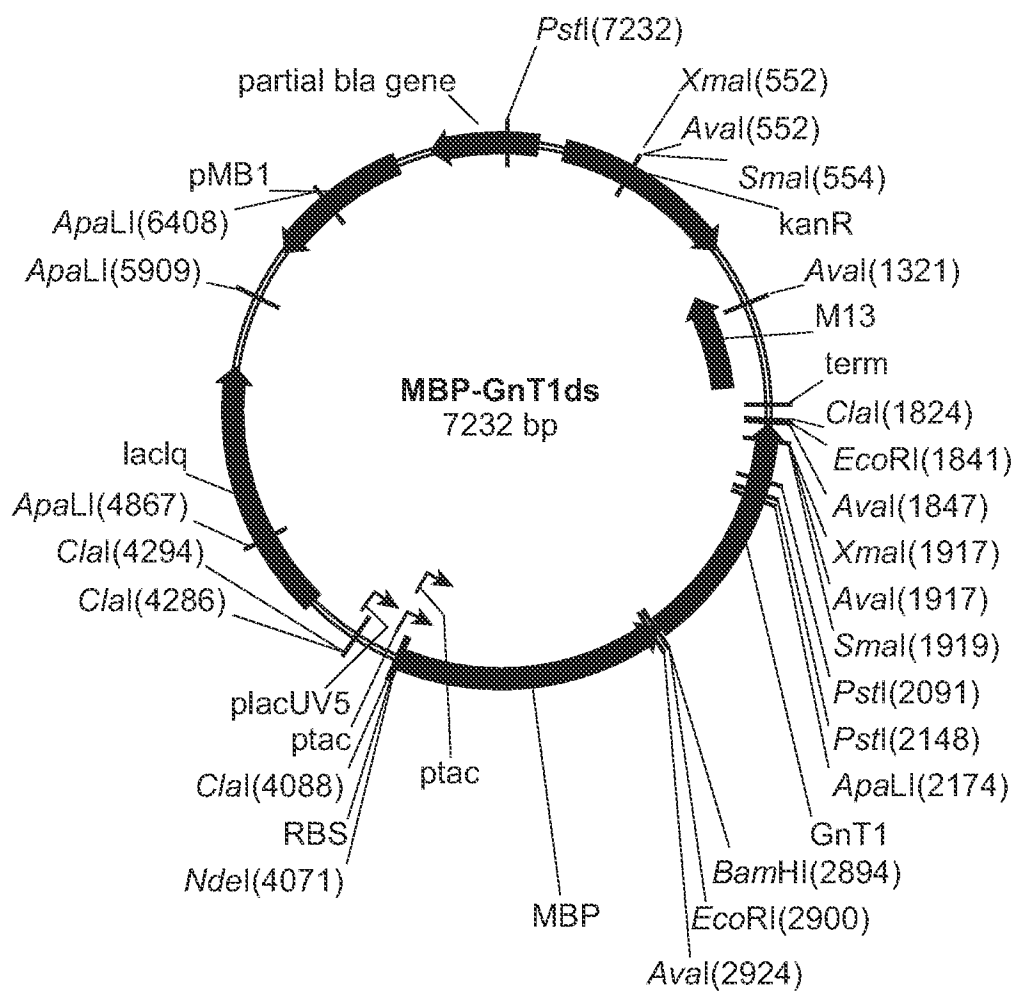


FIG. 6

7 / 12

Resin /pH	Source 30Q pool (10 mL)		Hydroxyapatite pool		Recovery	
	Mass (Purity)	Activity* (Specific activity**)	Mass (Purity)	Activity* (specific activity**)	Mass	Activity
HA Type I /pH 8	0.97 mg (89%)	4.02 U (4.14 U/mg)	0.79 mg (99%)	3.55 U (4.49 U/mg)	81%	92%
HA Type I /pH 9	0.97 mg (89%)	4.02 U (4.14 U/mg)	0.76 mg (98%)	3.31 U (4.36 U/mg)	78%	82%
HA Type II /pH 8	0.97 mg (89%)	4.02 U (4.14 U/mg)	0.77 mg (98%)	3.34 U (4.34 U/mg)	79%	83%
FA Type II /pH 8	0.97 mg (89%)	4.02 U (4.14 U/mg)	0.66 mg (99%)	2.76 U (4.18 U/mg)	68%	69%

FIG. 7

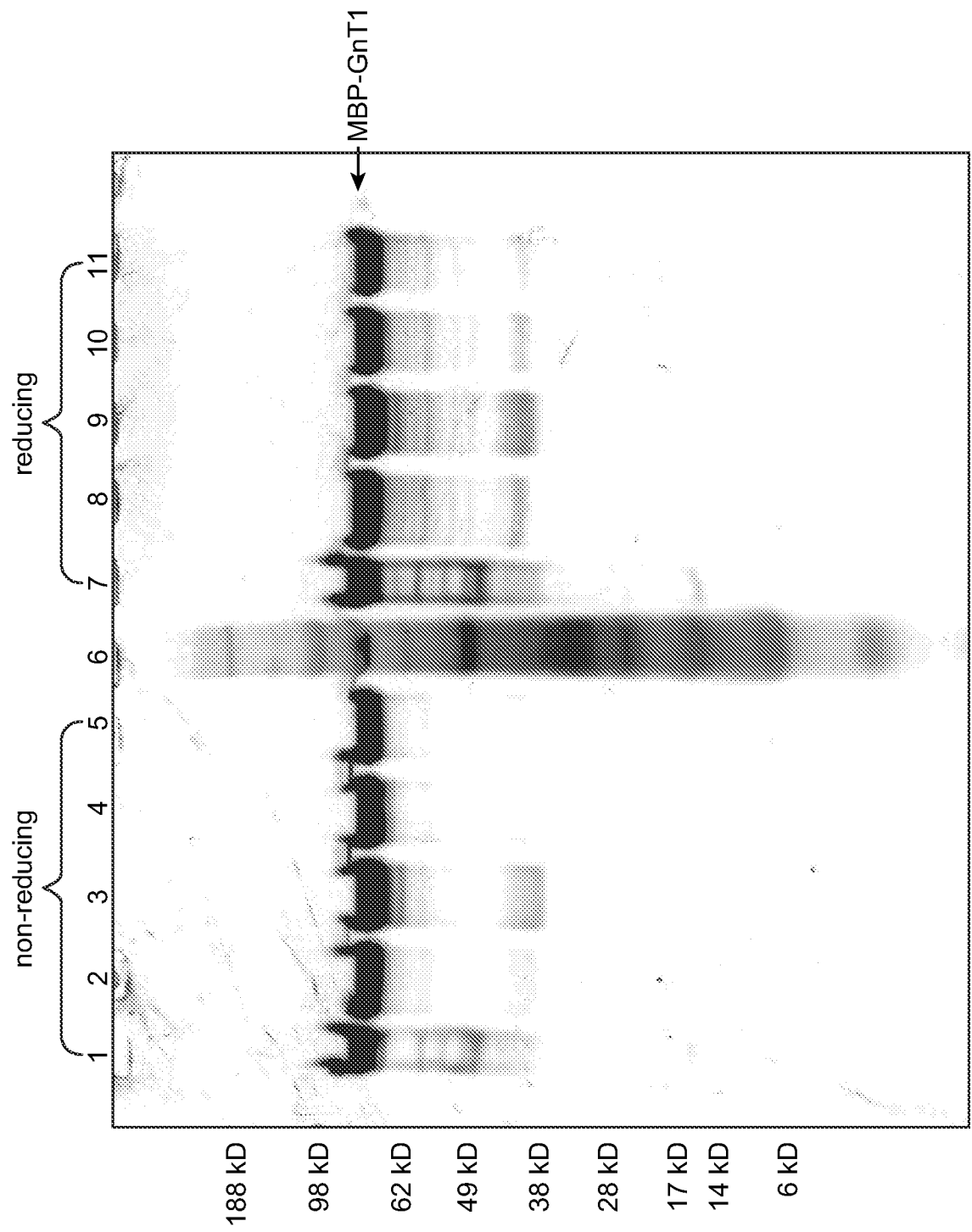


FIG. 8

9 / 12

**THE PURIFICATION OF MBP-GnT-1 BY HYDROXYAPATITE TYPE I
FROM THREE SEPARATE 15 L PRODUCTION RUNS**

Batch	Source 30Q pool			Hydroxyapatite pool			Recovery	
	Mass (Purity)	Activity* (Specific activity**)	Endotoxin	Mass (Purity)	Activity* (specific activity**)	Endotoxin	Mass	Activity
1 (061208)	45.97 mg (93%)	107.56 U (2.34 U/mg)	611 Eu/mg	36.50 mg (100%)	99.34 U (2.72 U/mg)	31 Eu/mg	79%	92%
2 (061212)	39.03 mg (91%)	134.85 U (3.46 U/mg)	ND	31.82 mg (100%)	124.78 U (3.92 U/mg)	ND	82%	93%
3 (070215)	22.44 mg (86%)	68.6 U (3.06 U/mg)	ND	15.5 mg (99.4%)	62.3 U (4.02 U/mg)	ND	69%	91%

* The MBP-GnT1 activity was analyzed using a spectrophotometric assay.

** The specific activity was defined as Units/mg total protein.

ND: Not Determined

FIG. 9

10 / 12

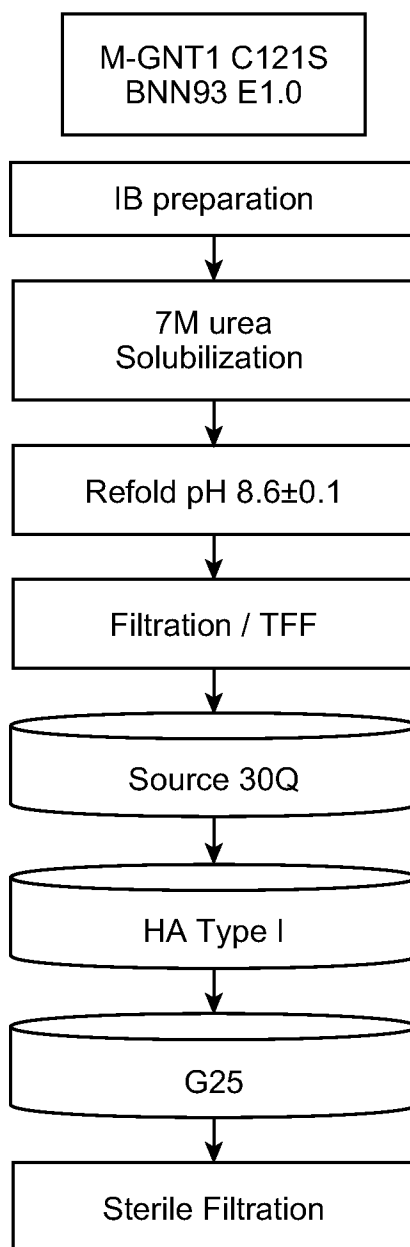
GNT1 Process

FIG. 10

11 / 12

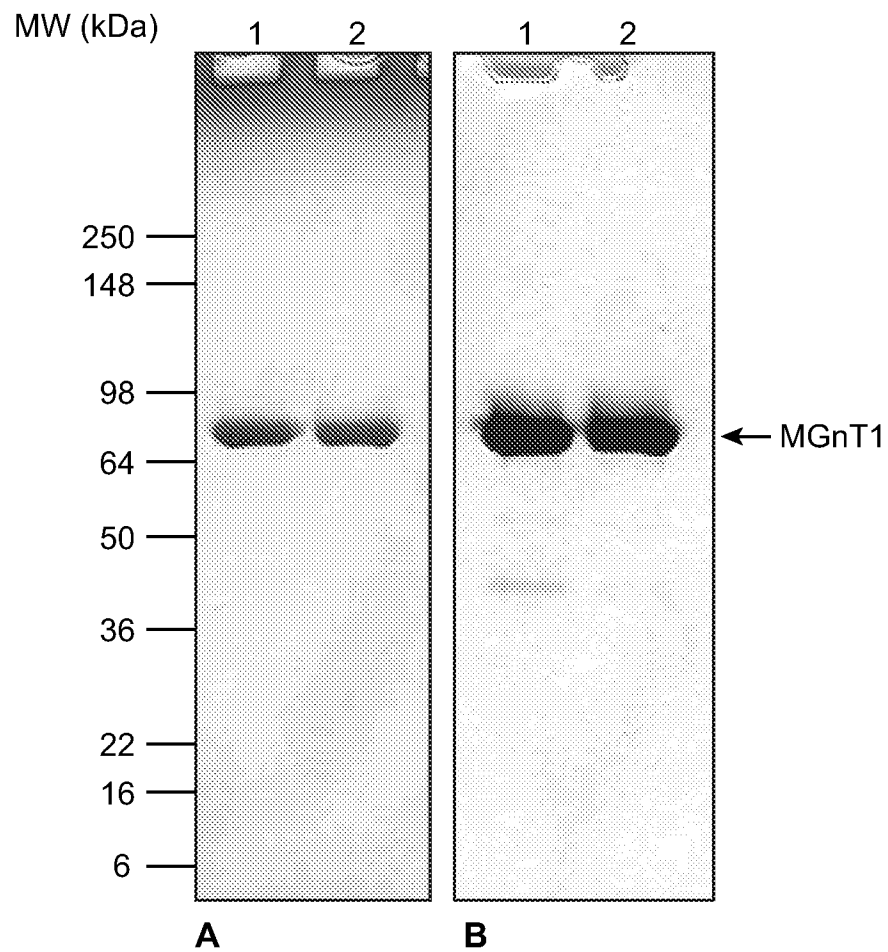


FIG. 11

12 / 12

Fraction	G25 load ^a	G25 pool
Activity (U/L)	542	391
MGnT-1(RP-HPLC) (mg/mL)	0.134	0.098
Vol (mL)	100	121
Total act. (U)	54.2	47.4
Act. step yield (%)	N/A	87%
Total mass (mg)	13.4	11.9
Mass Yield (%)	N/A	88%
Sp Act. (U/mg)	4.05	3.99
purity (%)	99.4	98.6
monomer (%)	100	100

FIG. 12