(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 July 2023 (13.07.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/133283 A1

(51) International Patent Classification:
C12N 15/62 (2006.01) *C12N 15/52* (2006.01)
C12N 9/10 (2006.01) *C12N 15/67* (2006.01)

(21) International Application Number:
PCT/US2023/010330

(22) International Filing Date:
06 January 2023 (06.01.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/297,419 07 January 2022 (07.01.2022) US

(71) Applicant: **CORNELL UNIVERSITY** [US/US]; Center for Technology Licensing ("CTL") at, Cornell University, 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(72) Inventors: **DELISA, Matthew**; 216 Wait Avenue, Ithaca, New York 14850 (US). **JAROENTOMEECHAI, Thapakorn**; c/o Center for Technology Licensing ("CTL") at, Cornell University, 395 Pine Tree Rd., Suite 310, Ithaca, New York 14850 (US). **MIZRACHI, Dario**; c/o Center for Technology Licensing ("CTL") at, Cornell University, 395 Pine Tree Rd., Suite 310, Ithaca, New York 14850 (US).

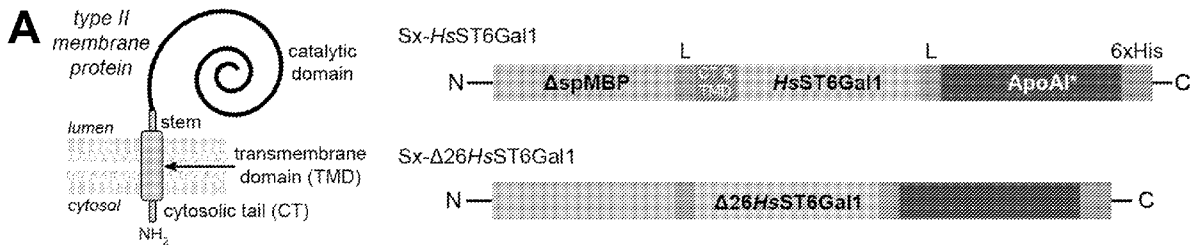
(74) Agent: **BLOCK, Olivia, K.T.** et al.; Troutman Pepper Hamilton Sanders LLP, 70 Linden Oaks, Suite 210, Rochester, New York 14625 (US).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: COMPOSITIONS AND METHODS FOR SOLUBILIZING GLYCOSYLTRANSFERASES



FIGS. 3A

(57) Abstract: The present disclosure relates to a nucleic acid construct having a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein. The chimeric nucleic acid molecule includes a first nucleic acid moiety encoding an amphipathic shield domain protein; a second nucleic acid moiety encoding a glycosyltransferase; and a third nucleic acid moiety encoding a water soluble expression decoy protein. The first nucleic acid moiety is coupled to the second nucleic acid moiety's 3' end and the third nucleic acid moiety is coupled to the second nucleic acid moiety's 5' end. The coupling may be direct or indirect. The present disclosure further relates to an expression vector, a host cell, and a tripartite glycosyltransferase fusion protein encoded by the nucleic acid construct. Also disclosed are methods of recombinantly producing a tripartite glycosyltransferase fusion protein in soluble form and methods of cell-free glycan remodeling.

WO 2023/133283 A1

COMPOSITIONS AND METHODS FOR SOLUBILIZING GLYCOSYLTRANSFERASES

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 63/297,419, filed January 7, 2022, which is hereby incorporated by reference in its entirety.

[0002] This invention was made with government support under HDTRA1-14-10052 awarded by Defense Threat Reduction Agency; CBET-1159581, CBET-1264701, CBET-1936823, and MCB 1413563 awarded by National Science Foundation; and 1R01GM137314, and 1R01GM127578 awarded by National Institutes of Health. The government has certain rights in the invention.

FIELD

[0003] The present disclosure relates to compositions and methods for making and using water-soluble glycosyltransferase fusion proteins.

BACKGROUND

[0004] Glycosylation – the process by which carbohydrate-based compounds known as glycans are covalently attached to acceptor molecules, typically proteins and lipids – is fundamental to all life (Varki, A., "Biological Roles of Glycans," *Glycobiology* 27:3-49 (2017) and Varki, A. et al., "Essentials of Glycobiology," (Cold Spring Harbor (NY) (2015)). Following conjugation to biomolecules, glycans add an additional layer of information and play important roles in numerous biological processes (Moremen et al., "Vertebrate Protein Glycosylation: Diversity, Synthesis and Function," *Nat. Rev. Mol. Cell Biol.* 13:448-62 (2012)) including cell adhesion and signaling (Crocker, P.R., "Siglecs: Sialic-acid-binding Immunoglobulin-like Lectins in Cell-cell Interactions and Signaling," *Curr. Opin. Struct. Biol.* 12:609-15 (2002) and Cummings, R.D., "Stuck on Sugars - How Carbohydrates Regulate Cell Adhesion, Recognition, and Signaling," *Glycoconj. J.* 36:241-257 (2019)), cell growth and development (Haltiwanger and Lowe, "Role of Glycosylation in Development," *Annu. Rev. Biochem.* 73:491-537 (2004)), and immune recognition/response (Zhou and Cobb, "Glycans in Immunologic Health and Disease," *Annu. Rev. Immunol.* 39:511-536 (2021) and Rudd et al., "Glycosylation and the Immune System," *Science* 291:2370-6 (2001)), among others. Moreover, structural remodeling of protein-linked glycans can improve therapeutic properties in a number of ways such as extending activity and stability both *in vitro* and *in vivo* (Sola and Griebenow, "Effects of Glycosylation on the Stability of Protein Pharmaceuticals," *J. Pharm. Sci.* 98:1223-

45 (2009) and Sinclair and Elliott, "Glycoengineering: The Effect of Glycosylation on the Properties of Therapeutic Proteins," *J. Pharm. Sci.* 94:1626-35 (2005), modulating interactions with specific immune receptors (Rothman et al., "Antibody-dependent Cytotoxicity Mediated by Natural Killer Cells is Enhanced by Castanospermine-induced Alterations of IgG Glycosylation," *Molecular Immunology* 26:1113-23 (1989)), and targeting to specific cells or tissues (Friedman et al., "A Comparison of the Pharmacological Properties of Carbohydrate Remodeled Recombinant and Placental-derived Beta-glucocerebrosidase: Implications for Clinical Efficacy in Treatment of Gaucher Disease," *Blood* 93:2807-16 (1999)).

[0005] As appreciation for the biological roles and therapeutic potential of glycans continues to grow, so too does the need for reliable, user-friendly technologies that enable their synthesis and remodeling. However, quantitative preparation of structurally-defined glycans and glycoconjugates remains technically challenging and represents a critical technology gap that limits widespread access to this important biomolecule class (Transforming Glycoscience: A Roadmap for the Future (The National Academies Press) (2012)). A major reason for this difficulty is the lack of template encoding in glycan biosynthesis, which distances carbohydrate structure and function from gene sequence. Hence, unlike nucleic acids and proteins, glycans cannot be directly produced from recombinant DNA technology. Instead, glycan biosynthesis is controlled by the availability, abundance, and activities of glycoenzymes, in particular, glycosyltransferases (GTs) that catalyze formation of specific glycosidic linkages by transferring sugar molecules from donor substrates (*e.g.*, nucleotide sugar or lipid-linked sugar) to hydroxyl groups of acceptor molecules (Lairson et al., "Glycosyltransferases: Structures, Functions, and Mechanisms," *Annu. Rev. Biochem.* 77:521-55 (2008) and Taniguchi et al., "Handbook of Glycosyltransferases and Related Genes," (Springer, Tokyo, Japan (2014)) and glycosyl hydrolases (GHs) that cleave glycan structures during oligosaccharide maturation (Davies and Henrissat, "Structures and Mechanisms of Glycosyl hydrolases," *Structure* 3:853-9 (1995)).

[0006] GTs exhibit unique catalytic specificities for a wide range of sugar donors and acceptor substrates and generate products with distinct anomeric configurations, which helps to explain the vast structural diversity of "glycospace". In mammals alone, it is estimated that there are approximately 7,000 oligosaccharide structures (Cummings, R.D., "The Repertoire of Glycan Determinants in the Human Glycome," *Mol. Biosyst.* 5:1087-104 (2009)) whose generation involves more than 200 GTs (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018) from 45 different protein families that have been annotated in the carbohydrate-active enzymes (CAZy) database (Lombard et al., "The Carbohydrate-active Enzymes Database (CAZy) in 2013," *Nucleic Acids Res* 42:D490-5 (2014)). Moreover, GTs are proficient at replicating the

diversity of naturally occurring glycans and glycoconjugates in unnatural contexts, leading to their emergence as powerful synthetic tools for building complex glycomolecules in the laboratory. Much of the progress in this regard exploits sugar nucleotide-dependent GTs of mammalian and bacterial origin for synthesis of complex carbohydrates, glycoconjugates, and glycosylated natural products, which are generated by functionally reconstituting artificial networks of these glycoenzymes within model cellular systems (Clausen et al., "Glycosylation Engineering in Essentials of Glycobiology (eds. rd et al.) 713-728 (Cold Spring Harbor (NY) (2015); Natarajan et al., "Metabolic Engineering of Glycoprotein Biosynthesis in Bacteria," *Emerg. Top Life Sci.* 2:419-432 (2018); Williams et al., "Metabolic Engineering of Capsular Polysaccharides," *Emerg. Top Life Sci.* 2:337-348 (2018); and Pandey et al., "Metabolic Engineering of Glycosylated Polyketide Biosynthesis," *Emerg. Top Life Sci.* 2:389-403 (2018)) or in cell-free, one-pot reaction systems (Na et al., "Recent Progress in Synthesis of Carbohydrates With Sugar Nucleotide-dependent Glycosyltransferases," *Curr. Opin. Chem. Biol.* 61:81-95 (2021); Jaroentomeechai et al., "Cell-free Synthetic Glycobiology: Designing and Engineering Glycomolecules Outside of Living Cells," *Front. Chem.* 8:645 (2020); and Kightlinger et al., "Synthetic Glycobiology: Parts, Systems, and Applications," *ACS Synth. Biol.* 9:1534-1562 (2020)).

[0007] These developments notwithstanding, broad access to GTs for fundamental and applied research is bottlenecked by difficulties associated with their recombinant expression. A major reason for this difficulty is that many GTs catalyze reactions at membrane interfaces (*e.g.*, between the cytoplasm and periplasm in Gram-negative bacteria or between the cytosol and endoplasmic reticulum (ER)/Golgi organelles within eukaryotes). As such, these enzymes are typically either secretory proteins or integral membrane proteins (IMPs) that need post-translational modifications (PTMs) (*e.g.*, disulfide bonds, *N*-linked glycosylation) and/or specialized chaperones to achieve proper folding, membrane translocation/insertion, and function. Efforts to express GTs in the absence of required PTMs or chaperones, or in the presence of single- or multi-pass transmembrane domains (TMDs) or terminal signal peptides (*e.g.*, N-terminal export signals, C-terminal retention signals), are often met with non-functional protein aggregates. This is particularly pronounced for expression of mammalian GTs in bacterial hosts, with successful reports often involving time- and labor-intensive searches for solubility-enhancing fusion partners and molecular chaperones, optimal host strains and culture conditions, and compatible detergents and denaturants for IMP solubilization and *in vitro* refolding from inclusion bodies, respectively (Skretas et al., "Expression of Active Human Sialyltransferase ST6GalNAcI in Escherichia coli," *Microb. Cell Fact* 8:50 (2009); Rao et al., "Structural Insight Into mammalian Sialyltransferases," *Nat. Struct. Mol. Biol.* 16:1186-8 (2009);

and Ramakrishnan and Qasba, "Crystal Structure of Lactose Synthase Reveals a Large Conformational Change in its Catalytic Component, the Beta1,4-galactosyltransferase-I," *J. Mol. Biol.* 310:205-18 (2001)).

[0008] For these reasons, functional expression of mammalian GTs in bacteria remains rare. Instead, eukaryotic cells remain the preferred host for producing recombinant glycoenzymes albeit with most studies involving small-scale expression of just one or a few GTs (Taniguchi et al., "Handbook of Glycosyltransferases and Related Genes," (Springer, Tokyo, Japan (2014)). To date, there are only a few reports of larger-scale expression campaigns involving significant numbers of GTs: one such study describes expression of 51 human GTs as fusions to the yeast cell wall Pir proteins to enable immobilization on the surface of *Saccharomyces cerevisiae* (Shimma et al., "Construction of a Library of Human Glycosyltransferases Immobilized in the Cell Wall of Saccharomyces cerevisiae," *Appl. Environ. Microbiol.* 72:7003-12 (2006)) while a second study describes expression of 339 human glycoenzymes as fusions to a solubility-enhancing GFP domain in either mammalian cells (HEK293) or baculovirus-infected insect cells (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018)). Interestingly, the authors of this latter study explore the potential of *Escherichia coli* for human glycoenzyme expression but report that all GTs expressed in this host accumulate as insoluble aggregates (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018)). Thus, the biosynthetic capacity and versatility of simple *E. coli* bacteria, one of the most important model organisms in biology and biotechnology (Blount, Z.D., "The Unexhausted Potential of E. coli," *Elife* 4(2015)), is yet to be unlocked for functional expression of GTs on a large scale.

[0009] The present disclosure is directed to overcoming these and other deficiencies in the art.

SUMMARY

[0010] A first aspect of the present disclosure relates to a nucleic acid construct. The nucleic acid construct includes a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein. The chimeric nucleic acid molecule includes a first nucleic acid moiety encoding an amphipathic shield domain protein; a second nucleic acid moiety encoding a glycosyltransferase; and a third nucleic acid moiety encoding a water soluble expression decoy protein. The first nucleic acid moiety is coupled to the second nucleic acid

moiety's 3′ end and the third nucleic acid moiety is coupled to the second nucleic acid moiety's 5′ end. The coupling may be direct or indirect.

[0011] Another aspect of the present disclosure relates to an expression vector including the nucleic acid construct according to the present disclosure.

[0012] Another aspect of the present disclosure relates to a host cell comprising the nucleic acid construct of the present disclosure.

[0013] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein produced by a host cell according to the present disclosure.

[0014] Another aspect of the present disclosure relates to a cell-free protein expression system. The cell-free protein expression system comprises a cell lysate or extract and a nucleic acid construct according to the present disclosure.

[0015] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein produced by the cell-free expression system according to the present disclosure.

[0016] Another aspect of the present disclosure relates to a method of recombinantly producing a tripartite glycosyltransferase fusion protein in water soluble form. This method involves providing a host cell according to the present disclosure or a cell-free expression system according to the present disclosure. The method further involves culturing the host cell or using the cell-free expression system under conditions effective to express the tripartite glycosyltransferase fusion protein in a water soluble form within the host cell cytoplasm or the cell-free expression system.

[0017] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein produced by the methods of recombinantly producing a tripartite glycosyltransferase fusion protein according to the present disclosure.

[0018] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein comprising: an amino terminal water soluble expression decoy protein; a glycosyltransferase; and a carboxyl terminal amphipathic shield domain protein.

[0019] Another aspect of the present disclosure relates to a method of cell-free glycan remodeling. This method involves providing a glycan primer; providing one or more tripartite glycosyltransferase fusion protein(s) according to the present disclosure; and incubating the glycan primer with the one or more tripartite glycosyltransferase fusion protein(s) under conditions effective to transfer a glycosyl group to the glycan primer to produce a modified glycan structure.

[0020] The Examples of the present disclosure describe a generalizable workflow for efficient production of structurally diverse GTs using standard *E. coli* expression strains. At the heart of this workflow is a protein engineering method called SIMPLEx (solubilization of

integral membrane proteins with high levels of expression) (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat Commun* 6:6826 (2015), which is hereby incorporated by reference in its entirety) that enables topological conversion of secretory and membrane-bound proteins into water-soluble variants. In the Examples of the present disclosure, this conversion is achieved for GTs by modifying their N-termini with a decoy protein that prevents membrane insertion and their C-termini with an amphipathic protein that effectively shields hydrophobic surfaces from the aqueous environment (FIG. 3A). Using this approach, soluble expression of nearly 100 GTs, including many of human origin, directly within the *E. coli* cytoplasm at titers in the 5-10 mg/L range is demonstrated. Importantly, this large-scale expression platform furnishes functional glycoenzymes that can subsequently be used to remodel the structures of diverse glycan donors, leading to the formation of a variety of important glycoforms including human complex-type *N*-glycans on the therapeutic monoclonal antibody (mAb) trastuzumab. It is anticipated that SIMPLEx-remodeled GTs will help to deepen the understanding of glycoenzymes from all kingdoms of life and accelerate the assembly of these enzymes into cell-based and cell-free systems that enable biosynthesis of important glycomolecules.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] FIGS. 1A–1B are tables providing the strains, cell lines, and plasmids used in Examples 1 –9. Table 1A provides strains, cell lines, and plasmids from various sources. Table 1B provides plasmids designed and evaluated in Examples 1–9. Reference 1: Dyson et al., "Production of Soluble Mammalian Proteins in *Escherichia coli*: Identification of Protein Features that Correlate with Successful Expression," *BMC Biotechnol.* 4:32 (2004); Reference 2: Glasscock et al., "A Flow Cytometric Approach to Engineering *Escherichia coli* for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488–495 (2018); Reference 3: Hamilton et al., "A Library of Chemically Defined Human N-Glycans Synthesized from Microbial Oligosaccharide Precursors," *Sci. Rep.* 7: 15907 (2017); Reference 4: Mizrachi et al., "Making Water-Soluble Integral Membrane Proteins in vivo Using an Amphipathic Protein Fusion Strategy," *Nat. Commun.* 6:6826 (2015); Reference 5: Stark et al., "BioBits Bright: A Fluorescent Synthetic Biology Education Kit," *Sci. Adv.* 4:eaat5107 (2018); Reference 6: Dodev et al., "A Tool Kit for Rapid Cloning and Expression of Recombinant Antibodies," *Sci. Rep.* 4:5885 (2014); and Reference 7: Feldman et al., "Engineering N-Linked Protein Glycosylation with Diverse O Antigen Lipopolysaccharide Structures in *Escherichia coli*," *Proc. Natl. Acad. Sci. USA* 102:3016–21 (2005), each of which are hereby incorporated by reference in their entirety.

[0022] FIGS. 2A–2D are tables showing the results of experiments carried out with glycosyltransferase enzymes according to the present disclosure. FIG. 2A is a table providing the name, glycosyltransferase family, Uniprot ID (*which is hereby incorporated by reference in its entirety), protein structure/topology, and *E. coli* host strain of the glycosyltransferase enzymes used in Examples 1–9. FIG. 2B is a table providing the 3D structure availability (UNIPROT), SIMPLEX expression score, unfused expression score, pI – full length (ExPASy), MW-full length (ExPASy), solubility prediction-full length, pI- truncated (ExPASy), MW-truncated (ExPASy), solubility prediction-truncated, MW-SIMPLEx, pI-SIMPLEx, pI-SIMPLEx (ExPASy), and solubility prediction-SIMPLEx for the glycosyltransferase enzymes used in Examples 1–9. FIG. 2C is a table providing the full length sequence (FASTA) of the glycosyltransferase enzymes used in Examples 1–9 (SEQ ID NOs. 1–100). FIG. 2D is a table providing the truncated sequence (FASTA) of the glycosyltransferase enzymes used in Examples 1–9 (SEQ ID NOs: 101–174).

[0023] FIGS. 3A–3D demonstrate SIMPLEx-mediated expression of biologically-active *Hs*ST6Gal1. FIG. 3A is a schematic showing membrane topology of type II transmembrane proteins and molecular architecture of SIMPLEx constructs. Each construct consisted of N-terminal ΔspMBP and C-terminal ApoAI* that flanked *Hs*ST6Gal1. Intervening flexible linker (L) connects ΔspMBP and ApoAI* to the GT domains while the 6xHis tag was placed at the C-terminus to facilitate detection and purification. *Hs*ST6Gal1 domain variants studied were wild-type (wt) *Hs*ST6Gal1 (top) and truncated Δ26*Hs*ST6Gal1 (bottom), in which the cytoplasmic tail (CT) and transmembrane domain (TMD) were removed. FIG. 3B shows immunoblot analysis of the soluble (S), detergent-solubilized (D), and insoluble (I) fractions prepared from *E. coli* SHuffle T7 Express *lysY* cells carrying plasmid pET28a(+) encoding each of the indicated constructs. An equivalent amount of total protein was loaded in each lane. Blots were probed with anti-polyhistidine antibody (αHis). Control blots were generated by probing with anti-GroEL antibody. Results are representative of three biological replicates. Molecular weight (M_w) markers are shown at left. FIG. 3C shows kinetic analysis of purified Sx-Δ26*Hs*ST6Gal1 and commercial human ST6Gal1 performed using asialofetuin as acceptor substrate and CMP-Neu5Ac as donor substrate. A standard phosphate curve was generated (see FIG. 9B) to convert the initial raw absorbance reading to the enzymatically released inorganic phosphate from CMP-Neu5Ac. Values for V_{max} and K_m values were determined using Prism 9. Data are the mean of three biological replicates +/- SEM. FIG. 2D shows functional characterization of sialyltransferase-mediated chemoenzymatic remodeling of protein-linked glycans using bioorthogonal click chemistry-based assay. Fluorescence (501/523 nm ex/em) measured in clarified lysates prepared from *E. coli* cells expressing: Sx-Δ26*Hs*ST6Gal1 (Sx-ST6), ΔspMBP-

Δ26*Hs*ST6Gal1 (ΔspMBP-ST6), Δ26*Hs*ST6Gal1-ApoAI (ST6-ApoAI), or Δ26*Hs*ST6Gal1 (ST6), as indicated. Lysates from *E. coli* cells carrying empty pET28a(+) plasmid was used as a negative control (empty lysate). Fluorescence data, corresponding to the extent of chemoenzymatic modification, are the mean of three biological replicates (starting from freshly transformed cells) +/- SEM.

[0024] FIGS. 4A–4J demonstrate soluble expression of Sx-GT constructs in the *E. coli* cytoplasm. 98 GTs were evaluated for soluble, cytoplasmic expression in the SIMPLEx framework. Immunoblot analysis of soluble fractions derived from either BL21(DE3) or SHuffle T7 Express cells carrying plasmids for Sx-GT (top blot in each panel) or unfused GT (bottom blot in each panel) constructs. GTs were clustered according to origin and activity as follows: human glucosyltransferases (*Hs*GlcTs) (FIG. 4A); human galactosyltransferases (*Hs*GalTs) (FIG. 4B); human mannosyltransferases (*Hs*ManTs) (FIG. 4C); human *N*-acetylglucosaminyltransferases (*Hs*GlcNAcTs) (FIG. 4D); human *N*-acetylgalactosaminyltransferases (*Hs*GalNAcTs) (FIG. 4E); human fucosyltransferases (*Hs*FucTs) (FIG. 4F); human sialyltransferases (*Hs*SiaTs) (FIG. 4G); other human GTs (*Hs*GTs) (FIG. 4H); eukaryotic GTs (*Euk*GTs) (FIG. 4I); and bacterial GTs (*Bac*GTs) (FIG. 4J). The expression strain and sequence for each GT including information about truncation of TMD domains are provided in FIG. 2. Graphical representations of monosaccharide substrates are presented according to symbol nomenclature for glycans (Symbol Nomenclature for Graphical Representation of Glycans, *Glycobiology* 25: 1323-1324 (2015), which is hereby incorporated by reference in its entirety). An equivalent amount of total protein was loaded in each lane and blots were probed with anti-polyhistidine antibody (αHis) to detect GTs. To confirm equivalent loading, the same samples were probed with anti-GroEL antibody (see FIGS. 10A–10J). Blots are representative of three biological replicates. Molecular weight (M_w) markers are indicated at left.

[0025] FIGS. 5A–5C demonstrate compatibility of SIMPLEx reformatting with diverse expression platforms. Immunoblot analysis of the soluble (S) and insoluble (I) fractions derived from: cell-free protein synthesis (CFPS) using crude S30 extract prepared from *E. coli* BL21(DE3) cells (FIG. 5A); and cell-based expression using *S. cerevisiae* strain SBY49 (FIG. 5B) or HEK 293T cells (FIG. 5C) as indicated. All three systems involved plasmids for expressing either Sx-Δ26*Hs*ST6Gal1 (Sx-ST6) or unfused Δ26*Hs*ST6Gal1 (ST6). Empty plasmid was used as a negative control (empty) in each case. Blots were probed with anti-polyhistidine (αHis) antibody to detect GT expression, with longer exposures (αHis - long) provided to better identify protein products with low expression. An equivalent amount of total protein was loaded in each lane and confirmed by probing blots with antibodies specific for

GroEL, Tubulin, and GAPDH, which are housekeeping proteins in *E. coli*, yeast, and mammalian cells, respectively. Results are representative of three biological replicates. Molecular weight (M_w) markers are shown on the left.

[0026] FIGS. 6A–6B demonstrate cell-free construction of hybrid- and complex-type *N*-glycans using Sx-GTs. FIG. 6A is a schematic of bioenzymatic routes to hybrid- and complex-type *N*-glycan structures. $Man_3GlcNAc_2$ glycan (M3; glycan 1) derived from glycoengineered *E. coli* cells equipped with biosynthesis pathway for eukaryotic trimannosyl core *N*-glycan was used as primer for glycan construction. Subsequent cell-free glycan elaboration reactions yielded the following *N*-glycan structures: 2 G0-GlcNAc; 3 G0; 4 G2; 5 G2S; 6 G2S2; 7 G0F; 8 G2F; 9 G2S1F; and 10 G2S2F. Glycan naming follows shorthand notation for IgG glycans. For complete glycan list with chemical structures, see FIG. 23. Synthesis steps: (i) non-enzymatic acid hydrolysis; (ii) Sx-Δ29*Hs*GnTI; (iii) Sx-Δ29*Hs*GnTII; (iv) Sx-Δ30*Hs*FucT8; (v) Sx-Δ44*Hs*β4GalT1; and (vi, vii) Sx-Δ26*Hs*ST6Gal1. All Sx-GTs were produced using *E. coli* BL21(DE3) or its derivative SHuffle T7 Express *lysY*. FIG. 6B shows MALDI-TOF MS spectra of glycans 1-10, where glycan 1 served as primer that was used as starting material to generate enzymatically-derived product glycans 2-10.

[0027] FIGS. 7A–7B show remodeling of IgG-Fc *N*-glycans on trastuzumab using Sx-GTs. FIG. 7A is a schematic of bioenzymatic routes to hybrid- and complex-type *N*-glycan structures linked to asparagine 297 (N297) of the trastuzumab antibody. Trastuzumab bearing $Man_5GlcNAc_2$ glycan (M5; glycan 11) derived from glycoengineered HEK293F lacking GnTI activity was used as a glycan primer. Subsequent cell-free glycan remodeling reactions yielded the following *N*-glycan structures: 12 (M5+GlcNAc); 2 (G0-GlcNAc); 3 (G0); 4 (G2); and 6 (G2S2). Glycan notation follows IgG glycan short naming system. For complete glycan list with chemical structures, see FIG. 23. SIMPLEx-reformatted GTs and glycosidase for each synthesis step are provided above reaction arrow. FIG. 7B shows deconvoluted LC-MS spectra in 140–160 kDa range using intact antibody analysis of trastuzumab bearing glycan 11 as starting material and enzymatically-derived product glycans 2-4, 6, and 12. Structures of anticipated *N*-glycan products are provided in each spectrum. Full MS spectra (0-200 kDa) for all structures are provided in FIG. 20.

[0028] FIGS. 8A–8B demonstrates that SIMPLEx architecture promotes soluble expression of difficult-to-express proteins (DTEPs). FIG. 8A shows the results of immunoblot analysis of the soluble fraction prepared from *E. coli* cells expressing human-derived DTEPs as SIMPLEx fusions (left panel) or as unfused proteins (right panel). *E. coli* BL21(DE3) was used to express human transcription factors (GATA2, JUN, and FOS), human cyclin-dependent kinase 4 (CDK4), and human cyclin-dependent kinase inhibitor 2A (CDKN2A), while SHuffle

T7 Express *lysY* strain was used to express human epidermal growth factor receptor tyrosine kinase domain (EGFRTK), human matrix metallopeptidase 1 (MMP1), and human proinsulin (ProIns). Human gene constructs were based on work of Dyson et al1. Dashed line indicates cropped membrane to remove empty lane. FIG. 8B shows a Coomassie-stained SDS-PAGE (left) and immunoblot analysis (right) of whole cell lysates derived from *E. coli* SHuffle T7 Express *lysY* cells carrying plasmid pET28a(+) encoding each of the indicated constructs. For immunoblots in FIG. 8A, an equivalent amount of total protein was loaded in each lane. For SDS-PAGE gel and immunoblot in FIG. 8B, samples were normalized by culture OD_{600} such that an equivalent number of cells were loaded in each lane. Immunoblots in FIGS. 8A–8B were probed with anti-polyhistidine antibody (αHis). Control blots were generated by probing membranes with anti-GroEL antibody. Results are representative of three biological replicates. Molecular weight (M_w) markers are shown on left of each blot/gel. Red arrows denote expression products.

[0029] FIGS. 9A–9B demonstrate the functional characterization of Sx-Δ26*Hs*ST6Gal1. FIG. 9A shows the results of Coomassie-stained SDS-PAGE gel analysis of fractions corresponding to purification of Sx-Δ26*Hs*ST6Gal1 by Ni-NTA chromatography. Gel is representative of three biological replicates. FIG. 9B shows the specific activity of sialyltransferase determined using malachite green phosphate reagents. Absorbance readings (OD_{600}) correspond to amount of released inorganic phosphate from CMP-Neu5Ac after glycosylation. Data are the mean of three biological replicates +/- SEM. FIG. 9C is a schematic of a bioorthogonal click chemistry-based ST assay. Purified human alpha-1 antitrypsin (A1AT) was treated with α2-3,6,8,9 neuraminidase (NA) to remove native sialic acid and used as substrate to evaluate Sx-Δ26*Hs*ST6Gal1-mediated installation of azido-Neu5Ac. Depicted glycans are representative glycoforms of native human A1AT. Azido (N3-) functional groups on Neu5Ac provide a chemical handle on A1AT for conjugation with carboxyrhodamine 110 (CR110) fluorophore or PEG4-biotin reporters via strain-promoted azide-alkyne cycloaddition (SPAAC) using dibenzocyclooctyne group (DBCO) as reactive alkyne. FIG. 9D shows representative SDS-PAGE and immunoblot of reaction products of *in vitro* ST assay. After labeling with CR110, reaction mixtures were separated on SDS-PAGE gel and fluorescence signal of labeled glycoproteins was measured at 501/523 nm $\lambda_{ex}/\lambda_{em}$. Coomassie-stained gel served as loading control. Results are representative of three biological replicates. Molecular weight (M_w) markers are shown at left. FIG. 9E shows fluorescence corresponding to *in vitro* ST activity of purified Sx-Δ26*Hs*ST6Gal1 and commercial *Hs*ST6Gal1 as function of enzyme concentration. Data are the mean of three biological replicates +/- SEM. Inset is logarithmic representation of same data.

[0030] FIGS. 10A–10J show loading control blots for FIGS. 4A–4J. Control immunoblots corresponding to each panel in FIGS. 5A–5C that were generated by loading an identical amount of each sample and probing with anti-GroEL antibody. Immunoblot analysis of soluble fractions derived from either BL21(DE3) or SHuffle T7 Express cells carrying plasmids for Sx-GT (top blot in each panel) or unfused GT (bottom blot in each panel) constructs. GTs were clustered according to origin and activity as follows: human glucosyltransferases (*Hs*GlcTs) (FIG. 10A); human galactosyltransferases (*Hs*GalTs) (FIG. 10B); human mannosyltransferases (*Hs*ManTs) (FIG. 10C); human *N*-acetylglucosaminyltransferases (*Hs*GlcNAcTs) (FIG. 10D); human *N*-acetylgalactosaminyltransferases (*Hs*GalNAcTs) (FIG. 10E); human fucosyltransferases (*Hs*FucTs) (FIG. 10F); human sialyltransferases (*Hs*SiaTs) (FIG. 10G); other human GTs (*Hs*GTs) (FIG. 10H); eukaryotic GTs (*Euk*GTs) (FIG. 10I); and bacterial GTs (*Bac*GTs) (FIG. 10J). Results are representative of three biological replicates. Molecular weight (*Mw*) markers are indicated at left.

[0031] FIGS. 11A–11I show subcellular fractionation analysis of SIMPLEx-reformatted GT expression. Western blot analysis of the soluble (S), detergent-solubilized (D), and insoluble (I) fractions prepared from either BL21(DE3) or SHuffle T7 Express cells carrying plasmid pET28a(+) encoding either Sx-GT (left blot in each panel) or GT (right blot in each panel) constructs corresponding to the following GTs: Δ29*Hs*GnTI (FIG. 11A); Δ36HsFucT7 (FIG. 11B); Δ35*Hs*ST6 GalNAc1 (FIG. 11C); Δ40*Hs*Alg11 (FIG. 11D); Δ30*Nt*GnTI (FIG. 11E); *Cj*CsTII (FIG. 11F); *Hp*β4GalT1 (FIG. 11G); *Ng*β4GalT1 (FIG. 11H); and *Nm*β4GalT1 (FIG. 11I). An equivalent amount of total protein was loaded in each lane. Blots were probed with anti-polyhistidine antibody (αHis). Results are representative of three biological replicates. Molecular weight (M_w) markers are shown on the left.

[0032] FIG. 12 shows the cell density of *E. coli* cultures expressing GT enzymes. Representative data points for cultures expressing SIMPLEx-reformatted GT fusions (Sx-GT; blue) or unfused GTs (red). All data points for each GT were plotted on the same axis. Final cell density values were recorded as OD_{600} readings taken after 16-18 hours of growth. Graph depicts three biological replicates for each construct.

[0033] FIGS. 13A–13C show yield quantification for select SIMPLEx-reformatted GTs. FIG. 13A shows the results of Coomassie-stained SDS-PAGE gel analysis of three representative Sx-GTs (Sx-Δ26*hs*ST6Gal1, Sx-Δ29*hs*GnT1, Sx-Δ30*hs*FUT8) following expression and purification by Ni-NTA chromatography. An equivalent amount of total protein was loaded in each lane. Results are representative of three biological replicates. Molecular weight (M_w) marker is shown on the left. Red arrows denote full-length expression products. Soluble protein yields determined on a mass basis (mg/L) (FIG. 13B) and a molar basis (nM) (FIG. 13C). Each

indicated protein was purified as in FIG. 13A using Ni-NTA chromatography from 1-L cultures of *E. coli* carrying each SIMPLEx-reformatted GT construct as indicated. *E. coli* cultures expressing unfused GT constructs served as controls. Yield values are representative of three biological replicates and error bars represent SEM. Red asterisks denote purified proteins depicted in FIG. 13A.

[0034] FIG. 14A–14B demonstrate the physicochemical properties of GTs that correlate with successful expression. FIG. 14A shows immunoblot analysis of SIMPLEx-reformatted GTs (Sx-GTs) and unfused GTs (GTs) and demonstrates expression score assignment. Blots were probed with anti-polyhistidine antibody (αHis) and results are representative of three biological replicates. Based on relative band intensities from immunoblot analysis, each GT as a SIMPLEx construct or unfused enzyme was categorized as non-expressor (score 0; grey circle), weak expressor (score 1; cyan circle), medium expressor (score 2; light blue circle), and strong expressor (score 3; dark blue circle). Bar graph summarizes total number of GTs for each category, comparing SIMPLEx and unfused formats as indicated. FIG. 14B shows scatter plots relating the following physicochemical properties: (i) protein molecular weight (Mw) excluding added mass from ΔspMBP and ApoAI* domains; (ii) protein isoelectric point (pI); and (iii) protein solubility score as calculated by Protein-Sol server. Individual data points colored according to their expressor score category. Plots were generated using R version 3.4.2 software. All data used for analysis are provided in FIG. 2.

[0035] FIGS. 15A–15B demonstrate the relationship between soluble expression, protein size and isoelectric point. FIG. 15A shows the expression scores for unfused GTs (GTs) or SIMPLEx-reformatted GTs (Sx-GTs) as a function of protein molecular weight to provide average expression score ($\overline{Ex}$) for small (<40 kDa), medium (40-60 kDa), and large (>60 kDa) proteins. Note that the added molecular weight from ΔspMBP and ApoAI* domains of the SIMPLEx construct was excluded from size classification. Graphs depict mean of expression scores determined in FIG. 14A +/- SEM. Statistical significance was determined by Welch's two-sided t-test ($p < 0.05$ considered significant; ns, not significant), leading to the following p-values: 0.0423 for Sx-GTs <40 kDa vs. >60 kDa; 0.0467 for Sx-GTs 40-60 kDa vs. >60 kDa; and 0.0464 for GTs <40 kDa vs. 40-60 kDa. FIG. 15B is a scatter plot of protein isoelectric point (pI) of unfused and SIMPLEx-fused GTs. Data points were labeled according to change in expression score difference between SIMPLEx-fused and unfused GTs as indicated in legend. Diagonal line represents no change in pI due to SIMPLEx fusion. Arrows signify shift towards a more basic or more acidic protein following SIMPLEx fusion. Plots were generated using R version 3.4.2 software. All data used for analysis are provided in FIG. 2.

[0036] FIGS. 16A–16C shows MS analysis of sialylated *N*-glycans from cell-free remodeling. FIG. 16A is an HILIC-LC-MS chromatogram of cell-free reaction to install sialic acid on glycan 4 using Sx-Δ26*Hs*ST6Gal1. Schematic for stepwise conversion of glycan 4 to 5 and 6 is shown, with the preferred substrate for *Hs*ST6Gal1, the α1–3Man branch, highlighted in green and the less-preferred α1–6Man branch highlighted in red. FIG. 16B shows an MS (left panel) and an MS/MS (right panel) spectra of the doubly charged glycan at m/z 966.8463. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S1 product. FIG. 16C shows an MS (left panel) and an MS/MS (right panel) spectra of the doubly charged glycan at m/z 1112.3943. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S2 product.

[0037] FIGS. 17A–17C show MS analysis of sialylated, core-fucosylated *N*-glycans from cell-free remodeling. FIG. 17A is an HILIC-LC-MS chromatogram of cell-free reaction to install sialic acid on glycan 8 using Sx-Δ26*Hs*ST6Gal1. Schematic for stepwise conversion of glycan 8 to 9 and 10 is shown, with the preferred substrate for HsST6Gal1, the α1–3Man branch, highlighted in green and the less-preferred α1–6Man branch highlighted in red. FIG. 17B shows an MS (left panel) and an MS/MS (right panel) spectra of the doubly charged glycan at m/z 1039.8713. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S1 product. FIG. 17C shows an MS (left panel) and an MS/MS (right panel) spectra of the doubly charged glycan at m/z 1185.4182. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S2 product.

[0038] FIGS. 18A–18D demonstrates remodeling glycans on therapeutic glycoproteins with Sx-GTs. Cell-free reactions catalyzing: sialyltransferase reaction on *N*-glycan of neuraminidase (NA)-treated A1A1 using Sx-*Cj*CstII (FIG. 18A); fucosyltransferase reaction on *N*-glycan of NA-treated A1AT using Sx-Δ36*Hs*FucT7 (FIG. 18B); sialyltransferase reaction on O-glycan of NA-treated bovine submaxillary mucin using Sx-Δ34*Hs*ST3Gal1 (FIG. 18C); and *N*-acetylglucosaminyl transferase reaction on $Man_3GlcNAc_2$ *N*-glycan of MBP-glucagon fusion protein2 using Sx-Δ29*Hs*GnT1 (FIG. 18D). In all cases, reactions were performed using nucleotide-activated sugars modified with azide group as indicated. Following Sx-GT-catalyzed reactions, glycoprotein products were labeled with either dibenzocyclooctyne-carboxyrhodamine 110 (DBCO-CR110) or DBCO-PEG_4-biotin reporters via SPACC reaction. In-gel fluorescence analysis was used to detect glycoproteins modified with fluorescent CR110 reporter while immunoblot analysis using streptavidin-HRP was used to detect glycoproteins modified with biotin reporter. Results are representative of three biological replicates. Molecular weight (M_w) markers are shown at left.

[0039] FIGS. 19A–19D demonstrate glycosidase sensitivity of *N*-glycans on

trastuzumab. Deconvoluted MS analysis of intact trastuzumab derived from Expi293F™ GnTI- cells following incubation with: PBS (FIG. 19A); endoglycosidase S2 (FIG. 19B); endoglycosidase F1 (FIG. 19C); and endoglycosidase F3 (FIG. 19D). All reactions were carried out at 37°C for 16 hours in 10-µL reaction volume. Full spectra in the range of 25-200 kDa are shown in the panels on the left. Red box indicates region between 140-160 kDa and the spectra for this mass range is provided in the panels on the right. Structures of anticipated *N*-glycoprotein products are provided within each spectrum.

[0040] FIG. 20 shows an MS spectra for trastuzumab glycoforms. Full MS spectra in the range of 0-200 kDa corresponding to each glycoform of trastuzumab detected in FIGS. 7A–7B. Structures of anticipated *N*-linked glycoprotein products are provided in each spectrum.

[0041] FIGS. 21A–21B demonstrate remodeling of IgG-Fc *N*-glycans on trastuzumab using Sx-GTs. FIG. 21A is an extended schematic of bioenzymatic routes to hybrid- and complex-type *N*-glycan structures linked to N297 of trastuzumab. Trastuzumab bearing $Man_5GlcNAc_2$ glycan (M5; glycan 11) derived from glycoengineered HEK293F lacking GnTI activity was used as a glycan primer. Subsequent cell-free glycan remodeling reactions yielded the following *N*-glycan structures: 1 (M3); 2 (G0-GlcNAc); 3 (G0); 4 (G2); 6 (G2S2); 7(G0F); 12 (M5+GlcNAc); 13 (M5+GlcNAcGal); 14 (M3+GlcNAc); and 15 (G0F-GlcNAc). For complete glycan list with chemical structures, see FIG. 23. SIMPLEx-reformatted GTs and glycosidases for each synthesis step are provided above reaction arrow. FIG. 21B shows deconvoluted LC-MS spectra in 140–160 kDa range using intact antibody analysis of enzymatically-derived product glycans 1, 7, and 13-15. Structures of the anticipated *N*-glycoprotein products are provided in each spectrum. Asterisk indicates unidentified product.

[0042] FIGS. 22A–22B demonstrates remodeling of IgG-Fc *N*-glycans on trastuzumab using Sx-GTs. FIG. 22A is a schematic of bioenzymatic routes to unnatural *N*-glycan structures linked to N297 of trastuzumab. Trastuzumab bearing $Man_5GlcNAc_2$ glycan (M5; glycan 11) was used as a glycan primer to generate *N*-glycan 16 (M5-GlcNAz). Subsequent strain-promoted cycloaddition reaction was performed to install either DBCO-PEG4-Biotin or DBCO-carboxyrhodamine (CR110) to the azide group, yielding glycan 17 and glycan 18, respectively. FIG. 22B shows deconvoluted LC-MS spectra in 140–160 kDa range using intact antibody analysis of enzymatically-derived product glycans 16–18. Structures of the anticipated *N*-glycoprotein products are provided in each spectrum.

[0043] FIG. 23 is a table providing *N*-glycan structures produced in Examples 1–9.

DETAILED DESCRIPTION

[0044] A first aspect of the present disclosure relates to a nucleic acid construct. The nucleic acid construct includes a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein. The chimeric nucleic acid molecule includes a first nucleic acid moiety encoding an amphipathic shield domain protein; a second nucleic acid moiety encoding a glycosyltransferase; and a third nucleic acid moiety encoding a water soluble expression decoy protein. The first nucleic acid moiety is coupled to the second nucleic acid moiety's 3′ end and the third nucleic acid moiety is coupled to the second nucleic acid moiety's 5′ end. The coupling may be direct or indirect.

[0045] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein produced by the methods of recombinantly producing a tripartite glycosyltransferase fusion protein according to the present disclosure.

[0046] The nucleic acid molecules encoding the various polypeptide components of a tripartite glycosyltransferase fusion protein can be ligated together along with appropriate regulatory elements that provide for expression of the tripartite glycosyltransferase fusion protein. Typically, the nucleic acid construct encoding the chimeric protein can be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art and further described *infra*.

[0047] As used herein, "nucleic acid", refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The nucleic acid construct may be a synthetic nucleic acid construct. As used herein "synthetic" nucleic acid construct refers to a nucleic acid construct that is artificially produced and/or that does not exist in nature. As described in more detail herein, the nucleic acid constructs of the present disclosure are utilized to make water-soluble glycosyltransferases using an amphipathic protein fusion strategy. In particular, the nucleic acid constructs are part of a new strategy for the solubilization of glycosyltransferases based on the affinity for hydrophobic surfaces displayed by amphipathic proteins.

[0048] As used herein, the term "glycosyltransferase" (GT) includes an enzyme or fragment thereof which catalyzes the transfer of a donor glycosyl moiety from a glycosyl donor to an acceptor. Suitable glycosyl donors include, without limitation, CMP-sialic acid, GDP-fucose, GDP-mannose, UDP-glucose, UDP-galactose, UDP-xylose, UDP-*N*-acetylglucosamine,

UDP-*N*-acetylgalactosamine, UDP-glucuronic acid, Dolichol-P-glucose, Dolichol-P-mannose, Dolichol-P-P-(glucose$_3$-mannose$_9$-GlcNAc$_2$), and undecaprenyl-P-P-*N*-acetylmuramic acid-pentapeptide-GlcNAc). Suitable acceptor moieties include, without limitation, oligosaccharides, monosaccharides, polypeptides, proteins, lipids such as ceramides, small organic molecules, and nucleic acid molecules such as DNA.

[0049] GTs may be classified as (i) single-pass transmembrane proteins with C-termini in the cytoplasm (type I transmembrane protein); (ii) single-pass transmembrane proteins with N-termini in the cytoplasm (type II transmembrane protein); (iii) multi-pass transmembrane proteins; (iv) secretory proteins with N-terminal signal peptides and C-terminal ER retention domains; and (v) cytosolic proteins. In some embodiments, the glycosyltransferase is selected from the group consisting of (i) a single-pass transmembrane protein with C-terminus in cytoplasm (type I transmembrane protein); (ii) a single-pass transmembrane protein with N-terminus in cytoplasm (type II transmembrane protein); (iii) a multi-pass transmembrane protein; and (iv) a secretory protein with N-terminal signal peptide and C-terminal ER retention domain.

[0050] In some embodiments, the glycosyltransferase is a full-length glycosyltransferase. Accordingly the second nucleic acid moiety encodes a full-length glycosyltransferase. In accordance with such embodiments, the second nucleic acid moiety comprises a full-length GT gene.

[0051] For example, the full-length GT may contain an internal single-pass or multi-pass TMD (*e.g.*, human Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase (*Hs*Alg3), human Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase (*Hs*Alg12), human GPI mannosyltransferase 1 (*Hs*PIGM), human GPI mannosyltransferase 3 (*Hs*PIGB), human GPI mannosyltransferase 4 (*Hs*PIGZ), human dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase (*Hs*Alg6), human probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase (*Hs*Alg8), human Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase (*Hs*Alg10), human ceramide glucosyltransferase (*Hs*UGCG), *E. coli* undecaprenyl-phosphate alpha-*N*-acetylglucosaminyl 1-phosphate transferase (*Ec*WecA), yeast beta-1,4-mannosyltransferase OS=*Saccharomyces cerevisiae* (*Sc*Alg1), and yeast GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Sc*Alg11) (FIG. 2A)).

[0052] In other embodiments, the full-length GT may be a predicted cytosolic GT (*e.g.*, human isoform 2 of putative UDP-*N*-acetylglucosamine transferase (*Hs*Alg13), human dolichol-phosphate mannosyltransferase subunit 1 (*Hs*DPM1), human glycogenin-1 (*Hs*GLYG), *Campylobacter jejuni* CsTII (*Cj*CsTII), *Neisseria meningitidis* polysialic acid O-acetyltransferase (*Nm*PolysiaT), *Campylobacter jejuni* beta-1,3-galactosyltransferase (*Cj*CgtB),

Helicobacter pylori (strain 51) beta-4-galactosyltransferase (*Hp*LgtB), *Neisseria meningitidis* serogroup B (strain MC58) lacto-*N*-neotetraose biosynthesis glycosyltransferase LgtB (*Nm*LgtB), *Neisseria gonorrhoeae* lacto-*N*-neotetraose biosynthesis glycosyltransferase (*Ng*LgtB), *E.coli* galactoside 2-alpha-L-fucosyltransferase WbgL (*Ec*FUT), *Legionella pneumophila* subsp. Pneumophila subversion of eukaryotic traffic protein A (*Lp*SetA), and *Neisseria meningitidis* Alpha-2,9-polysialyltransferase (*Nm*SynE) (FIG. 2A)).

[0053] As described *infra* N-/C-terminal transmembrane domains (TMDs) as well as C-terminal ER retention domains in mammalian GTs are used as membrane anchors and are dispensable for catalytic activity (Harduin-Lepers et al., "The Human Sialyltransferase Family," *Biochimie* 83727-83737 (2001), which is hereby incorporated by reference in its entirety). Thus, in some embodiments, the glycosyltransferase is a truncated glycosyltransferase. The truncated glycosyltransferase may exclude a GT C-terminal ER retention domain, a terminal TMD anchor, or both a C-terminal ER retention domain and a terminal TMD anchor. In some embodiments, the truncated glycosyltransferase excludes an N-terminal signal peptides. Various exemplary truncated GTs are provided in FIG. 2D.

[0054] Glycosyltransferases play vital roles in glycosylation and glycan remodeling. The tripartite glycosyltransferase fusion proteins according to the present disclosure are water soluble following extraction from their native environment (*e.g.*, a cellular membrane) without the use of detergents and/or detergent-like amphiphiles, overproduction using recombinant systems, protein engineering, and/or mutations to the GT itself, thereby allowing for improved functional and structural studies of GTs as well as *in vitro* reconstitution of enzymatic activity or *in vitro* reconstitution of a biological pathway involving water soluble GT enzymes and engineering of biological/metabolic pathways involving the water soluble GTs.

[0055] The GTs according to the present disclosure may be prokaryotic glycosyltransferases or eukaryotic glycosyltransferase (*e.g.*, human glycosyltransferases, rodent glycosyltransferases, yeast glycosyltransferases). Suitable exemplary prokaryotic and eukaryotic glycosyltransferases are identified in FIG. 2A.

[0056] The glycosyltransferase may be selected from the group consisting of fucosyltransferases (FucTs), galactosyltransferases (Gals), glucosyltransferases (GlcTs), mannosyltransferases (ManTs), *N*-acetylgalactosyltransferases (GalNAcTs), *N*-acetylglucosaminyltransferases (GlcNAcTs), and sialyltransferases (SiaTs).

[0057] Fucosyltransferases (FucTs) catalyze the transfer a fucose sugar from a donor substrate to an acceptor substrate. Suitable FucTs include, without limitation, human galactoside 2-alpha-L-fucosyltransferase 1 (*Hs*FUT1), human galactoside 2-alpha-L-fucosyltransferase 2 (*Hs*FUT2), HUMAN Galactoside 3(4)-L-fucosyltransferase (*Hs*FUT3),

human alpha-(1,3)-fucosyltransferase 4 (*Hs*FUT4), human alpha-(1,3)-fucosyltransferase 5 (*Hs*FUT5), human alpha-(1,3)-fucosyltransferase 6 (*Hs*FUT6), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human alpha-(1,6)-fucosyltransferase (*Hs*FUT8), human alpha-(1,3)-fucosyltransferase 9 (*Hs*FUT9), human alpha-(1,3)-fucosyltransferase 10 (*Hs*FUT10), human alpha-(1,3)-fucosyltransferase 11 (*Hs*FUT11), and human GDP-fucose protein *O*-fucosyltransferase 1 (*Hs*POFUT1) (*see*, *e.g.*, FIG. 2A).

[0058] Galactosyltransferases (Gals) catalyze the transfer of a galactose sugar from a donor substrate to an acceptor substrate. Suitable Gals include, without limitation, human beta-1,3-galactosyltransferase 1 (*Hs*B3GalT1), human beta-1,3-galactosyltransferase 2 (*Hs*B3GalT2), human beta-1,4-galactosyltransferase 1 (*Hs*B4GalT1), human beta-1,4-galactosyltransferase 2 (*Hs*B4GalT2), human beta-1,4-galactosyltransferase 3 (*Hs*B4GalT3), human beta-1,4-galactosyltransferase 4 (*Hs*B4GalT4), human beta-1,4-galactosyltransferase 5 (*Hs*B4GalT5), and human beta-1,4-galactosyltransferase 6 (*Hs*B4GalT6) (*see*, *e.g.*, FIG. 2A).

[0059] Glucosyltransferases (GlcTs) catalyze the transfer of a glucose sugar from a donor substrate to an acceptor substrate. Suitable GlcTs include, without limitation, human dolichyl-phosphate beta-glucosyltransferase (*Hs*Alg5), human dolichyl pyrophosphate $man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg6), human probable dolichyl pyrophosphate $Glc1Man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg8), human Dol-P-Glc:$Glc_2Man_9GlcNAc_2$-PP-Dol alpha-1,2-glucosyltransferase (*Hs*Alg10), human ceramide glucosyltransferase (*Hs*UGCG), human beta-1,3-glucosyltransferase (*Hs*B3GLCT), and human protein O-glucosyltransferase 1 (*Hs*POGLUT1) (*see*, *e.g.*, FIG. 2A).

[0060] Mannosyltransferases (ManTs) catalyze the transfer of a mannose sugar from a donor substrate to an acceptor substrate. Suitable ManTs include, without limitation, human chitobiosyldiphosphodolichol beta-mannosyltransferase (*Hs*Alg1), human alpha-1,3/1,6-mannosyltransferase (*Hs*Alg2), human Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase (*Hs*Alg3), human GDP-man:man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Hs*Alg11), human dol-p-man:man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase (*Hs*Alg12), human dolichol-phosphate mannosyltransferase subunit 1 (*Hs*DPM1), human GPI mannosyltransferase 1 (*Hs*PIGM), human GPI mannosyltransferase 3 (*Hs*PIGB), human GPI mannosyltransferase 4 (*Hs*PIGZ), yeast beta-1,4-mannosyltransferase OS=*Saccharomyces cerevisiae* (*Sc*Alg1), and yeast GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Sc*Alg11) (*see*, *e.g.*, FIG. 2A).

[0061] *N*-acetylgalactosyltransferases (GalNAcTs) catalyze the transfer of an *N*-acetylgalactosamine to an acceptor substrate. Suitable GalNAcTs include, without limitation, human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*Hs*ST6GalNAc1), human

alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2 (*Hs*ST6GalNAc2), human alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase (*Hs*ST6GalNAc4), human polypeptide N-acetylgalactosaminyltransferase 1 (*Hs*ppGalNAcT1), human polypeptide N-acetylgalactosaminyltransferase 2 (*Hs*ppGalNAcT2), human polypeptide N-acetylgalactosaminyltransferase 3 (*Hs*ppGalNAcT3), human polypeptide N-acetylgalactosaminyltransferase 4 (*Hs*ppGalNAcT4), human polypeptide N-acetylgalactosaminyltransferase 5 (*Hs*ppGalNAcT5), human polypeptide N-acetylgalactosaminyltransferase 6 (*Hs*ppGalNAcT6), human N-acetylgalactosaminyltransferase 7 (*Hs*ppGalNAcT7), human probable polypeptide N-acetylgalactosaminyltransferase 8 (*Hs*ppGalNAcT8), human polypeptide N-acetylgalactosaminyltransferase 9 (*Hs*ppGalNAcT9), human polypeptide N-acetylgalactosaminyltransferase 10 (*Hs*ppGalNAcT10), and human UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1 (*Hs*B3GALNT1) (*see*, *e.g.*, FIG. 2A).

[0062] *N*-acetylglucosaminyltransferases (GlcNAcTs) catalyze the transfer of an *N*-acetylglucosamine to an acceptor substrate. Suitable GlcNAcTs include, without limitation, human alpha-1,3-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-mannosyl-glycoprotein 4-beta-*N*-acetylglucosaminyltransferase (*Hs*GnTIII/MGAT3), human alpha-1,3-mannosyl-glycoprotein 4-beta-*N*-acetylglucosaminyltransferase a (*Hs*GnTIV/MGAT4), human beta-1,3-galactosyl-*O*-glycosyl-glycoprotein beta-1,6-*N*-acetylglucosaminyltransferase (*Hs*GCNT1), human *N*-acetyllactosaminide beta-1,6-*N*-acetylglucosaminyltransferase (*Hs*GCNT2), human *N*-acetyllactosaminide beta-1,3-*N*-acetylglucosaminyltransferase 2 (*Hs*B3GNT2), human acetylgalactosaminyl-*O*-glycosyl-glycoprotein beta-1,3-*N*-acetylglucosaminyltransferase (*Hs*B3GNT6), human phosphatidylinositol *N*-acetylglucosaminyltransferase subunit A (*Hs*PIGA), *Nicotiana tabacum* alpha-1,3-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase (*Nt*GnTI), and *Nicotiana tabacum* alpha-1,6-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase-like (*Nt*GnTII) (*see*, *e.g.*, FIG. 2A).

[0063] Sialyltransferases (SiaTs) catalyze the transfer of sialic acid to an acceptor substrate. Suitable SiaTs include, without limitation, human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase (*Hs*ST3Gal3), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4 (*Hs*ST3Gal4), human type 2 lactosamine alpha-2,3-sialyltransferase (*Hs*ST3Gal6), human beta-galactoside alpha-2,6-sialyltransferase 1 (*Hs*ST6Gal1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*Hs*ST6GalNAc1), human alpha-N-acetylgalactosaminide alpha-2,6-

sialyltransferase 2 (*Hs*ST6GalNAc2), human alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase (*Hs*ST6GalNAc4), human alpha-N-acetylneuraminide alpha-2,8-sialyltransferase (*Hs*ST8Sia1), human alpha-2,8-sialyltransferase 8b (*Hs*ST8Sia2), human sia-alpha-2,3-gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase (*Hs*ST8Sia3), human CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase (*Hs*ST8Sia4), and *Neisseria meningitidis* alpha-2,9-polysialyltransferase (*Nm*SynE) (*see*, *e.g.*, FIG. 2A).

[0064] In some embodiments, the glycosyltransferase is selected from the group consisting of human galactoside 2-alpha-L-fucosyltransferase 1 (*Hs*FUT1), human galactoside 2-alpha-L-fucosyltransferase 2 (*Hs*FUT2), HUMAN Galactoside 3(4)-L-fucosyltransferase (*Hs*FUT3), human alpha-(1,3)-fucosyltransferase 4 (*Hs*FUT4), human alpha-(1,3)-fucosyltransferase 5 (*Hs*FUT5), human alpha-(1,3)-fucosyltransferase 6 (*Hs*FUT6), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human alpha-(1,6)-fucosyltransferase (*Hs*FUT8), human alpha-(1,3)-fucosyltransferase 9 (*Hs*FUT9), human alpha-(1,3)-fucosyltransferase 10 (*Hs*FUT10), human alpha-(1,3)-fucosyltransferase 11 (*Hs*FUT11), human GDP-fucose protein O-fucosyltransferase 1 (*Hs*POFUT1), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase (*Hs*ST3Gal3), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4 (*Hs*ST3Gal4), human type 2 lactosamine alpha-2,3-sialyltransferase (*Hs*ST3Gal6), human beta-galactoside alpha-2,6-sialyltransferase 1 (*Hs*ST6Gal1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*Hs*ST6GalNAc1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2 (*Hs*ST6GalNAc2), human alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase (*Hs*ST6GalNAc4), human alpha-N-acetylneuraminide alpha-2,8-sialyltransferase (*Hs*ST8Sia1), human alpha-2,8-sialyltransferase 8b (*Hs*ST8Sia2), human sia-alpha-2,3-gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase (*Hs*ST8Sia3), human CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase (*Hs*ST8Sia4), human polypeptide N-acetylgalactosaminyltransferase 1 (*Hs*ppGalNAcT1), human polypeptide N-acetylgalactosaminyltransferase 2 (*Hs*ppGalNAcT2), human polypeptide N-acetylgalactosaminyltransferase 3 (*Hs*ppGalNAcT3), human polypeptide N-acetylgalactosaminyltransferase 4 (*Hs*ppGalNAcT4), human polypeptide N-acetylgalactosaminyltransferase 5 (*Hs*ppGalNAcT5), human polypeptide N-acetylgalactosaminyltransferase 6 (*Hs*ppGalNAcT6), human N-acetylgalactosaminyltransferase 7 (*Hs*ppGalNAcT7), human probable polypeptide N-acetylgalactosaminyltransferase 8 (*Hs*ppGalNAcT8), human polypeptide N-acetylgalactosaminyltransferase 9 (*Hs*ppGalNAcT9), human polypeptide N-acetylgalactosaminyltransferase 10 (*Hs*ppGalNAcT10), human UDP-

GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1 (*Hs*B3GALNT1), human beta-1,4 N-acetylgalactosaminyltransferase 1 (*Hs*B4GALNT1), human histo-blood group ABO system transferase (*Hs*-A-group), human lactosylceramide 4-alpha-galactosyltransferase (*Hs*A4GalT), human beta-1,3-galactosyltransferase 1 (*Hs*B3GalT1), human beta-1,3-galactosyltransferase 2 (*Hs*B3GalT2), human beta-1,4-galactosyltransferase 1 (*Hs*B4GalT1), human beta-1,4-galactosyltransferase 2 (*Hs*B4GalT2), human beta-1,4-galactosyltransferase 3 (*Hs*B4GalT3), human beta-1,4-galactosyltransferase 4 (*Hs*B4GalT4), human beta-1,4-galactosyltransferase 5 (*Hs*B4GalT5), human beta-1,4-galactosyltransferase 6 (*Hs*B4GalT6), human histo-blood group ABO system transferase (*Hs*-B-group), human 2-hydroxyacylsphingosine 1-beta-galactosyltransferase (*Hs*UGT8), human glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 (*Hs*C1GLT), human C1GALT1-specific chaperone 1 (*Hs*COSMC), human chitobiosyldiphosphodolichol beta-mannosyltransferase (*Hs*Alg1), human alpha-1,3/1,6-mannosyltransferase (*Hs*Alg2), human Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase (*Hs*Alg3), human GDP-man:man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Hs*Alg11), human dol-p-man:man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase (*Hs*Alg12), human isoform 2 of putative UDP-N-acetylglucosamine transferase (*Hs*Alg13), human UDP-N-acetylglucosamine transferase subunit alg14 homolog (*Hs*Alg14), human dolichol-phosphate mannosyltransferase subunit 1 (*Hs*DPM1), human GPI mannosyltransferase 1 (*Hs*PIGM), human GPI mannosyltransferase 3 (*Hs*PIGB), human GPI mannosyltransferase 4 (*Hs*PIGZ), human dolichyl-phosphate beta-glucosyltransferase (*Hs*Alg5), human dolichyl pyrophosphate $man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg6), human probable dolichyl pyrophosphate $Glc1Man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg8), human Dol-P-Glc:$Glc_2Man_9GlcNAc_2$-PP-Dol alpha-1,2-glucosyltransferase (*Hs*Alg10), human ceramide glucosyltransferase (*Hs*UGCG), human beta-1,3-glucosyltransferase (*Hs*B3GLCT), human glycogenin-1 (*Hs*GLYG), human protein O-glucosyltransferase 1 (*Hs*POGLUT1), human alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (*Hs*GnTIII/MGAT3), human alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase a (*Hs*GnTIV/MGAT4), human beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (*Hs*GCNT1), human N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase (*Hs*GCNT2), human N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2 (*Hs*B3GNT2), human acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase (*Hs*B3GNT6), human phosphatidylinositol N-acetylglucosaminyltransferase subunit A

(*Hs*PIGA), human xyloside xylosyltransferase 1 (*Hs*XXLT1), human UDP-glucuronosyltransferase 1-1 (*Hs*UGT1A1), human beta-1,4-glucuronyltransferase 1 (*Hs*B4GAT1), human UDP-glucuronosyltransferase 1-3 (*Hs*UGT1A3), *Campylobacter jejuni* CsTII (*Cj*CstII), *Neisseria meningitidis* polysialic acid *O*-acetyltransferase (*Nm*Pst), *Campylobacter jejuni* beta-1,3-galactosyltransferase (*Cj*CgtB), *Helicobacter pylori* (strain 51) beta-4-galactosyltransferase (*Hp*LgtB), *Neisseria meningitidis* serogroup B (strain MC58) lacto-N-neotetraose biosynthesis glycosyltransferase LgtB (*Nm*LgtB), *Neisseria gonorrhoeae* lacto-N-neotetraose biosynthesis glycosyltransferase (*Ng*LgtB), *E.coli* galactoside 2-alpha-L-fucosyltransferase WbgL (*Ec*WbgL), *E.coli* undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (*Ec*WecA), *Legionella pneumophila* subsp. Pneumophila Subversion of eukaryotic traffic protein A (*Lp*SetA), *Neisseria meningitidis* alpha-2,9-polysialyltransferase (*Nm*SynE), yeast beta-1,4-mannosyltransferase OS=*saccharomyces cerevisiae* (*Sc*Alg1), yeast GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Sc*Alg11), *Nicotiana tabacum* alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Nt*GnTI), *Nicotiana tabacum* alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like (*Nt*GnTII), *Bos taurus* n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Bt*GGTA1), mouse n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Mm*GGTA1), rat n-acetyllactosaminide alpha-1,3-galactosyltransferase (RnGGTA1), and *Bos taurus* beta-1,4-galactosyltransferase 1 (BtB4GalT1) (*see*, *e.g.*, FIG. 2A).

[0065] In some embodiments, the nucleic acid molecule encodes a second nucleic acid moiety encoding a glycosyltransferase having the amino acid sequence of any one of SEQ ID NOs: 1–174 (*see* FIGS. 2C–2D).

[0066] The Examples of the present disclosure demonstrate the use of tripartite glycosyltransferase fusion proteins (*e.g.*, Sx-*Cj*CstII , Sx-Δ36*Hs*FucT7, Sx-Δ34*Hs*ST3Gal1, Sx-Δ29*Hs*GnTI, Sx-Δ29*Hs*GnTII, Sx-Δ44*Hs*β4GalT1, Sx-Δ26*Hs*ST6Gal1, Sx-Δ44*Hs*FucT8) to catalyze the formation of a spectrum of homogenous *N*-glycan structures on intact glycoproteins. Thus, in some embodiments, the glycosyltransferase is selected from the group consisting of *Campylobacter jejuni* CsTII (*Cj*CstII), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-galactosyltransferase 1 (*Hs*β4GalT1), human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1), and human alpha-(1,6)-fucosyltransferase (*Hs*FUT8).

[0067] As used herein, the term "amphipathic shield domain protein" includes any protein that displays both hydrophilic and hydrophobic surfaces and is often associated with lipids as membrane anchors or involved in their transport as soluble particles. The amphipathic shield domain protein, in one embodiment, serves as a molecular shield to sequester large lipophilic surfaces of the glycosyltransferase from water.

[0068] In various other embodiments, the amphipathic shield domain protein is selected from the group consisting of apolipoprotein A (ApoA), apolipoprotein B (ApoB), apolipoprotein C (ApoC), apolipoprotein D (ApoD), apolipoprotein E (ApoE), apolipoprotein H (ApoH), truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*), and a peptide self-assembly mimic (PSAM). In particular, the amphipathic shield domain protein may be apolipoprotein A1 (ApoAI). As used herein, ApoAI avidly binds phospholipid molecules and organizes them into soluble bilayer structures or discs that readily accept cholesterol. ApoAI contains a globular amino-terminal (N-terminal) domain (residues 1–43) and a lipid-binding carboxyl-terminal (C-terminal) domain (residues 44–243). In some embodiments, the amphipathic shield domain protein is human apolipoprotein A1. The apolipoprotein A1 may be a truncated human apolipoprotein A1. Truncated variants of ApoA1 include, but are not limited to, human ApoAI lacking its 43-residue globular N-terminal domain (ApoA1*).

[0069] As used herein, ApoA1 exhibits remarkable structural flexibility, and may adopt a molten globular-like state for lipid-free ApoAI under conditions that may allow it to adapt to the significant geometry changes of the lipids with which it interacts. The present disclosure provides tripartite fusion proteins in which, for example, ApoAI* may be genetically fused to the carboxyl terminus of a glycosyltransferase (or truncated glycosyltransferase). As described herein, expression of such tripartite glycosyltransferase fusion proteins may yield appreciable amounts of globular, water-soluble tripartite glycosyltransferase fusion proteins that are stabilized in a hydrophobic environment and retain structurally relevant conformations. The approach provides, *inter alia*, a facile method for efficiently solubilizing structurally diverse glycosyltransferases, for example in both prokaryotic and eukaryotic cells, without the need for detergents or lipid reconstitutions.

[0070] As used herein, the term "water soluble expression decoy protein" includes any protein which serves to direct an glycosyltransferase into cellular cytoplasm. The water soluble expression decoy protein may assist in "tricking" a hydrophobic glycosyltransferase into thinking that it is not hydrophobic. The water soluble expression decoy protein may be selected from the group consisting of outer surface protein (OspA) lacking its native export signal peptide, DnaB lacking its native export signal peptide, and maltose-binding protein (MBP)

lacking its N-terminal signal peptide. In some embodiments, the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide.

[0071] In some embodiments of the nucleic acid constructs and the tripartite glycosyltransferase fusion proteins according to the present disclosure, the amphipathic shield domain protein is truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*) and the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide. For example, the nucleic acid construct may comprise a chimeric nucleic acid molecule comprising a first nucleic acid moiety encoding truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*), a second nucleic acid moiety encoding human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1) or a truncated *Hs*ST6Gal1 variant in which 26 amino acids from the N-terminus of HsST6Gal1 comprising its CT and TMD were deleted (Δ26*Hs*ST6Gal1), and a third nucleic acid moiety encoding maltose-binding protein (MBP) lacking its N-terminal signal peptide (ΔspMBP). Such embodiments are described in Example 1, where ΔspMBP-*Hs*ST6Gal1-ApoAI* (abbreviated as Sx-*Hs*ST6Gal1) and ΔspMBP-Δ26*Hs*ST6Gal1-ApoAI* (abbreviated as Sx-Δ26*Hs*ST6Gal1) are shown to accumulate almost exclusively in the soluble cytoplasmic fraction of *E. coli* cells. The importance of the amphipathic shield domain and water soluble expression decoy proteins is evidenced by expression of unfused *Hs*ST6Gal1 and Δ26*Hs*ST6Gal1, which were not observed to accumulate in the soluble fraction and were only observed in minimal amounts in the insoluble and detergent-solubilized fractions.

[0072] In some embodiments, the construct further includes a promoter and a termination sequence, where the promoter and the termination sequence are operatively coupled to the chimeric nucleic acid molecule.

[0073] The chimeric nucleic acid molecules of the present disclosure include DNA molecules (*e.g.*, linear, circular, cDNA, chromosomal, genomic, or synthetic, double stranded, single stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hair-pinned, circular, or in a padlocked conformation) and RNA molecules (*e.g.*, tRNA, rRNA, mRNA, genomic, or synthetic) and analogs of the DNA or RNA molecules of the described as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native inter-nucleoside bonds, or both.

[0074] In some embodiments, the first nucleic acid moiety, the second nucleic acid moiety, and/or the third nucleic acid moiety may be free of naturally flanking sequences (*i.e.*, sequences located at the 5' and 3' ends of the first nucleic acid moiety, the second nucleic acid moiety, and/or the third nucleic acid moiety) in the chromosomal DNA of the organism from

which the first nucleic acid moiety, the second nucleic acid moiety, and/or the third nucleic acid moiety was derived, respectively.

[0075] In various embodiments, the first nucleic acid moiety, the second nucleic acid moiety, and/or the third nucleic acid moiety may contain less than about 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, 50 bp, 25 bp or 10 bp of naturally flanking nucleotide chromosomal DNA sequences of the microorganism from which the first nucleic acid moiety, the second nucleic acid moiety, and/or the third nucleic acid moiety was derived, respectively.

[0076] The chimeric nucleic acid molecules may further include one or more linker nucleic acid moieties coupling the first, second, and/or third nucleic acid moieties together.

[0077] The tripartite glycosyltransferase fusion proteins according to the present disclosure include a continuous polymer of amino acids which comprise the full or partial sequence of three or more distinct proteins. The construction of fusion proteins is well-known in the art. Two or more amino acids sequences may be joined chemically, for instance, through the intermediacy of a crosslinking agent. For example, a fusion protein may be generated by expression of a nucleic acid construct comprising a chimeric nucleic acid molecule according to the present disclosure in a host cell. Such nucleic acid constructs may generally also contain replication origins active in host cells and one or more selectable markers encoding, for example, drug or antibiotic resistance.

[0078] The tripartite glycosyltransferase fusion proteins of the present disclosure can be generated as described herein or using any other standard technique known in the art. For example, the tripartite glycosyltransferase fusion proteins can be prepared by translation of a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein according to the present disclosure. The chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein is inserted into an expression vector which is used to transform or transfect a host cell.

[0079] Different chimeric nucleic acid molecules encoding unique tripartite glycosyltransferase fusion proteins may be present on separate nucleic acid constructs or on the same nucleic acid construct. Inclusion of different chimeric nucleic acid molecules encoding unique tripartite glycosyltransferase fusion proteins on the same nucleic acid molecule is advantageous, in that uptake of only a single species of nucleic acid by a host cell is sufficient to introduce sequences encoding the tripartite glycosystransferase(s) into the host cell. By contrast, when different chimeric nucleic acid molecules encoding unique tripartite glycosyltransferase fusion proteins are present on different nucleic acid constructs, both nucleic acid molecules are taken up by a particular host cell for the assay to be functional.

[0080] A nucleic acid construct comprising a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion proteins may be inserted into an expression system to which the nucleic acid construct is heterologous. The heterologous nucleic acid construct may be inserted into the expression system or vector in proper sense (5'→3') orientation relative to the promoter and any other 5' regulatory molecules, and correct reading frame. The preparation of the nucleic acid constructs can be carried out using standard cloning methods well known in the art as described by Sambrook et al., Molecular Cloning: *A Laboratory Manual,* Cold Springs Laboratory Press, Cold Springs Harbor, New York (1989), which is hereby incorporated by reference in its entirety. U.S. Patent No. 4,237,224 to Cohen and Boyer, which is hereby incorporated by reference in its entirety, also describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase.

[0081] Another aspect of the present disclosure is directed to tripartite glycosyltransferase fusion proteins produced by the host cells described herein.

[0082] As described herein, a variety of prokaryotic expression systems can be used to express the tripartite glycosyltransferase fusion proteins of the present disclosure. Expression vectors can be constructed which contain a promoter to direct transcription, a ribosome binding site, and a transcriptional terminator. Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway (Yanofsky et al., "Repression is Relieved Before Attenuation in the trp Operon of *Escherichia coli* as Tryptophan Starvation Becomes Increasingly Severe," *J. Bacteria.* 158:1018-1024 (1984), which is hereby incorporated by reference in its entirety) and the leftward promoter of phage lambda (N) (Herskowitz et al., "The Lysis-lysogeny Decision of Phage Lambda: Explicit Programming and Responsiveness," *Ann. Rev. Genet.,* 14:399-445 (1980), which is incorporated by reference in its entirety). Vectors used for expressing foreign genes in bacterial hosts generally will contain a sequence for a promoter which functions in the host cell. Plasmids useful for transforming bacteria include pBR322 (Bolivar et al., "Construction and Characterization of New Cloning Vehicles II. A Multipurpose Cloning System," *Gene* 2:95-113 (1977), which is hereby incorporated by reference in its entirety), the pUC plasmids (Messing, "New M13 Vectors for Cloning," *Meth. Enzymol.* 101:20-77 (1983), Vieira et al., "New pUC-derived Cloning Vectors with Different Selectable Markers and DNA Replication Origins," *Gene* 19:259-268 (1982) which are hereby incorporated by reference in their entirety), and derivatives thereof. Plasmids may contain both viral and bacterial elements. Methods for the recovery of the proteins in biologically active form are discussed in U.S. Patent Nos. 4,966,963 to Patroni and 4,999,422 to Galliher, which are incorporated herein by reference in their entirety. Suitable expression vectors

include those which contain replicon and control sequences that are derived from species compatible with the host cell. For example, if *E. coli* is used as a host cell, plasmids such as pUC19, pUC18 or pBR322 may be used. Alternatively, plasmids such as pET28a and pMALc2x may be used. Other suitable expression vectors are described in *Molecular Cloning: a Laboratory Manual:* 3rd edition, Sambrook and Russell, 2001, Cold Spring Harbor Laboratory Press, which is hereby incorporated by reference in its entirety. Many known techniques and protocols for manipulation of nucleic acids, for example in preparation of nucleic acid constructs, mutagenesis, sequencing, introduction of DNA into cells and gene expression, and analysis of proteins, are described in detail in *Current Protocols in Molecular Biology*, Ausubel et al. eds., (1992), which is hereby incorporated by reference in its entirety.

[0083] Different genetic signals and processing events control many levels of gene expression (*e.g.*, DNA transcription and messenger RNA ("mRNA") translation) and subsequently the amount of fusion protein that is displayed on the ribosome surface. Transcription of DNA is dependent upon the presence of a promoter, which is a DNA sequence that directs the binding of RNA polymerase, and thereby promotes mRNA synthesis. Promoters vary in their "strength" *(i.e.*, their ability to promote transcription). For the purposes of expressing a cloned gene, it is desirable to use strong promoters to obtain a high level of transcription and, hence, expression and surface display. Therefore, depending upon the host system utilized, any one of a number of suitable promoters may also be incorporated into the expression vector carrying the deoxyribonucleic acid molecule encoding the protein of interest coupled to a stall sequence. For instance, when using *E. coli*, its bacteriophages, or plasmids, promoters such as the T7 phage promoter, *lac* promoter, *trp* promoter, *rec*A promoter, ribosomal RNA promoter, the P_R and P_L promoters of coliphage lambda and others, including but not limited, to *lac*UV5, *omp*F, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lac*UV5 (*tac*) promoter or other *E. coli* promoters produced by recombinant DNA or other synthetic DNA techniques may be used to provide for transcription of the inserted gene.

[0084] Translation of mRNA in prokaryotes depends upon the presence of the proper prokaryotic signals, which differ from those of eukaryotes. Efficient translation of mRNA in prokaryotes requires a ribosome binding site called the Shine-Dalgarno ("SD") sequence on the mRNA. This sequence is a short nucleotide sequence of mRNA that is located before the start codon, usually AUG, which encodes the amino-terminal methionine of the protein. The SD sequences are complementary to the 3'-end of the 16S rRNA (ribosomal RNA) and probably promote binding of mRNA to ribosomes by duplexing with the rRNA to allow correct positioning of the ribosome. For a review on maximizing gene expression, *see* Roberts and

Lauer, "Maximizing Gene Expression on a Plasmid Using Recombination *In Vitro*," *Methods in Enzymology* 68:473-82 (1979), which is hereby incorporated by reference in its entirety.

[0085] In accordance with this and other aspects of the present disclosure, the amphipathic shield domain protein, glycosyltransferase, and/or water soluble expression decoy proteins are linked either directly or via a linker located adjacent to each other within the construct, coupled to each other in tandem or separated by at least one linker. In one embodiment, the chimeric nucleic acid molecule includes a linker coupling the nucleic acid moieties together. Likewise, the tripartite glycosyltransferase fusion proteins may include a linker coupling the amphipathic shield domain protein, the glycosyltransferase (or truncated glycosyltransferase), and the water soluble expression decoy protein together. The amphipathic shield domain protein, the glycosyltransferase (or truncated glycosyltransferase), and the water soluble expression decoy protein may be linked by a covalent linkage or may be linked by methods known in the art for linking peptides.

[0086] Linkers may include synthetic sequences of amino acids that are commonly used to physically connect polypeptide domains to each other or to biologically relevant moieties. Most linker peptides are composed of repetitive modules of one or more of the amino acids glycine and serine. Peptide linkers have been well-characterized and shown to adopt unstructured, flexible conformations. For example, linkers comprised of Gly and Ser amino acids have been found to not interfere with assembly and binding activity of the domains it connects. Freund et al., "Characterization of the Linker Peptide of the Single-chain Fv Fragment of an Antibody by NMR Spectroscopy," *FEBS* 320:97 (1993), which is hereby incorporated by reference in its entirety.

[0087] The nucleic acid constructs and tripartite glycosyltransferase fusion proteins of the present disclosure may include a flexible polypeptide linker separating the amphipathic shield domain protein, glycosyltransferase (or truncated glycosyltransferase), and/or water soluble expression decoy proteins and allowing for their independent folding. The linker is optimally 15 amino acids or 60 Å in length (~4 Å per residue) but may be as long as 30 amino acids but preferably not more than 20 amino acids in length. It may be as short as 3 amino acids in length, but more preferably is at least 6 amino acids in length. To ensure flexibility and to avoid introducing steric hindrance that may interfere with the independent folding of the fragment domain of reporter protein and the members of the putative binding pair, the linker should be comprised of small, preferably neutral residues such as Gly, Ala, and Val, but also may include polar residues that have heteroatoms such as Ser and Met, and may also contain charged residues. The first, second, and third proteins may be linked via a short polypeptide linker sequence. Suitable linkers include peptides of between about 2 and about 40 amino acids in length and may

include, for example, glycine residues Gly185 and Gly186. Preferred linker sequences include glycine-rich (*e.g.* G_{3-5}), serine-rich (*e.g.*, GSG, GSGS, GSGSG, GS_NG), or alanine rich (*e.g.*, TSAAA) linker sequences. Other exemplary linker sequences have a combination of glycine, alanine, proline and methionine residues such as AAAGGM; AAAGGMPPAAAGGM (SEQ ID NO: 175); AAAGGM; and PPAAAGGMM. Linkers may have virtually any sequence that results in a generally flexible chimeric protein.

[0088] Another aspect of the present disclosure relates to an expression vector including the nucleic acid construct of the present disclosure. Suitable nucleic acid vectors include, without limitation, plasmids, baculovirus vectors, bacteriophage vectors, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral vectors (for example, viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, and the like), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and other vectors. In some embodiments of the present disclosure, vectors suitable for use in prokaryotic host cells. Accordingly, exemplary vectors for use in prokaryotes such as *Escherichia coli* include, but are not limited to, pACYC184, pBeloBac11, pBR332, pBAD33, pBBR1MCS and its derivatives, pSC101, SuperCos (cosmid), pWE15 (cosmid), pTrc99A, pBAD24, vectors containing a ColE1 origin of replication and its derivatives, pUC, pBluescript, pGEM, and pTZ vectors.

[0089] Another aspect of the present disclosure relates to a host cell comprising the nucleic acid construct of the present disclosure. In accordance with this and other aspects of the present disclosure, suitable host cells include both eukaryotic and prokaryotic cells.

[0090] In some embodiments, the host cell is eukaryotic. Eukaryotic host cells, include without limitation, animal cells, fungal cells, insect cells, plant cells, and algal cells. In some embodiments, the eukaryotic host cells are selected from the group consisting of human cells, yeast, cells, and cell lines. Suitable eukaryotic host cells include, but are not limited to, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thennotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii*, and the like. In some embodiments, the eukaryotic host cell is a yeast cell and the yeast cell strain is SBY49. In some embodiments, the eukaryotic host cell is a human cell. Exemplary human cells lines include, without limitation, HEK293T (ATCC), FreeStyle™ 293-F (Thermo Fisher), and Expi293F™ GnTI- (Thermo Fisher).

[0091] In accordance with the present disclosure, the host cell may be prokaryotic, such as a bacterial cell. Such cells serve as a host for expression of recombinant proteins for production of recombinant therapeutic proteins of interest. Suitable microorganisms include *Pseudomonas sp.* such as *Pseudomonas aeruginosa, Escherichia sp., Escherichia coli* and other Enterobacteriaceae, *Salmonella sp.* such as *Salmonella gastroenteritis (typhimirium), S. typhi, S. enteriditis, Shigella sp.* such as *Shigella flexneri, S. sonnie, S dysenteriae, Neisseria sp.* such as *Neisseria gonorrhoeae, N. meningitides, Haemophilus sp.* including *Haemophilus influenzae H. pleuropneumoniae, Pasteurella sp.* including *Pasteurella haemolytica, P. multilocida, Legionella sp.* such as *Legionella pneumophila, Treponema pallidum, T. denticola, T. orales, Borrelia burgdorferi, Borrelia spp. Leptospira interrogans, Klebsiella sp.* such as *Klebsiella pneumoniae, Proteus vulgaris, P. morganii, P. mirabilis, Rickettsia prowazeki, R.typhi, R. richettsii, Porphyromonas (Bacteriodes) gingivalis, Chlamydia psittaci, C. pneumoniae, C. trachomatis, Campylobacter sp.* such as *Campylobacter jejuni, C. intermedis, C. fetus, Helicobacter sp.* such as *Helicobacter pylori, Francisella sp.* such as *Francisella tularenisis, Vibrio cholerae, Vibrio parahaemolyticus, Bordetella sp.* including *Bordetella pertussis, Burkholderia sp.* such as *Burkholderie pseudomallei, Brucella sp.* including *Brucella abortus, B. susi, B. melitens is, B. can is, Spirillum minus, Pseudomonas mallei, Aeromonas sp.* such as *Aeromonas hydrophila, A salmonicida,* and *Yersinia sp.* such as *Yersinia pestis.* Additional microorganisms include *Wolinella sp., Desulfovibrio sp. Vibrio sp., Bacillus sp., Listeria sp., Staphylococcus sp., Streptococcus sp., Peptostreptococcus sp., Megasphaera sp., Pectinatus sp., Selenomonas sp., Zymophilus sp., Actinomyces sp., Arthrobacter sp., Frankia sp., Micromonospora sp., Nocardia sp., Propionibacterium sp., Streptomyces sp., Lactobacillus sp., Lactococcus sp., Leuconostoc sp., Pediococcus sp., Acetobacterium sp., Eubacterium sp., Heliobacterium sp., Heliospirillum sp., Sporomusa sp., Spiroplasma sp., Ureaplasma sp., Erysipelothrix, sp., Corynebacterium sp. Enterococcus sp., Clostridium sp., Mycoplasma sp., Mycobacterium sp., Actinobacteria sp., Moraxella sp., Stenotrophomonas sp., Micrococcus sp., Bdellovibrio sp., Hemophilus sp., Proteus mirabilis, Enterobacter cloacae, Serratia sp., Citrobacter sp., Proteus sp., , Acinetobacter sp., Actinobacillus sp., Capnocytophaga sp., Cardiobacterium sp., Eikenella sp., Kingella sp., Flavobacterium sp. Xanthomonas sp., Plesiomonas sp.,* and alpha-proteobacteria such as *Wolbachia sp.*, cyanobacteria, spirochaetes, green sulfur and green non-sulfur bacteria, Gram-negative cocci, Gram negative bacilli which are fastidious, Enterobacteriaceae -glucose-fermenting gram-negative bacilli, Gram negative bacilli - non-glucose fermenters, Gram negative bacilli - glucose fermenting, oxidase positive. In some embodiments, the prokaryotic host cells is an *E. coli* cells such as DH5α, BL21 (DE3), SHuffle® T7 Express *lysY*, and Origami2(DE3) *gmd::kan ΔwaaL*.

[0092] Methods for transforming / transfecting host cells with expression vectors are well-known in the art and depend on the host system selected as described in Sambrook et al., Molecular Cloning: *A Laboratory Manual,* Cold Springs Laboratory Press, Cold Springs Harbor, New York (1989), which is hereby incorporated by reference in its entirety.

[0093] The present disclosure is also directed to tripartite glycosyltransferase fusion proteins produced by the host cells of the present disclosure.

[0094] When the nucleic acid construct is assembled in a host cell, the host cell may be cultured in a suitable culture medium optionally supplemented with one or more additional agents, such as an inducer (*e.g.*, where a nucleotide sequence encoding a chimeric protein is under the control of an inducible promoter). The inducer may be, for example, isopropyl-β-D-thiogalactoside. In one embodiment of the present disclosure, a substrate is endogenous to the host cell and upon assembly of the nucleic acid construct in the host cell, the substrate is readily converted. In another embodiment, a substrate is exogenous to the host cell. In accordance with this embodiment, the culture medium is supplemented with a substrate or substrate precursor that can be readily taken up by the host cell and converted. Suitable substrates include, without limitation, proteins, nucleic acid molecules, organic compounds, lipids, and glycans.

[0095] In some embodiments of the present disclosure, the tripartite glycosyltransferase fusion protein is separated from other products, macromolecules, etc., which may be present in the cell culture medium, the cell lysate, or the organic layer. Separation of the tripartite glycosyltransferase fusion protein from other products that may be present in the cell culture medium, cell lysate, or organic layer is readily achieved using standard methods known in the art, *e.g.*, standard chromatographic techniques. Several methods are readily known in the art, including ion exchange chromatography, high performance liquid chromatography, hydrophobic interaction chromatography, affinity chromatography (*e.g.*, Ni^{2+} affinity chromatography), size exclusion chromatography, gel filtration, and reverse phase chromatography. The tripartite glycosyltransferase fusion protein is preferably produced in purified form (at least about 40% pure, at least about 50% pure, at least about 60% pure, at least about 70% pure, at least about 80% pure, at least about 90% pure, at least about 95% pure, at least about 98%, or more than 98% pure) by conventional techniques. Depending on whether the host cell is made to secrete the protein into growth medium (*see* U.S. Patent No. 6,596,509 to Bauer et al., which is hereby incorporated by reference in its entirety), the protein can be isolated and purified by centrifugation (to separate cellular components from supernatant containing the secreted protein) followed by sequential ammonium sulfate precipitation of the supernatant. The fraction containing the protein can be subjected to gel filtration in an appropriately sized dextran or polyacrylamide column to separate the protein from other cellular components and proteins. If

necessary, the protein fraction may be further purified by HPLC. Accordingly, the tripartite glycosyltransferase fusion protein produced by the present disclosure can be used to isolate and solubilize a glycosyltransferase in a purified form, *e.g.*, "pure" in the context of a tripartite glycosyltransferase fusion protein that is free from other intermediate or precursor products, macromolecules, contaminants, etc.

[0096] Expression of soluble tripartite glycosyltransferase fusion proteins may be increased by at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 100% (or two-fold), as compared to when the corresponding glycosyltransferase is expressed in the absence of the amphipathic shield domain protein and/or water soluble expression decoy protein. In other embodiments of the present disclosure, the expression of tripartite glycosyltransferase fusion proteins from the nucleic acid constructs disclosed herein is at least about 2.5-fold, at least about 3-fold, at least about 5-fold, at least about 7-fold, at least about 10-fold, at least about 15-fold, at least about 20-fold, at least about 50-fold, at least about 100-fold, or more, higher compared to the expression level of the glycosyltransferase from nucleic acid constructs lacking the first nucleic acid moiety encoding an amphipathic shield domain protein and/or the third nucleic acid moiety encoding a water soluble expression decoy protein. Likewise, the expression of tripartite glycosyltransferase fusion proteins from the nucleic acid constructs disclosed herein may be at least about 2.5-fold, at least about 3-fold, at least about 5-fold, at least about 7-fold, at least about 10-fold, at least about 15-fold, at least about 20-fold, at least about 50-fold, at least about 100-fold, or more, higher compared to the expression of a corresponding wild type glycosyltransferase protein, which is not fused to a heterologous amphipathic shield domain protein and/or a water soluble expression decoy protein.

[0097] Methods for transforming/transfecting host cells with expression vectors are well-known in the art and depend on the host system selected, as described in Sambrook et al., Molecular Cloning: *A Laboratory Manual,* Cold Springs Laboratory Press, Cold Springs Harbor, New York (1989), which is hereby incorporated by reference in its entirety. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation, and transfection using bacteriophage.

[0098] The simplest single-celled organisms are composed of central regions filled with an aqueous material and a variety of soluble small molecules and macromolecules. Enclosing this central region is a membrane which is composed of phospholipids arranged in a bilayer

structure. In more complex living cells, there are internal compartments and structures that are also enclosed by membranes. There are many protein molecules embedded or associated within these membrane structures, and these membrane proteins are often the most important to determining cell functions including communication and processing of information and energy. The largest problem in studying membrane proteins is that the inside of the phospholipid bilayer is hydrophobic and the embedded or anchored part of the membrane protein is itself also hydrophobic. In isolating these membrane proteins from their native membrane environments, the present disclosure overcomes the difficult task of preventing recombinant glycosyltransferases from forming inactive aggregates while remaining in a native configuration. In one embodiment of the present disclosure, a tripartite glycosyltransferase fusion protein is encoded by the nucleic acid construct of the present disclosure and, preferably, the tripartite glycosyltransferase fusion protein is in water soluble form. The term "solubilizing" according to the present disclosure includes dissolving a molecule in a solution. This aspect of the disclosure is carried out in substantially the same way as described above.

[099] The present disclosure is also directed to tripartite glycosyltransferase fusion proteins produced by the host cells of the present disclosure.

[0100] In addition to cell-based expression hosts/systems of the present disclosure, the tripartite glycosyltransferase fusion proteins may also be expressed using cell-free expression platforms. Thus, another aspect of the present disclosure relates to a cell-free protein expression system. The cell-free protein expression system comprises a cell lysate or extract and a nucleic acid construct according to the present disclosure. The cell lysate or extract may include a heterologous and/or recombinant RNA polymerase. In some embodiments, the cell lysate or extract is capable of (i) transcribing the nucleic acid construct or the vector to form a translation template and (ii) translating the translation template. In some embodiments, the cell lysate or extract is an *E. coli* lysate or extract. Examples of cell-free expression platforms include, but are not limited to, the PURExpress kit from NEB and S30 lysate high-expression kit from Promega, among others.

[0101] The present disclosure is also directed to tripartite glycosyltransferase fusion proteins produced by the cell-free protein expression systems of the present disclosure.

[0102] Another aspect of the present disclosure relates to a method of recombinantly producing a tripartite glycosyltransferase fusion protein in water soluble form. This method involves providing a host cell according to the present disclosure or a cell-free expression system according to the present disclosure. The method further involves culturing the host cell or using the cell-free expression system under conditions effective to express the tripartite

glycosyltransferase fusion protein in a water soluble form within the host cell cytoplasm or the cell-free expression system.

[0103] In some embodiments, the method further includes recovering the tripartite glycosyltransferase fusion protein from the host cell or the cell-free expression system following the culturing or the using, respectively. The tripartite glycosyltransferase fusion protein may be recovered from the cell's cytoplasm. The recovery of the tripartite glycosyltransferase fusion protein from the host cell is consistent with the recovery of proteins discussed *supra*.

[0104] In some embodiments where the host cell is provided, the recovering involves lysing the cell to form a cell lysate comprising a water soluble fraction and subjecting the water soluble fraction of the cell lysate to chromatography to isolate the tripartite glycosyltransferase fusion protein.

[0105] In some embodiments where the cell-free expression system is provided, the recovering involves subjecting the water soluble fraction of the cell lysate to chromatography to isolate the tripartite glycosyltransferase fusion protein.

[0106] In one embodiment of this aspect of the present disclosure, the tripartite glycosyltransferase fusion proteins are provided in a purified isolated form.

[0107] The tripartite glycosyltransferase fusion protein can be synthesized using standard methods of protein/peptide synthesis known in the art, including solid phase synthesis or solution phase synthesis. Alternatively, the tripartite glycosyltransferase fusion proteins can be generated using recombinant expression systems and purified using any method readily known in the art, including ion exchange chromatography, hydrophobic interaction chromatography, affinity chromatography, gel filtration, and reverse phase chromatography.

[0108] Nucleotide sequences encoding the tripartite glycosyltransferase fusion proteins may be modified such that the nucleotide sequence reflects the codon preference for a particular host cell. For example, when yeast host cells are utilized, the nucleotide sequences encoding the chimeric proteins can be modified for yeast codon preference (*see, e.g.*, Bennetzen and Hall, "Codon Selection in Yeast," *J. Biol. Chem*. 257(6):3026-3031 (1982), which is hereby incorporated by reference in its entirety). Likewise, when bacterial host cells are utilized, e.g., *E. coli* cells, the nucleotide sequences encoding the chimeric biological pathway proteins can be modified for *E. coli* codon preference (*see e.g.*, Gouy and Gautier, "Codon Usage in Bacteria: Correlation With Gene Expressivity," *Nucleic Acids Res*. 10(22):7055-7074 (1982); Eyre-Walker et al., "Synonymous Codon Bias is Related to Gene Length in *Escherichia coli*: Selection for Translational Accuracy?," *Mol. Biol. Evol*. 13(6):864-872 (1996) and Nakamura et al., "Codon Usage Tabulated From International DNA Sequence Databases: Status for the year 2000,"

Nucleic Acids Res. 28(1):292 (2000), which are hereby incorporated by reference in their entirety).

[0109] A variety of genetic signals and processing events that control many levels of gene expression (*e.g.*, DNA transcription and messenger RNA ("mRNA") translation) can be incorporated into the nucleic acid construct encoding the chimeric proteins to maximize protein production. For the purpose of expressing a cloned nucleic acid sequence encoding the desired tripartite glycosyltransferase fusion protein, it is advantageous to use strong promoters to obtain a high level of transcription. Depending upon the host system utilized, any one of a number of suitable promoters may be used. For instance, when cloning in *E. coli*, its bacteriophages, or plasmids, promoters such as the T7 phage promoter, *lac* promoter, *trp* promoter, *rec*A promoter, ribosomal RNA promoter, the P_R and P_L promoters of coliphage lambda and others, including but not limited, to *lac*UV5, *omp*F, *bla*, *lpp*, and the like, may be used to direct high levels of transcription of adjacent DNA segments. Additionally, a hybrid *trp-lac*UV5 (*tac*) promoter or other *E. coli* promoters produced by recombinant DNA or other synthetic DNA techniques may be used to provide for transcription of the inserted chimeric genetic construct. Common promoters suitable for directing expression in mammalian cells include, without limitation, SV40, MMTV, metallothionein-1, adenovirus Ela, CMV, immediate early, immunoglobulin heavy chain promoter and enhancer, and RSV-LTR. Common promoters suitable for directing expression in a yeast cell include constitutive promoters such as an ADH1 promoter, a PGK1 promoter, an ENO promoter, a PYK1 promoter and the like; or a regulatable promoter such as a GAL1 promoter, a GAL10 promoter, an ADH2 promoter, a PHO5 promoter, a CUP1 promoter, a GAL7 promoter, a MET25 promoter, a MET3 promoter, a CYC1 promoter, a HIS3 promoter, a PGK promoter, a GAPDH promoter, an ADC1 promoter, a TRP1 promoter, a URA3 promoter, a LEU2 promoter, an ENO promoter, a TP1 promoter, and a AOX1 promoter.

[0110] There are other specific initiation signals required for efficient gene transcription and translation in eukaryotic and prokaryotic cells that can be included in the nucleic acid construct to maximize chimeric protein production. Depending on the vector system and host utilized, any number of suitable transcription and/or translation elements, including constitutive, inducible, and repressible promoters, as well as minimal 5' promoter elements, enhancers, or leader sequences may be used. For a review on maximizing gene expression see Roberts and Lauer, "Maximizing Gene Expression On a Plasmid Using Recombination *In Vitro*," *Methods in Enzymology* 68:473–82 (1979), which is hereby incorporated by reference in its entirety.

[0111] A nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein of the present disclosure, a promoter molecule of choice, including, without limitation, enhancers, and leader sequences; a suitable 3' regulatory region to allow transcription in the host, and any

additional desired components, such as reporter or marker genes, are cloned into a vector of choice using standard cloning procedures in the art, such as described in Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL (Cold Springs Harbor 1989); Ausubel, SHORT PROTOCOLS IN MOLECULAR BIOLOGY (Wiley 1999), and U.S. Patent No. 4,237,224 to Cohen and Boyer, which are hereby incorporated by reference in their entirety. Suitable expression vectors include those described *supra*. Two or more nucleic acid constructs encoding two or more tripartite glycosyltransferase fusion proteins can be housed in the same or different expression vectors. In one embodiment of the present disclosure, two or more nucleic acid molecules encoding two or more tripartite glycosyltransferase fusion proteins are present in the same nucleic acid vector.

[0112] In some embodiments, the recovered tripartite glycosyltransferase fusion protein is conformationally correct.

[0113] Another aspect of the present disclosure relates to a tripartite glycosyltransferase fusion protein produced by the methods of recombinantly producing a tripartite glycosyltransferase fusion protein according to the present disclosure.

[0114] As will be apparent to one of skill in the art, the present disclosure allows for a broad range of *in vivo* or *in vitro* glycan remodeling. The constructs of the present disclosure allow for solubilized tripartite glycosyltransferase fusion proteins for use in methods of *in vivo* or *in vitro* glycan remodeling. Accordingly, another aspect of the present disclosure relates to a method of cell-free glycan remodeling. This method involves providing a glycan primer; providing one or more tripartite glycosyltransferase fusion protein(s) according to the present disclosure; and incubating the glycan primer with the one or more tripartite glycosyltransferase fusion protein(s) under conditions effective to transfer a glycosyl group to the glycan primer to produce a modified glycan structure.

[0115] The glycan primer may be a monosaccharide or an oligosaccharide. For example, the glycan primer may comprise $Man_3GlcNAc_2$ or $Man_5GlcNAc_2$.

[0116] In some embodiments, the glycan primer is attached to an amino acid residue such as an asparagine residue. In some embodiments, the glycan primer is attached to a protein. Accordingly, the glycan primer may be attached to a glycoprotein. The glycoprotein may comprise an *N*-glycosidic linkage. For example, the glycoprotein may comprises an *N*-acetylglucosamine (GlcNAc) linkage to asparagine.

[0117] The glycoprotein may be selected from the group consisting of an antibody or a hormone.

[0118] In some embodiments, the glycoprotein comprises an *O*-glycosidic linkage.

[0119] Suitably tripartite glycosyltransferase fusion proteins are described in detail *supra*. In some embodiments, the glycosyltransferase fusion protein is selected from the group consisting of Sx-Δ29*Hs*GnTI, Sx-Δ29*Hs*GnTII, Sx-Δ30*Hs*FucT8, Sx-Δ44*Hs*β4GalT1, Sx-Δ26*Hs*ST6Gal1, and combinations thereof.

[0120] In some embodiments, when the incubating step is carried out with a plurality of different tripartite glycosyltransferase fusion proteins, at least some of the different tripartite glycosyltransferase proteins being used sequentially during said incubating. In accordance with such embodiments, the incubating step produces a modified glycan primer. In some embodiments, the method may further involve incubating a modified glycan primer with one or more glycosyl hydrolases. In accordance with such embodiments, the one or more hydrolases may be used sequentially during said further incubating.

[0121] In some embodiments, when the incubating step is carried out with a plurality of different tripartite glycosyltransferase fusion proteins, at least some of the different tripartite glycosyltransferase proteins being used simultaneously during said incubating.

[0122] The above disclosure generally describes the present disclosure. A more specific description is provided below in the following examples. The examples are described solely for the purpose of illustration and are not intended to limit the scope of the present disclosure. Changes in form and substitution of equivalents are contemplated as circumstances suggest or render expedient. Although specific terms have been employed herein, such terms are intended in a descriptive sense and not for purposes of limitation.

EXAMPLES

Example 1 – Materials and Methods

Strains and Cell Lines

[0123] The bacterial, yeast, and mammalian cells used in Examples 1–9 are listed in FIGS. 1A–1B. *E. coli* strain DH5α was used for all molecular cloning and plasmid storage. *E. coli* strain BL21(DE3) and its derivative SHuffle T7 Express *lysY* (New England Biolabs) were used for all protein expression and purification. Luria-Bertani medium (LB) was used to culture *E. coli* in all experiments and was supplemented with appropriate antibiotics for plasmid maintenance. The final concentration for each antibiotic used was: 50 μg/mL kanamycin, 20 μg/mL chloramphenicol, and 100 μg/mL ampicillin. Yeast strain SBY49 was grown in complex yeast extract peptone dextrose (YPD) medium or yeast nitrogen base (YNB) medium without amino acids supplemented with uracil dropout amino acids (-URA media) for plasmid maintenance. HEK293T cells were obtained from ATCC (CRL-3216) and cultured in DMEM

supplemented with 10% FetalClone (VWR), 4.5 g/L glucose and L-glutamine, and 1% (w/v) penicillin-streptomycin-amphotericin B (Thermo Fisher Scientific). FreeStyle™ 293-F cells (HEK293F) were obtained from Thermo Fisher Scientific (Cat # R79007). Expi293-F™ GnTI$^-$ cells (HEK293F GnTI$^-$) were obtained from Thermo Fisher Scientific (Cat # A39240) and were cultured in Expi293™ Expression Medium supplemented with 1% (w/v) penicillin-streptomycin-amphotericin B (Thermo Fisher Scientific). All cells were maintained in a 37°C incubator with 5% CO_2 and 90% relative humidity. Authentication of each cell line used in this study included morphology analysis, PCR assays with species-specific primers, and STR profiling, the latter of which was performed using ATCC's human cell STR profiling service.

Cell Growth Analysis

[0124] To facilitate high-throughput cell growth measurements, three individual colonies corresponding to each construct were seeded into 96-deep well plates (Eppendorf) where each well contained 100 µL LB media. Culture plates were then sealed using plate sealer and placed in an incubator shaker at 37°C for 16 hours. Then, 5 µL of the overnight culture was subcultured into fresh 100 µL LB media and incubated for 8 hours, after which IPTG was supplemented to a final concentration of 0.1 mM. Protein expression proceeded at 16°C for 18 hours. To measure OD_{600}, 10 µL of each sample was mixed with 90 µL DI water in a Costar 96-well assay plate (Corning) and OD_{600} of all samples was measured in an Infinite M1000Pro spectrophotometer (Tecan).

Plasmid Construction.

[0125] All plasmids used in this study are listed in FIGS. 1A–1B. The collection of prokaryotic and eukaryotic glycoenzymes was selected from the CAZy database (Lombard et al., "The Carbohydrate-active Enzymes Database (CAZy) in 2013," *Nucleic Acids Res* 42:D490-5 (2014), which is hereby incorporated by reference in its entirety). Amino acid sequences were all extracted from the UniProt database (UniProt, C., "UniProt: A Worldwide Hub of Protein Knowledge," *Nucleic Acids Res* 47:D506-D515 (2019), which is hereby incorporated by reference in its entirety). Each GT coding region was examined for membrane domains using the UniProt database to determine the TMD topology. GTs with internal or multi-pass TMDs were expressed as full-length proteins. For type II transmembrane proteins, N- and C-terminal TMD segments were truncated while stem regions were generally retained. For other classes, N-terminal signal peptides and C-terminal ER retention signals were generally removed. Amino acid sequences of full-length and truncated variants of all GTs in this study are provided in FIGS. 2C–2D. All GT genes were codon-optimized for expression in *E. coli* using GeneArt software (Thermo Fisher Scientific). These genes were then synthesized and ligated into the previously described SIMPLEx plasmid (Mizrachi et al., "Making Water-soluble Integral

Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat Commun.* 6:6826 (2015), which is hereby incorporated by reference in its entirety) to generate plasmids encoding SIMPLEx-reformatted GTs having the form pET28a(+)-MBP-(*Nde*I)-GT-(*Eco*RI)-ApoAI*-6xHis. PCR was used to amplify each GT gene with flanking *Nco*I and *Not*I restriction sites, and then ligated into pET28a(+) vector to create plasmids for expression of unfused GT constructs having the form pET28a(+)-(*Nco*I)-GT-(*Not*I)-6xHis. All PCR reactions were performed using 0.1 μM gene-specific primers, 50 ng DNA template, and Phusion® High-Fidelity DNA Polymerase (New England Biolabs). Ligation products were used to chemically transform *E. coli* DH5α, and the transformation cultures were plated on LB-agar plates containing kanamycin. Clones were selected and screened by colony PCR using 2x-OneTaq Quickload master mix (New England Biolabs). Successful clones were confirmed by Sanger sequencing at the Cornell Biotechnology Resource Center. Due to incompatibility of DNA restriction sites, plasmids used for expression in yeast and mammalian cells were constructed using Gibson assembly. Briefly, standard PCR was used to amplify target genes containing 20-25 bp homologous regions with vector at both ends. 50 ng of linearized vector and 150 ng of amplified insert were then combined in a Gibson Assembly Master Mix (New England Biolabs) and incubated for 1 hour. Assembly reactions were then used to transform *E. coli* DH5α, after which clones were screened and confirmed according to a similar procedure as described above.

Small-Scale Expression and Subcellular Fractionation

[0126] Plasmids encoding Sx-GT and unfused GT constructs were used to transform either *E. coli* strain BL21(DE3) for GTs containing no disulfide bonds or SHuffle T7 Express *lysY* for GTs contain predicted or confirmed to contain disulfide bonds. Small 5-mL LB cultures of *E. coli* harboring either a Sx-GT or GT plasmid were grown to an optical density at 600 nm (OD_{600}) of approximately 0.6-0.8 and then induced with IPTG to a final concentration of 0.1 mM. Protein expression proceeded for 18 hours at 16°C, after which culture volumes equivalent to OD_{600} of 2.0 were harvested. Media was removed by centrifugation and the resulting cell pellet was resuspended in 1 mL phosphate buffer saline (PBS). Cells were lysed using a Q125 Sonicator (Qsonica) with a 3.175-mm diameter probe at a frequency of 20 kHz and 40% amplitude. Lysate was first centrifuged at 15,000 x g for 30 minutes at 4 °C. Supernatant was collected and centrifuged at 100,000 x g for 1 hour at 4°C. The supernatant from this ultracentrifugation step was collected as the soluble fraction. Pellet was then resuspended in 1 mL PBS containing 1% (v/v) Triton X-100. The suspension was incubated for 1 hour at 4°C to allow partitioning of membrane proteins into Triton X-containing buffer. Following ultracentrifugation at 100,000 x g for 1 hour at 4°C, supernatant was collected as the detergent-solubilized fraction, while the pellet was taken as the insoluble fraction.

Protein Purification and Yield Determination

[0127] A single colony of *E. coli* harboring plasmid DNA encoding a specific glycoenyzme was selected from a transformation plate and grown overnight in LB media at 37°C. The next day, cells were subcultured 5% into 1 L of fresh LB media. Cells were grown at 37°C until OD_{600} reached approximately 0.6-0.8, after which IPTG was supplemented into culture at 0.1 mM final concentration. Protein expression proceeded at 16°C for 18 hours. Unless otherwise noted, all purification procedures were performed at 4°C. Cells were harvested, resuspended in PBS supplemented with 10% (v/v) glycerol, and lysed by passing the cell suspension through an Emulsiflex C5 homogenizer (Avestin) twice at 15,000 psi maximum pressure. Supernatant was collected following centrifugation at 15,000 x g for 30 minutes and then incubated with 300 µL pre-washed HisPur™ Ni-NTA resin (Thermo Fisher Scientific) at 4°C for 1 hour. The suspension was loaded onto an Econo-Pac® gravity flow chromatography column (Bio-Rad) and resin was washed with 6 column volumes HisPur wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0). The target protein was eluted with HisPur elusion buffer (50 mM NaH_2PO_4, 300 mM NaCl, 300 mM imidazole, pH 8.0). Sample was then buffer exchanged into PBS using Zeba spin desalting columns, 7K MWCO (Thermo Fisher Scientific). Protein concentration was determined using Bradford assay (Bio-Rad). Purified protein fractions were subjected to standard Coomassie-blue staining of SDS-PAGE gels and purity of each was determined by densitometry analysis using BioRad Image Lab software (version 6.1.0 build 7), whereby the intensity of the band corresponding to the full-length Sx-GT construct was normalized to the intensity of all bands that appeared in the same lane of the gel. In general, purity of isolated Sx-GTs was approximately 50-80% following just a single-step Ni-NTA purification. Final yield values were tabulated based on both total protein concentration and purity, and were representative of three biological replicates starting from freshly transformed cells.

[0128] All other purification was performed as described above but with amylose resin (NEB) instead of Ni-NTA resin. Clarified lysate was incubated with 300 µL pre-washed amylose resin with rotation for 2 hours at 4°C. The suspension was loaded onto an Econo-Pac® gravity column (Bio-Rad) and resin was washed with 6 column volumes of amylose column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). The target protein was eluted with amylose elusion buffer (10 mM maltose in column buffer). Protein purity and concentration were determined by Coomassie staining and Bradford assay (both from Bio-Rad), respectively. Proteins were kept at 4°C for 2 weeks. For longer term storage at -80°C, protein solution was supplemented with 10% (v/v) glycerol and 0.02% (w/v) sodium azide as a cryogenic agent and bacteriostat, respectively.

[0129] For human MAN2A1 expression and purification, an expression construct encoding the truncated catalytic domain of human MAN2A1 (UniProt Q16706, residues 27-1144) was used (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018), which is hereby incorporated by reference in its entirety). This recombinant human MAN2A1 construct was expressed by transient transfection of suspension culture HEK293F cells, with soluble recombinant human MAN2A1 expressed as a soluble secreted product that was purified as described (Kadirvelraj et al., "Human N-acetylglucosaminyltransferase II Substrate Recognition Uses a Modular Architecture That Includes a Convergent Exosite," *Proc. Natl. Acad. Sci. USA* 115:4637-4642 (2018), which is hereby incorporated by reference in its entirety). Briefly, the conditioned culture medium was loaded on a Ni^{2+}-NTA Superflow column (Qiagen) equilibrated with 20 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 7.4, washed with column buffer, and eluted successively with column buffers containing stepwise increasing imidazole concentrations (40-300 mM). The eluted fusion protein was pooled, concentrated, and concurrently mixed with recombinant TEV protease and EndoF1 at ratios of 1:10 relative to the GFP-MGAT2 for each enzyme, respectively, and incubated at 4°C for 36 hours to cleave the tag and glycans. Dilution to lower the imidazole concentration was followed by passing the sample through a Ni^{2+}-NTA column to remove the fusion tag and His-tagged TEV protease and EndoF1. The protein was further purified on a Superdex 75 gel filtration column (GE Healthcare) and peak fractions of MGAT2 were collected. The protein buffer was exchanged by ultrafiltration and adjusted to 1 mg/mL with buffer containing 20 mM HEPES, 100 mM NaCl, pH 7.0, 0.05% sodium azide, and 10% glycerol and stored at -80°C until use.

[0130] For antibody expression and purification, glycoengineered HEK293F GnTI$^-$ cells were used as follows. After at least three passages, cells were washed and resuspended at 3 million cells per mL concentration. Plasmid pVITRO1-Trastuzumab-IgG1/κ (Addgene #61883) was prepared from *E. coli* culture and the purified plasmid was flowed through an endotoxin removal column to remove contaminating endotoxin. Plasmid DNA-cationic lipid complex was then generated using Lipofectamine™ Transfection Reagent (Thermo Fisher Scientific) and was slowly added into the culture media with gentle mixing. The amount of DNA, cationic-lipid reagents, and cells were scaled linearly according to the manufacturer's protocol. Cells were maintained in a 37°C incubator shaker for 24 hours prior to being supplemented with Expression Enhancer Reagents (Thermo Fisher Scientific). Cell cultures were maintained at the same condition for another 5 days to allow antibody accumulation in the culture supernatant. Cells were then removed by centrifugation at 1,000 x g for 5 minutes and supernatant was filtered through a 0.2-micron bottle-top filter. Supernatant was then mixed with 1x PBS at a 1:1 (v/v)

ratio. This solution was flowed through MabSelect SuRe resin (Sigma-Aldrich) twice to allow antibody capture on protein A/G beads. Following extensive washing with 1x PBS, captured antibodies were eluded using glycine solution (pH 2.0) directly into neutralizing buffer (Tris-HCl pH 8.5). The antibody product was then buffer exchanged into 1x PBS supplemented with 0.01% sodium azide. Antibody was stored at 4°C and was stable at the described conditions for at least a month.

Immunoblot Analysis

[0131] Prior to electrophoretic separation, samples were combined with NuPAGE™ 4X LDS Sample Buffer (Invitrogen) supplemented with 2.5% β-mercaptoethanol and then boiling at 100°C for 10 minutes. Samples equivalent to OD_{600} of 0.375 for small-scale expression or 15 µL of CFPS reaction were loaded into each well of Bolt™ 8% Bis-Tris Plus Gels (Thermo Fisher Scientific). Following electrophoretic separation and transfer to Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 µm), blots were washed with TBS buffer (80 g/L NaCl, 20 g/L KCl, and 30 g/L Tris-base) followed by a 1-hour incubation in blocking solution (50 g/L non-fat milk in TBS supplemented with 0.05% (v/v%) Tween-20; TBST). Blots were then washed 4 times with TBST in 10-minute intervals and probed with primary antibodies including rabbit polyclonal antibody to 6xHis epitope tag (Thermo Fisher Scientific; Cat # PA1-983B; 1:5,000 dilution), mouse monoclonal anti-GAPDH clone 6C5 (Calbiochem; Cat # CB1001; 1:10,000 dilution), rabbit polyclonal anti-GroEL (Sigma-Aldrich; Cat # G6532; 1:20,000 dilution), and rabbit anti-alpha tubulin clone EPR13799 (Abcam; Cat # ab184970; 1:10,000 dilution). Secondary antibodies were used as needed and these include goat anti-rabbit IgG H&L (HRP) (Abcam; Cat # ab6721; 1:5,000 dilution), rabbit anti-mouse IgG H&L (HRP) (Abcam; Cat # ab6728; 1:5,000 dilution), and ExtrAvidin®–Peroxidase (Sigma-Aldrich; Cat # E2886; 1:4,000 dilution). Blots were then washed as above. Imaging of blots was performed using a ChemiDoc™ XRS+ System following a brief incubation with Western ECL substrate (Bio-Rad).

Sialyltransferase Activity Assay

[0132] Kinetic analysis of sialytransferases was performed using a commercial sialytransferase activity kit (R&D Systems, Cat # EA002) according to manufacturer's protocols. Briefly, assays used 2 µg/mL of purified Sx-Δ26*Hs*ST6Gal1 or commercial human ST6Gal1 (amino acids 44-406) (R&D Systems; Cat # 7620-GT-010), 1.0 mg/mL of asialofetuin (Sigma-Aldrich; Cat # A4781-50MG) as acceptor substrate, and 0.02-0.8 mM of CMP-Neu5Ac as donor substrate. All reactions were incubated for 15 minutes at 37°C. Values for V_{max} and K_m were determined using Prism 9 for MacOS version 9.2.0. A conversion factor used for calculating the amount of enzymatically released inorganic phosphate from CMP-Neu5Ac was determined to be 3,833.5 pmol/OD_{620} using the phosphate standards included in the kit and was used for all data

analysis. Specific activity was calculated using 0.1 mM of CMP-Neu5Ac, 1.0 mg/mL of asialofetuin, and 0.04-0.23 μg of Sx-Δ26*Hs*ST6Gal1. A linear plot of absorbance (OD_{620}) versus amount of Sx-Δ26*Hs*ST6Gal1 was generated (FIG. 9B). The slope of this plot was transformed using the conversion factor and divided by the reaction time to calculate the specific activity in units of pmol/min/μg.

Bioorthogonal Click Chemistry-Based Chemoenzymatic Remodeling

[0133] Strain-promoted alkyne-azide cycloaddition was used to assess the ability of Sx-GTs to chemoenzymatically remodel glycoprotein substrates. In a typical reaction, a 1.5-mL microcentrifuge tube was charged with 20 μL of reaction mixture consisting of 1 μg purified Sx-GT or 50 μg cell lysate, 3 μg purified acceptor glycoprotein substrate, and 10 mM nucleotide-activated monosaccharide donor modified with an azide functional group. Depending on the GT reactions, the nucleotide-activated monosaccharide donors included UDP-GlcNAz, UDP-GalNAz, GDP-AzFuc, and CMP-AzNeu5Ac (all from R&D Systems). Following an incubation in a 37°C water bath for 1 hour, reaction mixtures were supplemented with 2-iodoacetamide (Sigma-Aldrich) at 100 mM final concentration and incubated in the dark at room temperature for 1 hour. Then, 100 mM final concentration of carboxyrhodamine 110 or biotin$(PEG)_4$ conjugated dibenzocyclooctyne-amines (Click Chemistry Tools) in *N*,*N*-dimethylformamide (DMF) was supplemented into the reaction mixture. Strain-promoted click reactions were carried out at 37°C for 2 hours. Samples were then combined with 4X LDS Sample Buffer (Invitrogen) supplemented with 2.5% β-mercaptoethanol and heated at 65 °C for 5 minutes. Following SDS-PAGE analysis, in-gel fluorescence from carboxyrhodamine110-linked glycans on glycoproteins was measured using a ChemiDoc™ MP Imaging System (Bio-Rad) with 501/523 nm $\lambda_{ex}/\lambda_{em}$. Biotin-linked glycans on glycoproteins were analyzed following immunoblot analysis using horseradish peroxidase conjugated streptavidin (Sigma-Aldrich) in a similar manner as described above for immunoblot analysis.

Cell-Free Protein Synthesis

[0134] *E. coli* lysate was prepared according to an established protocol (Kwon and Jewett, "High-throughput Preparation Methods of Crude Extract for Robust Cell-free Protein Synthesis," *Scientific Reports* 5:8663 (2015), which is hereby incorporated by reference in its entirety). Briefly, *E. coli* strain BL21(DE3) was cultured in 2xYTPG media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L potassium phosphate monobasic, 3 g/L potassium phosphate dibasic and 18 g/L glucose) at 37°C with 0.5 mM IPTG until OD_{600} reached approximately 1.0. Cells were then harvested and washed twice with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate and 60 mM potassium acetate). The resulting pellet was stored at -80°C until used. To prepare crude extract, pellets were thawed on ice and

resuspended with S30 buffer (1 mL per gram cell pellet). Cells were lysed using a Q125 Sonicator with a 3.175-mm diameter probe at a frequency of 20 kHz and 40% amplitude until the total energy input reached 1500 J. Lysate was then centrifuged twice at 30,000 x g at 4°C for 30 minutes. Supernatant was then collected, aliquoted, and stored at -80°C until used. Cell-free synthesis of Sx-GT and unfused GT constructs was performed using the modified PANOx-SP system (Jewett and Swartz, "Mimicking the Escherichia coli Cytoplasmic Environment Activates Long-lived and Efficient Cell-free Protein Synthesis," *Biotechnology and Bioengineering* 86:19-26 (2004), which is hereby incorporated by reference in its entirety). Specifically, S30 lysate was pre-conditioned with 750 μM iodoacetamide in the dark at room temperature for 30 minutes and then lysate was supplemented with 200 mM glutathione at a 3:1 ratio between oxidized and reduced forms. Then, 200 ng plasmid DNA was introduced into cell-free protein synthesis reaction containing 30% (v/v) S30 lysate and the following: 12 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 1.2 mM adenosine triphosphate (ATP), 0.85 mM guanosine triphosphate (GTP), 0.85 mM uridine triphosphate (UTP), 0.85 mM cytidine triphosphate (CTP), 0.034 mg/mL folinic acid, 0.171 mg/mL *E. coli* tRNA (Roche), 2 mM each of 20 amino acids, 30 mM phosphoenolpyruvate (PEP, Roche), 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 4 mM oxalic acid, 1 mM putrescine, 1.5 mM spermidine, and 57 mM HEPES. The synthesis reaction was carried out at 30°C for 6 hours, after which the sample was centrifuged at 15,000 x g for 30 minutes at 4°C. Supernatant was collected and stored at -20°C until further analysis.

Yeast and Mammalian Cell Expression

[0135] Yeast cells were transformed with plasmid pYS338 encoding Δ26*Hs*ST6Gal1 using the LiAc/single stranded carrier DNA/PEG method (Gietz and Schiestl, "High-efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method," *Nat. Protoc.* 2:31-4 (2007), which is hereby incorporated by reference in its entirety). For yeast expression, SBY49 cells were grown in -URA media at 30°C until OD_{600} reached approximately 0.6-0.8, after which protein expression was induced with galactose to a final concentration of 2% (w/v). Protein expression was performed for 22 hours at 30°C. Yeast cells were lysed by vortexing the cell suspension with glass beads in PBS containing zymolyase enzyme. For mammalian cell expression, 2.0 mL of HEK293T cells at approximately 80% confluency in a 6-well plate were transfected with 2 μg plasmid DNA using jetPRIME® transfection reagent (Polyplus Transfection). After transfection, cells were maintained in an incubator at 37°C with 5% CO_2 and 90% relative humidity for 36 hours, after which they were harvested. HEK293T cells were lysed by tip sonication. Subcellular fractionation analysis for yeast and HEK293T cells was performed similarly as described above. All samples were stored at -20°C until further analysis.

Cell-Free Bioenzymatic Glycan Synthesis

[0136] All glycans and nucleotide-activated sugar substrate solutions were prepared in sterile DI water and stored at –20°C. Glycan 1 was prepared as described (Hamilton et al., "A Library of Chemically Defined Human N-glycans Synthesized From Microbial Oligosaccharide Precursors," *Sci. Rep.* 7:15907 (2017), which is hereby incorporated by reference in its entirety). Briefly, dried cell pellets from a 250-mL culture of *E. coli* Origami2(DE3) *gmd*::kan Δ*waaL* cells carrying plasmid pConYCGmCB (Glasscock et al., "A Flow Cytometric Approach to Engineering Escherichia coli for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488-495 (2018), which is hereby incorporated by reference in its entirety) were resuspended in 2:1 chloroform: methanol, sonicated, and the remaining solids collected by centrifugation. This pellet was sonicated in water and collected by centrifugation. The resulting pellet was sonicated in 10:10:3 chloroform: methanol:water to isolate the lipid-linked oligosaccharides (LLOs) from the inner membrane. The LLOs were purified using acetate-converted DEAE anion exchange chromatography as they bind to the anion exchange resin via the phosphates that link the lipid and glycan. The resulting compound was dried and treated by mild acid hydrolysis to release glycans from the lipids. The released glycans were then separated from the lipid by a 1:1 butanol:water extraction, wherein the water layer contains the glycans. The glycans were then further purified with a graphitized carbon column using a 0–50% water: acetonitrile gradient. Following this procedure, approximately 750 μg of glycan 1 that was well resolved from contaminant peaks was reproducibly obtained (FIG. 6B). To synthesize glycan 2, 5 μg of glycan 1 was incubated with 20 μg/mL Sx- Δ29*Hs*GnTI and 10 mM UDP-GlcNAc (Sigma-Aldrich) in GnT buffer (20 mM HEPES, 50 mM NaCl, 10 mM $MnCl_2$, pH 7.2) at 37°C for 16 hours. To synthesize glycan 3, glycan 2 was incubated with 80 μg/mL Sx-Δ29*Hs*GnTII and 20 mM UDP-GlcNAc in GnT buffer at 37°C for 36 hours. Glycan 3 was then incubated with 20 μg/mL Sx- Δ44*Hs*β4GalT1 and 10 mM UDP-Gal (Sigma-Aldrich) in GalT buffer (20 mM HEPES, 150 mM NaCl, 10 mM $MnCl_2$, pH 7.5) at 37°C for 16 hours to produce glycan 4. Sialic acid terminal glycans 5 and 6 were synthesized by incubating glycan 4 with 20 μg/mL Sx-Δ26*Hs*ST6Gal1 and 20 mM CMP-Neu5Ac (Sigma-Aldrich) in SiaT buffer (50 mM sodium phosphate, 150 mM NaCl, 10 mM $MgCl_2$, pH 8.0) at 37°C for 16 hours. Glycan 7 was synthesized by incubating glycan 4 with 20 μg/mL Sx- Δ30*Hs*FucT8 and 10 mM GDP-fucose (Sigma-Aldrich) in FucT buffer (100 mM MES, 10 mM $MgCl_2$, pH 7.0) at 37°C for 16 hours. Glycans 8, 9, and 10 were synthesized sequentially from glycan 7 using Sx-Δ44*Hs*β4GalT1 and Sx-Δ26*Hs*ST6Gal1 as described above for glycans 4, 5, and 6. Following reaction clean-up and glycan purification, reaction progress was monitored by MALDI-TOF MS. Briefly, 1 μL (~25 ng) of partially purified glycan was co-crystalized with 1 μl matrix consists of 2,5-

dihydroxybenzoic acid (10 mg/ml) in 70% (v/v) acetonitrile. The sample was analyzed in positive mode MALDI-TOF (SCIEX TOF/TOF 5800) operated in linear mode with data acquisition at 2000 shots/spot in the 5-100-kDa mass range. Because sialic acid is subject to MS-induced in-source and metastable decay, successful biosynthesis of glycans 5, 6, 9, and 10 was verified by nano LC-MS/MS analysis as described below.

Cell-Free Bioenzymatic Glycan Remodeling on Glycoproteins

[0137] Unless noted otherwise, all glycoprotein remodeling reactions were performed at 37°C for 1 hour prior to bioorthogonal labeling reaction as described above. The sialytransferase activity of Sx-*Cj*CstII was assessed using human A1AT as glycoprotein acceptor substrate. A total of 3 μg of recombinant A1AT (R&D Systems) was treated with 20 U/μL α2-3,6,8,9 neuraminidase A (NEB) in a 10-μL reaction at 37°C for 2 hours to remove terminal sialic acid residues on A1AT glycans. Reaction mixtures were then heated at 85°C for 15 minutes to inactivate neuraminidase A. Neuraminidase A-treated A1AT was then incubated with Sx-*Cj*CstII and CMP-AzNec5Ac in SiaT buffer in a 37°C water bath for 1 hour. Sialyltransferase activity of Sx-Δ34*Hs*ST3Gal1 was evaluated in a similar manner but neuraminidase-treated bovine submaxillary glands mucin (Sigma-Aldrich) was used as the glycoprotein substrate. *N*-acetylglucosaminyltransferase activity of Sx-Δ29*Hs*GnTI was assessed using MBP-GCG^{DQNAT}, a fusion between *E. coli* MBP and human glucagon (residues 1-29) followed by a C-terminal DQNAT glycosylation tag (Glasscock et al., "A Flow Cytometric Approach to Engineering Escherichia coli for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488-495 (2018), which is hereby incorporated by reference in its entirety). The MBP-GCG^{DQNAT} construct was glycosylated with $Man_3GlcNAc_2$ using glycoengineered *E. coli* as described (Glasscock et al., "A Flow Cytometric Approach to Engineering Escherichia coli for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488-495 (2018), which is hereby incorporated by reference in its entirety). Briefly, Origami2(DE3) *gmd*::kan Δ*waaL* cells carrying plasmid pConYCGmCB along with plasmid pMAF10 (Feldman et al., "Engineering N-linked Protein Glycosylation With Diverse O Antigen Lipopolysaccharide Structures in Escherichia coli," *Proc Natl Acad Sci U S A* 102:3016-21 (2005), which is hereby incorporated by reference in its entirety) and pTrc-spDsbA-MBP-GCG^{DQNAT} (Glasscock et al., "A Flow Cytometric Approach to Engineering Escherichia coli for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488-495 (2018), which is hereby incorporated by reference in its entirety) were grown in 100 mL of LB at 37°C until OD_{600} reached ≈1.5. Culture temperature was reduced to 30°C and allowed to grow overnight at 30°C. The next day, cells were induced with 0.1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to initiate synthesis of the MBP-GCG^{DQNAT} acceptor protein. Protein expression proceeded for 8 hours at 30°C. Cells were then

harvested and subjected to subcellular fractionation. This involved pelleting and washing 100 mL of IPTG-induced culture with subcellular fractionation buffer (0.2 M Tris-Ac (pH 8.2), 0.25 mM EDTA, and 0.25 M sucrose, and 160 μg/mL lysozyme). Cells were resuspended in 1.5 mL subcellular fractionation buffer and then incubated for 5 minutes on ice and spun down. After addition of 60 μL of 1M $MgSO_4$, cells were incubated for 10 minutes on ice. Cells were spun down, and the supernatant was taken as the periplasmic fraction. To isolate glycoproteins, periplasmic fractions were subjected to affinity chromatography using HisPur™ Ni-NTA resin (Thermo Fisher Scientific). Eluates were collected, solubilized in Laemmli sample buffer containing 5% β-mercaptoethanol, and resolved on SDS-polyacrylamide gels. Purified MBP-GCG^{DQNAT} was incubated with Sx-Δ29*Hs*GnTI and UDP-GlcNAz in GnT buffer in a 37°C water bath for 1 hour. Fucosyltransferase activity was evaluated by incubating A1AT or neuraminidase A-treated A1AT with Sx-Δ36*Hs*FucT7 and GDP-AzFuc in FucT buffer in a 37°C water bath for 1 hour.

Endoglycosidase Sensitivity Assay

[0138] In a sterile Eppendorf microcentrifuge tube, 1 μg of purified trastuzumab bearing $Man_5GlcNAc_2$ glycan was incubated with: (i) *Streptococcus pyogenes* Endo S2 (Genovis # A0-GL8-020) in Glycobuffer 1 (NEB # B1727SVIAL); (ii) *Elizabethkingia meningosepticum* Endo F1 (Sigma-Aldrich #324725) in GlycoBuffer 4 (NEB #B1703); (iii) *Elizabethkingia miricola* Endo F3 (NEB #P0771S) in GlycoBuffer 4; or (iv) PBS control. Reaction mixtures were incubated at 37°C for 16 hours and the product was analyzed by LC-MS using intact protein MS mode.

Cell-Free Bioenzymatic Glycan Remodeling on Trastuzumab

[0139] Glycan remodeling on full-length mAb was performed in an on-column mode. 50 μg purified trastuzumab bearing $Man_5GlcNAc_2$ glycan was first incubated with MabSelect SuRe resin (Sigma-Aldrich) for 10 minutes to allow antibody capture on protein A/G beads. This mixture was then transferred to a spin column, followed by washing twice with PBS. The bottom of the spin column was then capped with rubber cap. In a separate tube, 50 μL of a specific glycan remodeling reaction mixture was prepared. For preparing *N*-acetylglucosaminyltransferase, galactosyltransferase, fucosyltransferase, and sialyltransferase reaction mix. UDP-GlcNAz substrate was used at the same concentration as UDP-GlcNAc. Reaction using β-*N*-acetylglucosaminidase S (NEB # P0744S) was performed in Glycobuffer 1 (NEB) at 37°C for 4 hours. Reactions using human Man2A1 mannosidase were performed in 50 mM sodium acetate buffer (pH 5.5) 1 mM $ZnCl_2$ at 37°C for 16 hours. Following each reaction step, the reaction mixture was removed by centrifugation at 300 x g for 2 minutes. Resin was then washed twice with PBS using the same centrifugation setting. In general, approximately

80-90% recovery yield of IgG was observed following purification as determined by NanoDrop spectrophotometer. Subsequent reaction mixture was then added to the column and the clean-up process was repeated for each reaction step. Final IgG product was eluted using glycine solution (pH 2.0) and analyzed immediately by LC-MS.

Chromatography and Mass Spectrometry

[0140] Hydrophilic interaction liquid chromatography (HILIC) was carried out using an Exion HPLC system with built-in autosampler (SCIEX). The free glycan samples were reconstituted in buffer A (80%: 20% acetonitrile: water), filtered with 0.22 μm spin filter (Corning) and loaded onto a Kinetex HILIC column (2.6 μm, 2.6x150 mm; Phenomenex) with 80% ACN/20% water as buffer A and 50 mM NH_4FA with pH 4.4 as buffer B. LC was performed using a 7-min gradient from 80 to 0% of buffer B at a flow rate of 400 μL/min.

[0141] All LC-MS/MS analysis was carried out using an X500B QTOF (SCIEX) mass spectrometer equipped with an electrospray ion source and coupled with an Exion HPLC system. Each reconstituted sample was injected onto a Kinetex HILIC column (2.6 μm, 2.6x 150 mm; Phenomenex). The free glycans were eluted in a 9-min gradient of 80% to 0% (80% ACN/20% water) at 400 nL/min followed by a 3-minute hold at 80% (80% ACN/20% water) for re-equilibration. The instrument was operated in positive ion mode with ESI voltage set at 5.0 kV, ion source gas 1, gas 2 = 50 psi, curtain gas = 35 and CAD gas = 7, and source temperature of 350°C. Calibration was done using positive calibrant with CDS system. For free glycan analysis, the instrument was operated in MS full-scan mode from *m/z* range from 2,00-2,000 followed by multiple reaction monitoring high-resolution (MRM-HR) scan from 0-12 minutes at two different collision energies of 20 and 35 V with DP = 20 V and accumulation time of 0.25 s. MS survey scans were performed for the mass range of *m/z* 2,00-2000 with DP = 20 V, CE = 7 V and accumulation time of 0.25 s and MS/MS MRM-HR scans were at the same DP voltage and CE = 20 V and with Q1 unit resolution. All MS and MS/MS raw spectra from each sample obtained by MRM-HR scan were analyzed by SCIEX OS 1.4 data analysis system. XIC spectra were extracted from MS full-scan with each MRM transition. The glycan structure was annotated manually using GlycanMass-ExPAsy tool.

Physicochemical Data Collection and Analysis

[0142] The name, amino acid sequence, structure availability (full-length or partial), and predicted post-translational modifications (*i.e.*, disulfide bonds, glycosylation) for each GT enzyme were retrieved from the UniProt database (UniProt, C., "UniProt: A Worldwide Hub of Protein Knowledge," *Nucleic Acids Res* 47:D506-D515 (2019), which is hereby incorporated by reference in its entirety). GT family members were annotated from the CAZy database (Lombard et al., "The Carbohydrate-active Enzymes Database (CAZy) in 2013," *Nucleic Acids*

Res 42:D490-5 (2014), which is hereby incorporated by reference in its entirety). Amino acid sequences of full length, truncated, and SIMPLEx-fused GTs were compiled in FASTA format. The M_w and pI were calculated using the ExPASy Bioinformatics resource portal in average resolution setting (Wilkins et al., "Protein Identification and Analysis Tools in the ExPASy Server," *Methods Mol. Biol.* 112:531-52 (1999), which is hereby incorporated by reference in its entirety). Solubility prediction score was calculated using CamSol Intrinsic version 2.1 (Sormanni et al., "The CamSol Method of Rational Design of Protein Mutants With Enhanced Solubility," *J. Mol. Biol.* 427:478-90 (2015), which is hereby incorporated by reference in its entirety). The expression scores for all constructs were annotated based on immunoblots in FIGS. 14A–14B. Correlation between protein properties (M_w, pI, solubility prediction score, and expression score) were analyzed using R software version 3.4.2. Specifically, scatter plots between protein properties, generated with data points colored according to expression score, were used to examine any possible correlations. For the correlation between expression score and pI, datasets were analyzed as a function of expression score movement. Scatter plots comparing the pI of Sx-GT versus GT constructs were created and the data were colored according to the change in expression score. A similar approach was used to analyze the correlation between expression score and solubility prediction score. Because no statistical significance was observed for the correlation between expression score and either pI or solubility prediction score, general observations from the plots were described instead. For the correlation between expression score and M_w, data were categorized into 3 groups: $M_w < 40$ kDa, $M_w = 40$-60 kDa, and $M_w > 60$ kDa, before average expression score for each group was calculated. Both Sx-GT and GT constructs were binned using the same criteria since the added mass from the SIMPLEx fusion was constant for all constructs. Welch's *t*-test was used to analyze statistical significance for categorical datasets.

Statistical Analysis and Reproducibility

[0143] To ensure robust reproducibility of all results, experiments were performed with at least three biological replicates and at least three technical measurements. Sample sizes were not predetermined based on statistical methods but were chosen according to the standards of the field (at least three independent biological replicates for each condition), which gave sufficient statistics for the effect sizes of interest. All data were reported as average values with error bars representing standard error of the mean (SEM). Statistical significance was determined by Welch's *t*-test and *p*-values of <0.05 were considered significant. All graphs were generated using Microsoft Excel, Prism 9 for MacOS version 9.2.0, or R software version 3.4.2. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Example 2 – SIMPLEx Promotes Soluble Expression of Human ST6Gal1

[0144] Towards the goal of developing a versatile and universal approach for large-scale GT production, it was hypothesized that SIMPLEx could relieve bottlenecks that have hampered GT expression in *E. coli*. The rationale for this hypothesis was based on two observations. First, the SIMPLEx strategy has previously been shown as a promising technique for converting IMPs into water-soluble proteins with retention of biological function (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat. Commun.* 6:6826 (2015) and Mizrachi et al., "A Water-soluble DsbB Variant That Catalyzes Disulfide-bond Formation In Vivo," *Nat. Chem. Biol.* 13:1022-1028 (2017), which are hereby incorporated by reference in their entirety). Second, SIMPLEx was able to rescue soluble expression of a diverse panel of globular proteins that were previously reported to be recalcitrant to soluble expression in *E. coli* (Dyson et al.., "Production of Soluble Mammalian Proteins in Escherichia coli: Identification of Protein Features That Correlate With Successful Expression," *BMC Biotechnol.* 4:32 (2004), which is hereby incorporated by reference in its entirety) (FIG. 3A). Collectively, these results highlight the capacity of SIMPLEx to shield large amounts of protein hydrophobicity that drive misfolding and aggregation and promote soluble expression of membrane and non-membrane proteins alike.

[0145] To see if the benefits of SIMPLEx could be leveraged for GT expression, the human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1), a sialytransferase belonging to the GT29 family, was chosen as a model GT for proof-of-concept experiments. *Hs*ST6Gal1 consists of a short N-terminal cytoplasmic tail (CT), a transmembrane domain (TMD), a stem region that serves as a linker, and a large C-terminal catalytic domain that adopts a variant GT-A fold containing a seven-stranded central β-sheet flanked by α-helices (FIG. 3A) (Kuhn et al., "The Structure of Human Alpha-2,6-sialyltransferase Reveals the Binding Mode of Complex Glycans," *Acta. Crystallogr. D. Biol. Crystallogr.* 69:1826-38 (2013), which is hereby incorporated by reference in its entirety). Overexpression of *Hs*ST6Gal1 has been documented in several cancer cell types (Garnham et al., "ST6GAL1: A Key Player in Cancer," *Oncol Lett* 18:983-989 (2019), which is hereby incorporated by reference in its entirety); hence, the ability to produce preparative amounts of *Hs*ST6Gal1 could help to understand its role in cancer biology and therapy. To express this enzyme in the SIMPLEx architecture, a tripartite *Hs*ST6Gal1 chimera in which its N-terminus was genetically fused to a water-soluble "decoy" protein, namely *E. coli* maltose-binding protein lacking its N-terminal signal peptide (ΔspMBP), while its C-terminus was fused to an amphipathic "shield" protein, namely truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*) was designed, yielding ΔspMBP-*Hs*ST6Gal1-ApoAI* (hereafter Sx-*Hs*ST6Gal1) (FIG. 3A). As removal of the

transmembrane anchor segment is a common practice to improve expression and solubility of mammalian GTs (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018), which is hereby incorporated by reference in its entirety), the Sx-Δ26*Hs*ST6Gal1 variant in which 26 amino acids from the N-terminus of *Hs*ST6Gal1, comprising its CT and TMD, were genetically removed was also generated (FIG. 3A). The *Hs*ST6Gal1 enzyme contains 3 disulfide bonds in its native structure (Kuhn et al., "The Structure of Human Alpha-2,6-sialyltransferase Reveals the Binding Mode of Complex Glycans," *Acta. Crystallogr. D. Biol. Crystallogr.* 69:1826-38 (2013), which is hereby incorporated by reference in its entirety). Therefore, the commercially available *E. coli* strain named SHuffle T7 Express (Lobstein et al., "SHuffle, a Novel Escherichia coli Protein Expression Strain capable of Correctly Folding Disulfide Bonded Proteins in its Cytoplasm," *Microb. Cell Fact* 1156 (2012), which is hereby incorporated by reference in its entirety), which has been engineered with a more oxidizing cytoplasmic environment and expresses a cytoplasmic version of the disulfide bond isomerase DsbC (Lobstein et al., "SHuffle, a Novel Escherichia coli Protein Expression Strain capable of Correctly Folding Disulfide Bonded Proteins in its Cytoplasm," *Microb. Cell Fact* 1156 (2012), which is hereby incorporated by reference in its entirety), was selected as an expression host to facilitate cytoplasmic disulfide bond formation. Following expression of the two engineered chimeras in SHuffle T7 Express cells, stable products corresponding to Sx-*Hs*ST6Gal1 and Sx-Δ26*Hs*ST6Gal1 that accumulated almost exclusively in the soluble cytoplasmic fraction were observed (FIG. 3B). In stark contrast, no detectable expression of unfused *Hs*ST6Gal1 or Δ26*Hs*ST6Gal1 was seen in the soluble fraction and only minimal amounts were observed in the insoluble and detergent-solubilized fractions (FIG. 3B), in agreement with previous findings that human sialyltransferases are poorly expressed in bacteria (Skretas et al., "Expression of Active Human Sialyltransferase ST6GalNAcI in Escherichia coli," *Microb. Cell Fact* 8:50 (2009) and Ortiz-Soto and Seibel, "Expression of Functional Human Sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli," *PLoS One* 11:e0155410 (2016), which are hereby incorporated by reference in their entirety). The large expression difference seen for the Sx-Δ26*Hs*ST6Gal1 fusion relative to its unfused counterpart was also clearly observed in whole cell lysates (FIG. 8B).

[0146] To demonstrate the importance of the decoy and shield domains, chimeras lacking each of these elements were also expressed. When the decoy protein was omitted, Δ26*Hs*ST6Gal1-ApoAI* partitioned almost entirely in the insoluble fraction (FIG. 3B) and was undetectable in whole cell lysates (FIG. 8B). Omission of the shield protein resulted in accumulation of ΔspMBP-Δ26*Hs*ST6Gal1 primarily in the soluble fraction, but with significant amounts also detected in the detergent-solubilized and insoluble fractions (FIG. 3B). Consistent

with earlier studies with integral membrane protein targets (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat. Commun.* 6:6826 (2015) and Mizrachi et al., "A Water-soluble DsbB Variant That Catalyzes Disulfide-bond Formation In Vivo," *Nat. Chem. Biol.* 13:1022-1028 (2017), which are hereby incorporated by reference in their entirety), results confirmed the importance of the decoy and shield in directing folding away from the membrane and promoting water solubility. Moreover, SIMPLEx-based expression in a redox-engineered bacterial host sidestepped the need for chaperones that occur uniquely in the mammalian secretory pathway and for *N*-linked glycosylation of the GT that is not required for activity but needed for folding, stability, and solubility of the enzyme (Chen and Colley, "Minimal Structural and Glycosylation Requirements for ST6Gal I Activity and Trafficking," *Glycobiology* 10:531-83 (2000) and Meng et al., "Enzymatic Basis for N-glycan Sialylation: Structure of rat Alpha2,6-sialyltransferase (ST6GAL1) Reveals Conserved and Unique Features for Glycan Sialylation," *J. Biol. Chem.* 288:34680-98 (2013), which are hereby incorporated by reference in their entirety).

Example 3 – Soluble *Hs*ST6Gal1 in the SIMPLEx Framework Retains Biological Activity

[0147] To determine whether soluble Sx-Δ26*Hs*ST6Gal1 was biologically active, the enzyme was purified (FIG. 9A) and subjected to kinetic analysis using a commercial kit for quantifying release of nucleotide cytidine 5'-monophosphate (CMP) from the donor substrate CMP-*N*-acetylneuraminic acid (CMP-Neu5Ac). From this assay, the apparent K_M and V_{max} values for Sx-Δ26*Hs*ST6Gal1 were determined as 0.19 ± 0.03 mM and 85.5 ± 6.3 pmol/min, respectively (FIG. 3C). These parameters were in reasonable agreement with the apparent kinetic parameters measured for the commercial human ST6Gal1 (produced recombinantly using N60 mouse myeloma cells) and that were measured previously (Mbua et al., "Selective Exo-enzymatic Labeling of N-glycans on the Surface of Living Cells by Recombinant ST6Gal I," *Angew Chem. Int. Ed. Engl.* 52:13012-5 (2013), which is hereby incorporated by reference in its entirety). The specific activity of the soluble Sx-Δ26*Hs*ST6Gal1 enzyme was 516.9 pmol/min/μg (FIG. 9B**)**, which was also consistent with previously published data for human ST6Gal1 (Houeix and Cairns, "Engineering of CHO Cells for the Production of Vertebrate Recombinant Sialyltransferases," *PeerJ* 7:e5788 (2019), which is hereby incorporated by reference in its entirety).

[0148] Upon confirming that Sx-Δ26*Hs*ST6Gal1 was enzymatically active, it was next sought to demonstrate its practical utility for chemoenzymatic remodeling of *N*-linked glycans present on glycoprotein substrates. To this end, a bioorthogonal click chemistry-based assay for quantifying sialyltransferase-mediated chemoenzymatic modification was developed (FIG. 9C).

Specifically, Sx-Δ26*Hs*ST6Gal1 enzyme preparations were evaluated for their ability to transfer azido-Neu5Ac from CMP-activated glycosyl donor onto terminal Gal residues of the alpha-1 antitrypsin (A1AT) serpin protein, which was first treated with neuraminidase to remove native sialic acids. The modified A1AT was then fluorescently labeled through a strain-promoted azide-alkyne cycloaddition (SPAAC) reaction using carboxyrhodamine 110 DBCO and separated by standard SDS-PAGE. Fluorescence intensity of the labeled A1AT proteins, which corresponded to the extent of chemoenzymatic remodeling by Sx-Δ26*Hs*ST6Gal1, was then directly visualized and quantified by in-gel fluorescence analysis.

[0149] Using clarified lysate generated from *E. coli* cells expressing Sx-Δ26*Hs*ST6Gal1 as a catalyst source, a strong fluorescence from the treated A1AT was detected (FIG. 3D). In contrast, clarified lysates containing either Δ26*Hs*ST6Gal1 or Δ26*Hs*ST6Gal1-ApoAI* yielded only a weak fluorescent signal (FIG. 3D), which was consistent with the barely detectable levels of soluble expression observed for these constructs that both lacked the ΔspMBP decoy (FIG. 3B). Interestingly, while addition of the ΔspMBP moiety alone was able to promote soluble expression of ΔspMBP-Δ26*Hs*ST6Gal1 (FIG. 3B and FIG. 8B), the clarified lysate containing this construct exhibited about 50% less activity than that measured for the Sx-Δ26*Hs*ST6Gal1 enzyme (FIG. 3D). A significant portion of the soluble ΔspMBP-Δ26*Hs*ST6Gal1 protein lacking ApoAI* was misfolded aggregates, consistent with previous findings (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat Commun* 6:6826 (2015), which is hereby incorporated by reference in its entirety) and indicative of the essential nature of both ΔspMBP and ApoAI* domains for producing this GT in a highly soluble, active conformation within the *E. coli* cytoplasm. Importantly, it was also confirmed that purified Sx-Δ26*Hs*ST6Gal1 catalyzed chemoenzymatic remodeling to an extent that was indistinguishable from that of commercial human ST6Gal1 (FIGS. 9D–9E). The fact that the fused ApoAI* domain did not measurably interfere with important C-terminal catalytic regions in Δ26*Hs*ST6Gal1 including sialyl motifs III (Tyr354–Gln357), S (Pro321–Phe343), and VS (His370–Glu375) suggests that its helical bundle structure is sufficiently flexible to promote solubility while still allowing proper protein-glycan interactions that are necessary for native-like enzyme function.

Example 4 – Large-Scale Soluble Expression of Diverse GTs Using SIMPLEx Platform

[0150] Encouraged by the ability of SIMPLEx to promote soluble expression of *Hs*ST6Gal1 in *E. coli* while preserving its biological activity, whether the strategy could be extended to a larger collection of structurally diverse GTs was next investigated. To this end, a library of 98 GT genes from diverse prokaryotic and eukaryotic organisms was compiled, with

an emphasis placed on those of human origin (FIGS. 2A–2D). These genes were organized according to their species of origin and activity, and included the following: human fucosyltransferases (*Hs*FucTs), human galactosyltransferases (*Hs*Gals), human glucosyltransferases (*Hs*GlcTs), human mannosyltransferases (*Hs*ManTs), human *N*-acetylgalactosyltransferases (*Hs*GalNAcTs), human *N*-acetylglucosaminyltransferases (*Hs*GlcNAcTs), human sialyltransferases (*Hs*SiaTs), and a collection of other human, eukaryotic, and prokaryotic GTs. Using the UniProt database (UniProt, C., "UniProt: The Universal Protein Knowledgebase in 2021," *Nucleic Acids Res.* 49:D480-D489 (2021), which is hereby incorporated by reference in its entirety), these GTs were annotated based on the following characteristics: (i) single-pass transmembrane protein with C-terminus in cytoplasm (type I transmembrane protein); (ii) single-pass transmembrane protein with N-terminus in cytoplasm (type II transmembrane protein); (iii) multi-pass transmembrane protein; (iv) secretory protein with N-terminal signal peptide and C-terminal ER retention domain; and (v) cytosolic protein. It is known that N-/C-terminal TMDs as well as C-terminal ER retention domains in mammalian GTs are used as membrane anchors and are dispensable for catalytic activity (Harduin-Lepers et al., "The Human Sialyltransferase Family," *Biochimie* 83727-83737 (2001), which is hereby incorporated by reference in its entirety), as was seen above for *Hs*ST6Gal1. Because SIMPLEx-mediated solubility enhancement of *Hs*ST6Gal1 was independent of whether the TMD was present or absent (FIG. 3B), these terminal TMD anchors were generally removed from the designed constructs. N-terminal signal peptides that natively route GTs to the secretory pathway were not necessary in the context of the disclosed bacterial cytoplasmic expression system and thus were also removed. GTs containing internal single-pass or multi-pass TMDs as well as predicted cytosolic GTs were designed as full-length genes. Each designed construct in the GT library (*see* FIGS. 2C–2D for amino acid sequences) was cloned into a T7 promoter-based expression vector as both a stand-alone GT (full-length or truncated) with C-terminal 6xHis tag (hereafter GT) and a tripartite SIMPLEx fusion (hereafter Sx-GT). The expression of all Sx-GT constructs was tested in small-scale, batch-mode microbial cultures. SHuffle T7 Express cells were used to produce enzymes containing previously observed or predicted disulfide bonds while BL21(DE3) cells were used to express enzymes without such bonds (FIG. 2A). Cytoplasmic expression of the Sx-GTs was profiled by immunoblot analysis of clarified lysates derived from *E. coli* cells expressing the respective constructs. Importantly, 95 of the Sx-GT constructs showed clearly visible accumulation in the soluble cytoplasmic fractions, with most exhibiting moderate to strong expression and only a few that were faintly expressed (FIGS. 4A–4J and FIG. 10A–10J). It should be noted that this success rate was achieved under standard bacterial expression conditions (starting $OD_{600} \approx 0.6$, induction with 0.1 mM β-D-1-

thiogalactopyranoside (IPTG) at 16°C for 16–20 hours in LB medium) that were identical for each construct and did not require any of the lengthy optimization trials that are commonly associated with expression campaigns employing bacteria. Conversely, only ~45% of the unfused GT constructs could be detected in the soluble fraction under the same conditions, and in most cases the level of soluble GT expression was visibly lower compared to its Sx-GT counterpart (FIGS. 4A–4J and FIGS. 10A–10J). Subcellular fractionation analysis of select 9 candidates revealed that all SIMPLEx constructs accumulated predominantly in the soluble fraction whereas unfused versions of the enzymes partitioned mostly in the insoluble or detergent-solubilized fractions (FIGS. 11A–11I), consistent with the solubility profiles observed above for Sx-*Hs*ST6Gal1 and Sx-Δ26*Hs*ST6Gal1. Additionally, aberrant expression products such as high-molecular-weight aggregates and proteolytic degradants were prevalently detected among the GT but not the Sx-GT constructs (FIGS. 4A–4J), highlighting the intrinsic ability of the SIMPLEx strategy to enhance intracellular stability and prevent off-pathway misfolding and aggregation of target enzymes.

[0151] Another advantage of expressing GTs in the SIMPLEx framework is the potential to relieve cellular stress that arises from high-level accumulation of severely misfolded proteins (*e.g.*, inclusion bodies) or destabilization of the cytoplasmic membrane caused by high-level expression of membrane proteins, phenomena that are both well-known to negatively impact cell growth and productivity. Indeed, cultures expressing Sx-GTs were consistently observed to reach higher final cell densities than those expressing unfused GTs (FIG. 12). Likewise, titers of selected Sx-GT candidates purified from 1-L cultures were also higher on both a mass and molar basis relative to unfused GTs, with all SIMPLEx constructs accumulating in the 5-10 mg/L range (FIGS. 13A–13C). Taken together, these results demonstrate (i) large-scale GT expression using *E. coli* as the host organism and (ii) SIMPLEx as a universal strategy for high-yield, soluble expression of GTs having diverse origins, structures, and activities.

Example 5 – Correlates of Successful GT Expression in *E. coli*

[0152] It was next sought to identify the protein features that correlated with soluble protein expression by comparing physicochemical properties of the proteins including molecular weight (M_w), isoelectric point (pI), and amino acid content. This involved assigning an expression score to each of the Sx-GT and GT constructs based on their soluble expression profiles (FIG. 14A). Next, expression scores were used to bin these proteins into four groups: non-expressor (score 0); weak expressor (score 1); medium expressor (score 2); and strong expressor (score 3). According to this classification, ~95% of the Sx-GTs were identified as expressible (score ≥ 1), with more than 50% falling into the medium-to-high expressor groups.

In stark contrast, over 50% of the GT constructs were identified as non expressors, with most of the others classifying as weak expressors (FIG. 14A). A scatter plot of protein M_w, excluding added mass from the ΔspMBP and ApoAI* domains, versus solubility score calculated by Protein-Sol, a web tool for predicting protein solubility from sequence (Hebditch et al., "Protein-Sol: A Web Tool for Predicting Protein Solubility From Sequence," *Bioinformatics* 33:3098-3100 (2017), which is hereby incorporated by reference in its entirety), revealed that expressible Sx-GT constructs clustered within a 25–60 kDa range whereas expressible GT constructs were clustered in a narrower 25–40 kDa range that was skewed to smaller proteins (FIG. 14B).

[0153] This observation prompted further investigation of the relationship between soluble expression of the protein and its M_w. To this end, all GTs were categorized into one of three size groups: small (M_w < 40 kDa), medium (M_w = 40–60 kDa), and large (M_w > 60 kDa). The average expression score ($\overline{Ex}$) was then calculated for each size group within the Sx-GT or GT datasets. For GT constructs, a significant decrease in $\overline{Ex}$was observed as protein M_w increased, with no soluble expression for large proteins (FIG. 15A), consistent with the observation that bacterial translation machinery has evolved to express shorter polypeptides (Netzer and Hartl, "Recombination of Protein Domains Facilitated by Co-translational Folding in Eukaryotes," *Nature* 388:343-9 (1997), which is hereby incorporated by reference in its entirety) and that expression of larger eukaryotic proteins in bacteria frequently leads to misfolding and aggregation (Dyson et al.., "Production of Soluble Mammalian Proteins in Escherichia coli: Identification of Protein Features That Correlate With Successful Expression," *BMC Biotechnol.* 4:32 (2004) and Marston, F.A., "The Purification of Eukaryotic Polypeptides Synthesized in Escherichia coli," *Biochem. J.* 240:1-12 (1986), which are hereby incorporated by reference in their entirety). On the contrary, $\overline{Ex}$ was high for all Sx-GT constructs, with no significant difference between small- and medium-sized proteins and only a small decrease in $\overline{Ex}$ for large-sized proteins (FIG. 15A). These results suggested that the SIMPLEx framework helps to overcome the protein size barrier that typically restricts successful expression in *E. coli*. Unfortunately, attempts to identify additional parameters such as protein pI and amino acid content that correlated with expressibility did not yield conclusive results (FIG. 14B and FIG. 15B). Nonetheless, the data presented here reveal important design parameters that could guide efforts to express even more GT enzymes in the future.

Example 6 – Efficient Production of Sx-GTs Across Diverse Expression Platforms

[0154] To further expand the utility of the platform and demonstrate its universality, SIMPLEx fusions in other popular expression platforms including: (i) *E. coli*-based cell-free protein synthesis (CFPS); (ii) *Saccharomyces cerevisiae* strain SBY49; and (iii) human

embryonic kidney (HEK) 293T cells were produced. Using appropriate expression vectors for each system, significant accumulation of the Sx-Δ26*Hs*ST6Gal1 construct in the soluble fractions derived from each of these three systems was observed (FIGS. 5A–5C). In contrast, little to no soluble expression of the unfused Δ26*Hs*ST6Gal1 construct lacking the ΔspMBP and ApoAI* domains was detected in any of these systems (FIGS. 5A–5C). While Sx-Δ26*Hs*ST6Gal1 was also found in the insoluble fraction derived from the CFPS system, the amount of this construct that partitioned in the soluble fraction was significantly higher (FIG. 5A). For cell-based expression of Sx-Δ26*Hs*ST6Gal1, both yeast and human cells yielded products that accumulated almost exclusively in the soluble fractions (FIGS. 5B–5C), in line with the *E. coli* cell-based expression results observed above. Importantly, these results highlight the cross-platform compatibility of the SIMPLEx strategy and the ease with which it was adapted to these microbial, mammalian, and cell-free expression systems.

Example 7 – Cell-Free Construction of Free Human *N*-Glycans Using Sx-GTs

[0155] To date, a growing number of cell-free bio/chemoenzymatic synthesis strategies have been reported that provide access to large repertoires of pure and chemically-defined glycans, especially complex structures that are otherwise difficult to obtain by conventional chemical synthesis (Hamilton et al., "A Library of Chemically Defined Human N-glycans Synthesized From Microbial Oligosaccharide Precursors," *Sci. Rep.* 7:15907 (2017); Li and Wang, "Chemoenzymatic Methods for the Synthesis of Glycoproteins," *Chem. Rev.* 118:8359-8413 (2018); and Li et al., "Strategies for Chemoenzymatic Synthesis of Carbohydrates," *Carbohydr. Res.* 472:86-97 (2019), which are hereby incorporated by reference in their entirety). Because these approaches generally depend on the availability of glycoenzymes, many of which cannot be recombinantly expressed or purified at scale, it was sought to demonstrate the practical utility of Sx-GTs as biocatalysts for constructing customized glycan structures via a previously described bioenzymatic synthesis approach (Hamilton et al., "A Library of Chemically Defined Human N-glycans Synthesized From Microbial Oligosaccharide Precursors," *Sci. Rep.* 7:15907 (2017), which is hereby incorporated by reference in its entirety). To this end, two multi-GT enzyme pathways for *de novo* biosynthesis of a library of human hybrid- and complex-type *N*-glycans starting from a mannose$_3$-*N*-acetylglucosamine$_2$ ($Man_3GlcNAc_2$) primer were devised (FIG. 6A). To generate this primer, a glycoengineered *E. coli* strain carrying a heterologous biosynthesis pathway for producing undecaprenyl-linked $Man_3GlcNAc_2$ glycan was leveraged (Valderrama-Rincon et al., "An Engineered Eukaryotic Protein Glycosylation Pathway in Escherichia coli," *Nat. Chem. Biol.* 8:434-6 (2012), which is hereby incorporated by reference in its entirety). Following glycolipid extraction from these cells, $Man_3GlcNAc_2$ (M3; glycan 1) was

removed from undecaprenol by mild acid hydrolysis and purified to homogeneity as confirmed by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis (FIG. 6B).

[0156] Using 1 as a primer, glycan elaboration with GlcNAc was carried out by sequential treatment with purified Sx-Δ29*Hs*GnTI and Sx-Δ29*Hs*GnTII, yielding hybrid-type glycan 2 (also known as G0-GlcNAc) and complex-type glycan 3 (G0), respectively, as evidenced by MALDI-TOF MS analysis of each reaction (FIG. 6B). Further elaboration of glycan 3 with galactose was achieved using Sx-Δ44*Hs*β4GalT1 to generate glycan 4 (G2), which was subsequently elaborated using Sx-Δ26*Hs*ST6Gal1 to produce glycan 5 (G2S1) and glycan 6 (G2S2), the mono- and di-sialylated complex-type *N*-glycans, respectively (FIG. 6B). Alternatively, glycan 3 was first fucosylated using Sx-Δ30*Hs*FucT8 to generate glycan 7 (G0F), which was then further elaborated to yield glycan 8 (G2F), glycan 9 (G2S1F), and glycan 10 (G2S2F) using a similar bioenzymatic strategy (FIG. 6B).

[0157] Overall, enzymatic conversion in each of these reactions was at or near 100% except in the cases involving the Sx-Δ26*Hs*ST6Gal1-catalyzed sialyation reactions. However, because the unstable nature of sialic acid-containing glycans in MALDI-TOF MS may have confounded the sialylation analysis, nano-scale reverse phase chromatography and tandem MS (nano LC-MS/MS) analysis were performed to confirm the abundance and identity of the sialylated glycans 5, 6, 9, and 10. While both mono- and di-sialylated products were clearly detected, this analysis revealed an approximate 5:1 ratio between the G2S1 and G2S2 glycans as well as the G2S1F and G2S2F glycans (FIGS. 16A–16C and FIGS. 17A–17C). It is worth pointing out that this phenomenon has been well documented (Barb et al., "Branch-specific Sialylation of IgG-Fc Glycans by ST6Gal-I," *Biochemistry* 48:9705-7 (2009) and Tayi and Butler, "Solid-Phase Enzymatic Remodeling Produces High Yields of Single Glycoform Antibodies," *Biotechnol. J.* 13:e1700381 (2018), which are hereby incorporated by reference in their entirety) and arises from the fact that human ST6Gal1 exhibits a preference for α1–3Man-β1,2-GlcNAc-β1,4-Gal (hereafter α1–3Man branch). As a result, ST6Gal1 readily installs Neu5Ac on this branch first, with subsequent sialyation of α1–6Man-β1,2-GlcNAc-β1,4-Gal (hereafter α1–6Man branch) known to be very slow (Barb et al., "Branch-specific Sialylation of IgG-Fc Glycans by ST6Gal-I," *Biochemistry* 48:9705-7 (2009), which is hereby incorporated by reference in its entirety).

Example 8 – Cell-Free Remodeling of Protein-Linked *N*- and *O*-Glycans Using Sx-GTs

[0158] Glycoform manipulation is an emerging strategy for improving pharmacokinetics and pharmacodynamics of therapeutic glycoproteins (Wang et al., "Glycoengineering of

Antibodies for Modulating Functions," *Annu. Rev. Biochem.* 88:433-459 (2019) and Wang and Lomino, "Emerging Technologies for Making Glycan-defined Glycoproteins," *ACS Chem. Biol.* 7:110-22 (2012), which are hereby incorporated by reference in their entirety). The remodeling of protein-linked glycans can be readily achieved using one or more GTs; however, the limited availability of requisite enzymes for customizing glycan structures represents a barrier to widespread adoption. To address this technology gap, members from the disclosed library of SIMPLEx-reformatted GTs were employed to alter the glycan profiles on several biomedically-relevant glycoproteins. Remodeling reactions included: (i) Sx-*Cj*CstII-mediated α2,3-sialylation of the *N*-glycoforms on α_1-antitrypsin (A1AT), a serpin used in prophylactic treatment of the genetic disorder α_1-antitrypsin deficiency; (ii) Sx-Δ36*Hs*FucT7-mediated fucosylation of the *N*-glycoforms on A1AT; (iii) Sx-Δ34*Hs*ST3Gal1-mediated α2,3-sialylation of the *O*-glycoforms on bovine submaxillary mucin (BSM), a glycoprotein with potential uses as a biocompatible material and drug delivery vehicle; and (iv) Sx-Δ29*Hs*GnTI-catalyzed GlcNAc transfer onto $Man_3GlcNAc_2$ glycans present on a neoglycoprotein variant of human glucagon (GCG). In all cases, Sx-GTs readily remodeled their glycoprotein substrates, installing respective monosaccharides in 1-hour reactions that were monitored using bioorthogonal click chemistry-based assays with either a fluorophore or biotin reporter for glycan labeling (FIGS. 18A–18D). It should be noted that significantly decreased activity was observed for Sx-Δ36*Hs*FucT7 when the *N*-glycans on A1AT were pre-treated with neuraminidase to remove native Neu5Ac residues. This observation was in line with earlier reports (Tu et al., "Development of Fucosyltransferase and Fucosidase Inhibitors," *Chem Soc Rev* 42:4459-75 (2013), which is hereby incorporated by reference in its entirety) and highlights how subtle differences in substrate specificity can be directly investigated using GTs within the SIMPLEx framework.

Example 9 – Remodeling IgG *N*-Glycans Using Sx-GTs.

[0159] *N*-glycans present on the Fc domain of IgG antibodies play a critical role in the structure and function of these important proteins, but understanding of how discrete glycan structures affect IgG behavior remains limited due to naturally occurring microheterogeneity. Hence, strategies for generating structurally-defined *N*-glycans on IgG-Fc are expected to improve the understanding of the roles played by these structures in human immunity and to open the door to creating better medicines through glycoengineering. To this end, members from the disclosed library of Sx-GTs were leveraged to generate a homogenously glycosylated variant of trastuzumab (FIG. 7A), an anti-human epidermal growth factor receptor 2 (HER2) mAb used to treat HER2-positive breast, gastroesophageal, and gastric cancers. This involved first preparing trastuzumab using a glycoengineered cell line, Expi293F™ GnTI$^-$, that

homogeneously produces *N*-glycoproteins bearing $Man_5GlcNAc_2$ glycans (FIG. 7A, glycan 11). Using a glycosidase sensitivity assay coupled with LC-MS analysis of the intact antibody, it was confirmed that the *N*-glycans on trastuzumab derived from Expi293F™ GnTI$^-$ were indeed $Man_5GlcNAc_2$ glycans (FIGS. 19A–19D). Next, Sx-Δ29*Hs*GnTI was used to install GlcNAc on the α1,3-man branch of 11 to generate $GlcNAcMan_5GlcNAc_2$ glycan (glycan 12) directly on trastuzumab (FIG. 7B). The two terminal Man residues on the α1,6-man branch of 12 were then removed using human Golgi Man2A1 (*Hs*Man2A1), yielding trastuzumab bearing glycan 2. Subsequent cell-free glycan remodeling reactions using Sx-Δ29*Hs*GnTII and Sx-Δ44*Hs*β4GalT1 furnished trastuzumab with glycans 3 and 4, respectively. Finally, Sx-Δ26*Hs*ST6Gal1 was used to cap glycan 4 with Neu5Ac, efficiently generating glycans 5 and, to a lesser extent, glycan 6 (FIG. 7B). Additional *N*-glycan structures including paucimannose (glycan 1), hybrid (glycan 13, 14), and complex (glycan 7, 15) types were also prepared directly on trastuzumab IgG-Fc using a variety of Sx-GTs (FIGS. 21A–21B). In most cases, glycan remodeling efficiency was near 100% following incubation with Sx-GTs for 16-36 hours at 37°C with approximately 80-90% recovery yield from purification between each step. Only the conversion to the disialylated *N*-glycan using Sx-Δ26HsST6Gal1 resulted in notably lower efficiency. While the reason for this inefficiency is unclear, it likely results from human ST6Gal1's known preference for the α1–3Man branch (Barb et al., "Branch-specific Sialylation of IgG-Fc Glycans by ST6Gal-I," *Biochemistry* 48:9705-7 (2009), which is hereby incorporated by reference in its entirety).

[0160] In addition to producing authentic, homogeneous human *N*-glycans, whether Sx-GTs could generate IgG-Fc bearing unnatural glycan structures was also investigated. To this end, Sx-Δ29*Hs*GnTI was used to elaborate trastuzumab *N*-glycans with *N*-azidoacetylglucosamine (GlcNAz), a synthetic monosaccharide containing an azide moiety (FIGS. 22A–22B, glycan 13) that served as a versatile chemical handle for regiospecific conjugation via bioorthogonal click chemistry. Indeed, it was possible to site-specifically modify this handle on trastuzumab with either a biotin group (glycan 17) or a fluorescent reporter (glycan 18), thereby providing a convenient route for extending the functional utility of Fc domain-linked glycans. Collectively, these results highlight the biocatalytic potential of SIMPLEx glycoenzymes in the construction of homogeneous glycans as both free and protein-linked structures and effectively paves the way for accelerating protein glycosylation studies as well as tailoring the biological, biophysical, and biomedical properties of glycoproteins.

Discussion of Examples 1 – 9

[0161] Examples 1–9 describe the creation of a universal expression platform for producing nearly 100 different GTs, predominantly of human origin, at relatively high titers

(approximately 5–10 mg/L) using standard bacterial culture. This platform leverages SIMPLEx to engineer GT chimeras that are rendered highly soluble in the cytoplasm of *E. coli* cells. Consistent with earlier works (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat. Commun.* 6:6826 (2015) and Mizrachi et al., "A Water-soluble DsbB Variant That Catalyzes Disulfide-bond Formation In Vivo," *Nat. Chem. Biol.* 13:1022-1028 (2017), which are hereby incorporated by reference in their entirety), SIMPLEx-reformatted GTs retained biological activity as exemplified by the human ST6Gal1 chimera that exhibited activity that was similar to a commercially sourced enzyme. The ability to solubilize such a large set of GTs without compromising function made it possible to remodel the structures of different free and protein-linked glycans including those found on the monoclonal antibody trastuzumab. Overall, the platform described *infra* represents a versatile addition to the synthetic glycobiology toolkit, providing easy access to a vast collection of transformative reagents that are expected to find use in structure-function studies of GTs and to fuel myriad applications where complex glycomolecules are featured.

[0162] Previous studies revealed the capacity of SIMPLEx to broadly transform all major classes of IMPs into water-soluble molecules (Mizrachi et al., "Making Water-soluble Integral Membrane Proteins In Vivo Using an Amphipathic Protein Fusion Strategy," *Nat. Commun.* 6:6826 (2015) and Mizrachi et al., "A Water-soluble DsbB Variant That Catalyzes Disulfide-bond Formation In Vivo," *Nat. Chem. Biol.* 13:1022-1028 (2017), which are hereby incorporated by reference in their entirety). These IMPs included proteins having both bitopic and polytopic α-helical structures such as glutamate receptor (GluA2) and bacteriorhodopsin (bR) as well as polytopic β-barrel structures such as voltage-dependent anion channel 1 (VDAC1). Here, this solubilization capacity was broadened to include polytopic α-helical GTs with multiple TMDs such as found in human mannosyltransferases Alg2, Alg3, and Alg12 and human glucosyltransferases Alg6, Alg8, and Alg10 as well as monotopic α-helical GTs with single-pass internal TMDs that could not be easily removed such as Alg2 and PigA. For these complex integral membrane proteins, introduction of an N-terminal decoy protein, MBP, prevented co-translational insertion of the polypeptide into the inner membrane through the signal recognition particle (SRP) pathway (Luirink and Sinning, "SRP-mediated Protein Targeting: Structure and Function Revisited," *Biochim. Biophys. Acta.* 1694:17-35 (2004), which is hereby incorporated by reference in its entirety) while the amphipathic ApoAI* domain effectively shielded the hydrophobic TMDs from the aqueous environment.

[0163] It is noteworthy that most of the GTs investigated (72 out of 98 total) were simpler type II transmembrane proteins. Type II GTs such as *Hs*ST6Gal1 possess just a single-pass TMD at their N- or C-termini (FIG. 3A), which is generally not required for activity and is

thus commonly removed during expression campaigns (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018), which is hereby incorporated by reference in its entirety), which is hereby incorporated by reference in its entirety). Hence, while the rationale for using SIMPLEx with full-length GTs including their TMDs was clear, it was not obvious that this solubilization method would benefit type II GTs lacking a TMD altogether. That said, removal of GTs from their transmembrane contexts and the lack of native interacting/folding partners can create difficulties in folding upon expression in heterologous hosts. Indeed, many N-terminally truncated GTs accumulated exclusively in the insoluble or detergent-solubilized fractions of *E. coli* cells, in agreement with numerous previous reports involving expression of truncated GTs in bacteria. One possible reason for this poor expression is that GTs contain several moderately hydrophobic segments including around the stem region, just after the TMD, that can trigger unwanted membrane targeting or otherwise drive misfolding and aggregation. To circumvent this issue, fusion of solubility-enhancing partners such as MBP to the N-termini of truncated GTs is often obligatory, even in mammalian cells where GTs fused with GFP expressed significantly better than unfused versions of the same glycoenzymes (Moremen Et Al., "Expression System For Structural And Functional studies of Human Glycosylation Enzymes," *Nat. Chem. Biol.* 14:156-162 (2018), which is hereby incorporated by reference in its entirety). However, introduction of fusion tags does not always lead to immediate success in terms of soluble GT expression and thus often requires lengthy optimization of growth and induction conditions as well as trial-and-error evaluation of different host strains and accessory factor (*e.g.*, molecular chaperone) co-expression strategies (Skretas et al., "Expression of Active Human Sialyltransferase ST6GalNAcI in Escherichia coli," *Microb Cell Fact* 8:50 (2009); Ortiz-Soto and Seibel, "Expression of Functional Human Sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli," *PLoS One* 11:e0155410 (2016); and Hidari et al., "Purification and Characterization of a Soluble Recombinant Human ST6Gal I Functionally Expressed in Escherichia coli," *Glycoconj J* 22:1-11 (2005), which are hereby incorporated by reference in their entirety). Even when appreciable solubility is achieved, it is quite common for the resulting MBP fusions to accumulate as soluble but heterogeneous multimeric aggregates in which only a small fraction of the fusion protein is properly folded and active (Skretas et al., "Expression of Active Human Sialyltransferase ST6GalNAcI in Escherichia coli," *Microb Cell Fact* 8:50 (2009); Ortiz-Soto and Seibel, "Expression of Functional Human Sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli," *PLoS One* 11:e0155410 (2016); and Pasek et al., "Galectin-1 as a Fusion Partner for the Production of Soluble and Folded Human Beta-1,4-galactosyltransferase-T7 in E. coli," *Biochem Biophys Res Commun* 394:679-84 (2010), which are hereby incorporated

by reference in their entirety). Along these lines, >50% less activity was observed for MBP-tagged Δ26*Hs*ST6Gal1 (lacking the native TMD) compared to its SIMPLEx counterpart, underscoring the essential contribution made by the ApoAI* domain in promoting solubility of a type II GT in the *E. coli* cytoplasm. It is therefore hypothesized that the amphipathic nature of ApoAI*, when expressed in the proximity of exposed hydrophobic patches in type II GT proteins, provides a stabilizing effect via hydrophobic interaction, akin to how ApoAI-based nanodiscs solubilize membrane proteins in solution (Grinkova et al., "Engineering Extended Membrane Scaffold Proteins for Self-assembly of Soluble Nanoscale Lipid Bilayers," *Protein Eng. Des. Sel.* 23:843-8 (2010), which is hereby incorporated by reference in its entirety).

[0164] Importantly, the SIMPLEx architecture enabled soluble expression for nearly 100 GTs (>95% "hit" rate) under standard, identically matched conditions without any optimization, thereby offering a universal solution to GT production in *E. coli* that has not been possible with stand-alone fusion tags such as MBP or other expression optimization techniques (Wagner et al., "Rationalizing Membrane Protein Overexpression," *Trends Biotechnol.* 24:364-71 (2006), which is hereby incorporated by reference in its entirety). An additional layer of universality stems from the compatibility of SIMPLEx-mediated GT solubilization with other commonly used expression hosts such as yeast and HEK293 cells as well as with *E. coli*-based cell-free protein synthesis (CFPS). Such platform flexibility is significant for several reasons. For one, each of these platforms is amenable to high-throughput profiling of protein expression and production that can be scaled up to larger volumes (Subedi et al., "High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension," *J. Vis. Exp.* e53568 (2015) and Spirin, A.S,. "High-throughput Cell-free Systems for Synthesis of Functionally Active Proteins," *Trends Biotechnol.* 22:538-45 (2004), which are hereby incorporated by reference in their entirety). Moreover, in the case of yeast and HEK293, the compatibility of SIMPLEx-reformatted GTs in these well-established eukaryotic hosts may provide access to protein folding networks and post-translational modifications including *N*- and *O*-linked glycosylation that may be important for the biological function of a subset of GTs (Mikolajczyk et al., "How Glycosylation Affects Glycosylation: The Role of N-glycans in Glycosyltransferase Activity," *Glycobiology* 30:941-969 (2020), which is hereby incorporated by reference in its entirety) but are natively lacking in standard *E. coli* strains. In the case of *E. coli*-based CFPS, the "open" nature and multiplexability of these systems, combined with their speed and simplicity, should provide opportunities for high-throughput screening of GT function (Kightlinger et al., "Design of Glycosylation Sites by Rapid Synthesis and Analysis of Glycosyltransferases," *Nat. Chem. Biol.* 14(6):627-635 (2018), which is hereby incorporated by reference in its entirety) as well as rapid discovery, prototyping, and optimization of

glycomolecule synthesis pathways (Karim and Jewett, "A Cell-free Framework for Rapid Biosynthetic Pathway Prototyping and Enzyme Discovery," *Metab. Eng.* 36:116-126 (2016) and Kightlinger et al., "A Cell-free Biosynthesis Platform for Modular Construction of Protein Glycosylation Pathways," *Nat. Commun.* 10:5404 (2019), which are hereby incorporated by reference in their entirety).

[0165] As proof of concept for the utility of the disclosed SIMPLEx pipeline, some of the solubilized products were used in coordinated cell-free reaction networks to catalyze the formation of chemically-defined *N*-glycans. In one instance, it was possible to transform quantitative amounts of a simple paucimannose precursor *N*-glycan, $Man_3GlcNAc_2$ derived from glycoengineered *E. coli* (Valderrama-Rincon et al., "An Engineered Eukaryotic Protein Glycosylation Pathway in Escherichia coli," *Nat. Chem. Biol.* 8:434-6 (2012) and Glasscock et al., "A Flow Cytometric Approach to Engineering Escherichia coli for Improved Eukaryotic Protein Glycosylation," *Metab. Eng.* 47:488-495 (2018), which are hereby incorporated by reference in their entirety), into complex biantennary *N*-glycans including those containing core-fucose and sialic acid caps using a set of SIMPLEx-reformatted GTs. This workflow to efficiently generate a library of complex *N*-glycans, starting from expression and purification and then finally utilization of SIMPLEx-reformatted GTs, could be completed in less than one week. Using an identical strategy, it was possible to generate a spectrum of homogenous *N*-glycan structures on intact glycoproteins including trastuzumab, a mAb therapy used to treat breast and stomach cancers. Akin to earlier engineering of an artificial cytoplasmic disulfide formation pathway involving a water-soluble SIMPLEx variant of DsbB (Mizrachi et al., "A Water-soluble DsbB Variant That Catalyzes Disulfide-bond Formation In Vivo," *Nat. Chem. Biol.* 13:1022-1028 (2017), which is hereby incorporated by reference in its entirety), ensembles of SIMPLEx-reformatted GTs could similarly be assembled into designer pathways, either *in vitro* or in living cells, for the on-demand biosynthesis of important glycans and glycoconjugates. Looking forward, it is anticipated that the constructs, expression systems, and workflows for glycoenzyme production described herein will find widespread use by those seeking to push the boundaries of our knowledge of glycobiology and glycochemistry and its application in health, energy, and materials science.

[0166] Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions, and the like can be made without departing from the spirit of the invention and these are therefore considered to be within the scope of the invention as defined in the claims which follow.

WHAT IS CLAIMED:

1. A nucleic acid construct comprising:

a chimeric nucleic acid molecule encoding a tripartite glycosyltransferase fusion protein, said chimeric nucleic acid molecule comprising:

a first nucleic acid moiety encoding an amphipathic shield domain protein;

a second nucleic acid moiety encoding a glycosyltransferase; and

a third nucleic acid moiety encoding a water soluble expression decoy protein, wherein said first nucleic acid moiety is coupled to said second nucleic acid moiety's 3′ end and said third nucleic acid moiety is coupled to said second nucleic acid moiety's 5′ end, said coupling being direct or indirect.

2. The nucleic acid construct according to claim 1, wherein the amphipathic shield domain protein is selected from the group consisting of apolipoprotein A (ApoA), apolipoprotein B (ApoB), apolipoprotein C (ApoC), apolipoprotein D (ApoD), apolipoprotein E (ApoE), apolipoprotein H (ApoH), truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*), and a peptide self-assembly mimic (PSAM).

3. The nucleic acid construct according to claim 1, wherein the amphipathic shield domain protein is human apolipoprotein A1.

4. The nucleic acid construct according to claim 3, wherein the human apolipoprotein A1 is a truncated human apolipoprotein A1.

5. The nucleic acid construct according to claim 4, wherein the truncated human apolipoprotein A1 protein is truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*).

6. The nucleic acid construct according to any one of claims 1, wherein the glycosyltransferase is a truncated glycosyltransferase.

7. The nucleic acid construct according to any one of claims 1 to 6, wherein the glycosyltransferase is a prokaryotic glycosyltransferase.

8. The nucleic acid construct according to any one of claims 1 to 6, wherein the glycosyltransferase is a eukaryotic glycosyltransferase.

9. The nucleic acid construct according to claim 8, wherein the glycosyltransferase is a human glycosyltransferase.

10. The nucleic acid construct according to any of claims 1 to 9, wherein the glycosyltransferase is selected from the group consisting of (i) a single-pass transmembrane protein with C-terminus in cytoplasm (type I transmembrane protein); (ii) a single-pass transmembrane protein with N-terminus in cytoplasm (type II transmembrane protein); (iii) a multi-pass transmembrane protein; and (iv) a secretory protein with N-terminal signal peptide and C-terminal ER retention domain.

11. The nucleic acid construct according to any of claims 1 to 9, wherein the glycosyltransferase is selected from the group consisting of fucosyltransferases (FucTs), galactosyltransferases (Gals), glucosyltransferases (GlcTs), mannosyltransferases (ManTs), N-acetylgalactosyltransferases (GalNAcTs), N-acetylglucosaminyltransferases (GlcNAcTs), and sialyltransferases (SiaTs).

12. The nucleic acid construct according to any of claims 1 to 9, wherein the glycosyltransferase is selected from the group consisting of human galactoside 2-alpha-L-fucosyltransferase 1 (*Hs*FUT1), human galactoside 2-alpha-L-fucosyltransferase 2 (*Hs*FUT2), HUMAN Galactoside 3(4)-L-fucosyltransferase (*Hs*FUT3), human alpha-(1,3)-fucosyltransferase 4 (*Hs*FUT4), human alpha-(1,3)-fucosyltransferase 5 (*Hs*FUT5), human alpha-(1,3)-fucosyltransferase 6 (*Hs*FUT6), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human alpha-(1,6)-fucosyltransferase (*Hs*FUT8), human alpha-(1,3)-fucosyltransferase 9 (*Hs*FUT9), human alpha-(1,3)-fucosyltransferase 10 (*Hs*FUT10), human alpha-(1,3)-fucosyltransferase 11 (*Hs*FUT11), human GDP-fucose protein O-fucosyltransferase 1 (*Hs*POFUT1), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase (*Hs*ST3Gal3), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4 (*Hs*ST3Gal4), human type 2 lactosamine alpha-2,3-sialyltransferase (*Hs*ST3Gal6), human beta-galactoside alpha-2,6-sialyltransferase 1 (*Hs*ST6Gal1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*Hs*ST6GalNAc1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2 (*Hs*ST6GalNAc2), human alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase (*Hs*ST6GalNAc4), human alpha-N-acetylneuraminide alpha-2,8-sialyltransferase (*Hs*ST8Sia1),

human alpha-2,8-sialyltransferase 8b (*Hs*ST8Sia2), human sia-alpha-2,3-gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase (*Hs*ST8Sia3), human CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase (*Hs*ST8Sia4), human polypeptide N-acetylgalactosaminyltransferase 1 (*Hs*ppGalNAcT1), human polypeptide N-acetylgalactosaminyltransferase 2 (*Hs*ppGalNAcT2), human polypeptide N-acetylgalactosaminyltransferase 3 (*Hs*ppGalNAcT3), human polypeptide N-acetylgalactosaminyltransferase 4 (*Hs*ppGalNAcT4), human polypeptide N-acetylgalactosaminyltransferase 5 (*Hs*ppGalNAcT5), human polypeptide N-acetylgalactosaminyltransferase 6 (*Hs*ppGalNAcT6), human N-acetylgalactosaminyltransferase 7 (*Hs*ppGalNAcT7), human probable polypeptide N-acetylgalactosaminyltransferase 8 (*Hs*ppGalNAcT8), human polypeptide N-acetylgalactosaminyltransferase 9 (*Hs*ppGalNAcT9), human polypeptide N-acetylgalactosaminyltransferase 10 (*Hs*ppGalNAcT10), human UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1 (*Hs*B3GALNT1), human beta-1,4 N-acetylgalactosaminyltransferase 1 (*Hs*B4GALNT1), human histo-blood group ABO system transferase (*Hs*-A-group), human lactosylceramide 4-alpha-galactosyltransferase (*Hs*A4GalT), human beta-1,3-galactosyltransferase 1 (*Hs*B3GalT1), human beta-1,3-galactosyltransferase 2 (*Hs*B3GalT2), human beta-1,4-galactosyltransferase 1 (*Hs*B4GalT1), human beta-1,4-galactosyltransferase 2 (*Hs*B4GalT2), human beta-1,4-galactosyltransferase 3 (*Hs*B4GalT3), human beta-1,4-galactosyltransferase 4 (*Hs*B4GalT4), human beta-1,4-galactosyltransferase 5 (*Hs*B4GalT5), human beta-1,4-galactosyltransferase 6 (*Hs*B4GalT6), human histo-blood group ABO system transferase (*Hs*-B-group), human 2-hydroxyacylsphingosine 1-beta-galactosyltransferase (*Hs*UGT8), human glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 (*Hs*C1GLT), human C1GALT1-specific chaperone 1 (*Hs*COSMC), human chitobiosyldiphosphodolichol beta-mannosyltransferase (*Hs*Alg1), human alpha-1,3/1,6-mannosyltransferase (*Hs*Alg2), human Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase (*Hs*Alg3), human GDP-man:man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Hs*Alg11), human dol-p-man:man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase (*Hs*Alg12), human isoform 2 of putative UDP-N-acetylglucosamine transferase (*Hs*Alg13), human UDP-N-acetylglucosamine transferase subunit alg14 homolog (*Hs*Alg14), human dolichol-phosphate mannosyltransferase subunit 1 (*Hs*DPM1), human GPI mannosyltransferase 1 (*Hs*PIGM), human GPI mannosyltransferase 3 (*Hs*PIGB), human GPI mannosyltransferase 4 (*Hs*PIGZ), human dolichyl-phosphate beta-glucosyltransferase (*Hs*Alg5), human dolichyl pyrophosphate $man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg6), human probable dolichyl pyrophosphate $Glc1Man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg8), human Dol-P-Glc:$Glc_2Man_9GlcNAc_2$-PP-Dol alpha-1,2-glucosyltransferase (*Hs*Alg10), human ceramide glucosyltransferase (*Hs*UGCG), human beta-1,3-glucosyltransferase (*Hs*B3GLCT),

human glycogenin-1 (*Hs*GLYG), human protein O-glucosyltransferase 1 (*Hs*POGLUT1), human alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (*Hs*GnTIII/MGAT3), human alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase a (*Hs*GnTIV/MGAT4), human beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (*Hs*GCNT1), human N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase (*Hs*GCNT2), human N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2 (*Hs*B3GNT2), human acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase (*Hs*B3GNT6), human phosphatidylinositol N-acetylglucosaminyltransferase subunit A (*Hs*PIGA), human xyloside xylosyltransferase 1 (*Hs*XXLT1), human UDP-glucuronosyltransferase 1-1 (*Hs*UGT1A1), human beta-1,4-glucuronyltransferase 1 (*Hs*B4GAT1), human UDP-glucuronosyltransferase 1-3 (*Hs*UGT1A3), *Campylobacter jejuni* CsTII (*Cj*CstII), *Neisseria meningitidis* polysialic acid O-acetyltransferase (*Nm*Pst), *Campylobacter jejuni* beta-1,3-galactosyltransferase (*Cj*CgtB), *Helicobacter pylori* (strain 51) beta-4-galactosyltransferase (*Hp*LgtB), *Neisseria meningitidis* serogroup B (strain MC58) lacto-N-neotetraose biosynthesis glycosyltransferase LgtB (*Nm*LgtB), *Neisseria gonorrhoeae* lacto-N-neotetraose biosynthesis glycosyltransferase (*Ng*LgtB), *E.coli* galactoside 2-alpha-L-fucosyltransferase WbgL (*Ec*WbgL), *E.coli* undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (*Ec*WecA), *Legionella pneumophila* subsp. Pneumophila Subversion of eukaryotic traffic protein A (*Lp*SetA), *Neisseria meningitidis* alpha-2,9-polysialyltransferase (*Nm*SynE), yeast beta-1,4-mannosyltransferase OS=*saccharomyces cerevisiae* (*Sc*Alg1), yeast GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Sc*Alg11), *Nicotiana tabacum* alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Nt*GnTI), *Nicotiana tabacum* alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like (*Nt*GnTII), *Bos taurus* n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Bt*GGTA1), mouse n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Mm*GGTA1), rat n-acetyllactosaminide alpha-1,3-galactosyltransferase (RnGGTA1), and *Bos taurus* beta-1,4-galactosyltransferase 1 (BtB4GalT1).

13. The nucleic acid construct according to any one of claims 1 to 12, wherein the glycosyltransferase is selected from the group consisting of *Campylobacter jejuni* CsTII (*Cj*CstII), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human alpha-1,3-mannosyl-

glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-galactosyltransferase 1 (*Hs*β4GalT1), human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1), and human alpha-(1,6)-fucosyltransferase (*Hs*FUT8).

14. The nucleic acid construct according to any one of claims 1 to 13, wherein the water soluble expression decoy protein is selected from the group consisting of outer surface protein (OspA) lacking its native export signal peptide, DnaB lacking its native export signal peptide, and maltose-binding protein (MBP) lacking its N-terminal signal peptide.

15. The nucleic acid construct according to any one of claims 1 to 14, wherein the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide.

16. The nucleic acid construct according to any one of claims 1 to 15, wherein the amphipathic shield domain protein is truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*) and the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide.

17. The nucleic acid construct according to any one of claims 1 to 16, wherein the nucleic acid construct further comprises:

a promoter and

a termination sequence, wherein said promoter and said termination sequence are operatively coupled to the chimeric nucleic acid molecule.

18. The nucleic acid construct according to any one of claims 1 to 17, wherein the chimeric nucleic acid molecule further comprises:

one or more linker nucleic acid moieties coupling said first, second, and/or third nucleic acid moieties together.

19. An expression vector comprising the nucleic acid construct of any one of claims 1 to 18.

20. A host cell comprising the nucleic acid construct of any one of claims 1 to 18 or the expression vector of claim 19.

21. The host cell according to claim 20, wherein the host cell is prokaryotic.

22. The host cell according to claim 21, wherein the prokaryotic cell is *E. coli*.

23. The host cell according to claim 20, where in the host cell is eukaryotic.

24. The host cell according to claim 23, wherein the eukaryotic cell is a yeast cell.

25. The host cell according to claim 23, wherein the eukaryotic cell is a human cell line.

26. A tripartite glycosyltransferase fusion protein produced by the host cell according to any one of claims 20-25.

27. A cell-free protein expression system comprising:
a cell lysate or extract and
the nucleic acid construct according to anyone of claims 1 to 18 or the expression vector according to claim 19.

28. The cell-free protein expression system according to claim 27, wherein the cell lysate or extract comprises a heterologous and/or recombinant RNA polymerase.

29. The cell-free protein expression system according to claim 27 or claim 28, wherein the cell lysate or extract is capable of (i) transcribing the nucleic acid construct or the vector to form a translation template and (ii) translating the translation template.

30. The cell-free protein expression system according to any of claim 27 to claim 29, wherein the cell lysate or extract is an *E. coli* lysate or extract.

31. A tripartite glycosyltransferase fusion protein produced by the cell-free expression system according to any one of claims 27-30.

32. A method of recombinantly producing a tripartite glycosyltransferase fusion protein in water soluble form, said method comprising:

providing the host cell of any one of claims 20 to 25 or the cell-free expression system of any one of claims 27 to 30 and

culturing the host cell or using the cell-free expression system under conditions effective to express the tripartite glycosyltransferase fusion protein in a water soluble form within the host cell cytoplasm or the cell-free expression system.

33. The method according to claim 32 further comprising:

recovering the tripartite glycosyltransferase fusion protein from the host cell or the cell-free expression system following said culturing or said using, respectively.

34. The method according to claim 33, wherein the host cell is provided and said recovering comprises:

lysing the host cell to form a cell lysate comprising a water soluble fraction; and

subjecting the water soluble fraction of the cell lysate to chromatography to isolate the tripartite glycosyltransferase fusion protein .

35. The method according to claim 33, wherein the cell-free expression system is provided and said recovering comprises:

subjecting the water soluble fraction of the cell lysate to chromatography to isolate the tripartite glycosyltransferase fusion protein.

36. The method according to any one of claims 33 to claim 35, wherein the recovered tripartite glycosyltransferase fusion protein is conformationally correct.

37. A tripartite glycosyltransferase fusion protein produced by the methods of any one of claims 32 to 36.

38. A tripartite glycosyltransferase fusion protein comprising:

an amino terminal water soluble expression decoy protein;

a glycosyltransferase; and

a carboxyl terminal amphipathic shield domain protein.

39. The tripartite glycosyltransferase fusion protein according to claim 38, wherein the amphipathic shield domain protein is selected from the group consisting of apolipoprotein A (ApoA), apolipoprotein B (ApoB), apolipoprotein C (ApoC), apolipoprotein D (ApoD), apolipoprotein E (ApoE), apolipoprotein H (ApoH), truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*), and a peptide self-assembly mimic (PSAM).

40. The tripartite glycosyltransferase fusion protein according to claim 38, wherein the amphipathic shield domain protein is human apolipoprotein A1.

41. The tripartite glycosyltransferase fusion protein according to claim 40, wherein the human apolipoprotein A1 is a truncated human apolipoprotein A1.

42. The tripartite glycosyltransferase fusion protein according to claim 41, wherein the truncated human apolipoprotein A1 protein is truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*).

43. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 42, wherein the glycosyltransferase is a truncated glycosyltransferase.

44. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 43, wherein the glycosyltransferase is a prokaryotic glycosyltransferase.

45. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 43, wherein the glycosyltransferase is a eukaryotic glycosyltransferase.

46. The tripartite glycosyltransferase fusion protein according to claim 45, wherein the glycosyltransferase is a human glycosyltransferase.

47. The tripartite glycosyltransferase fusion protein according to any of claims 38 to 46, wherein the glycosyltransferase is selected from the group consisting of (i) a single-pass transmembrane protein with C-terminus in cytoplasm (type I transmembrane protein); (ii) a single-pass transmembrane protein with N-terminus in cytoplasm (type II transmembrane protein); (iii) a multi-pass transmembrane protein; and (iv) a secretory protein with N-terminal signal peptide and C-terminal ER retention domain.

48. The tripartite glycosyltransferase fusion protein according to any of claims 38 to 46, wherein the glycosyltransferase is selected from the group consisting of fucosyltransferases (FucTs), galactosyltransferases (Gals), glucosyltransferases (GlcTs), mannosyltransferases (ManTs), N-acetylgalactosyltransferases (GalNAcTs), N-acetylglucosaminyltransferases (GlcNAcTs), and sialyltransferases (SiaTs).

49. The tripartite glycosyltransferase fusion protein according to any of claims 38 to 46, wherein the glycosyltransferase is selected from the group consisting of wherein the glycosyltransferase is selected from the group consisting of human galactoside 2-alpha-L-fucosyltransferase 1 (*Hs*FUT1), human galactoside 2-alpha-L-fucosyltransferase 2 (*Hs*FUT2), HUMAN Galactoside 3(4)-L-fucosyltransferase (*Hs*FUT3), human alpha-(1,3)-fucosyltransferase 4 (*Hs*FUT4), human alpha-(1,3)-fucosyltransferase 5 (*Hs*FUT5), human alpha-(1,3)-fucosyltransferase 6 (*Hs*FUT6), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human alpha-(1,6)-fucosyltransferase (*Hs*FUT8), human alpha-(1,3)-fucosyltransferase 9 (*Hs*FUT9), human alpha-(1,3)-fucosyltransferase 10 (*Hs*FUT10), human alpha-(1,3)-fucosyltransferase 11 (*Hs*FUT11), human GDP-fucose protein O-fucosyltransferase 1 (*Hs*POFUT1), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase (*Hs*ST3Gal3), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4 (*Hs*ST3Gal4), human type 2 lactosamine alpha-2,3-sialyltransferase (*Hs*ST3Gal6), human beta-galactoside alpha-2,6-sialyltransferase 1 (*Hs*ST6Gal1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (*Hs*ST6GalNAc1), human alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2 (*Hs*ST6GalNAc2), human alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase (*Hs*ST6GalNAc4), human alpha-N-acetylneuraminide alpha-2,8-sialyltransferase (*Hs*ST8Sia1), human alpha-2,8-sialyltransferase 8b (*Hs*ST8Sia2), human sia-alpha-2,3-gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase (*Hs*ST8Sia3), human CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase (*Hs*ST8Sia4), human polypeptide N-acetylgalactosaminyltransferase 1 (*Hs*ppGalNAcT1), human polypeptide N-acetylgalactosaminyltransferase 2 (*Hs*ppGalNAcT2), human polypeptide N-acetylgalactosaminyltransferase 3 (*Hs*ppGalNAcT3), human polypeptide N-acetylgalactosaminyltransferase 4 (*Hs*ppGalNAcT4), human polypeptide N-acetylgalactosaminyltransferase 5 (*Hs*ppGalNAcT5), human polypeptide N-acetylgalactosaminyltransferase 6 (*Hs*ppGalNAcT6), human N-acetylgalactosaminyltransferase 7 (*Hs*ppGalNAcT7), human probable polypeptide N-acetylgalactosaminyltransferase 8

(*Hs*ppGalNAcT8), human polypeptide N-acetylgalactosaminyltransferase 9 (*Hs*ppGalNAcT9), human polypeptide N-acetylgalactosaminyltransferase 10 (*Hs*ppGalNAcT10), human UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1 (*Hs*B3GALNT1), human beta-1,4 N-acetylgalactosaminyltransferase 1 (*Hs*B4GALNT1), human histo-blood group ABO system transferase (*Hs*-A-group), human lactosylceramide 4-alpha-galactosyltransferase (*Hs*A4GalT), human beta-1,3-galactosyltransferase 1 (*Hs*B3GalT1), human beta-1,3-galactosyltransferase 2 (*Hs*B3GalT2), human beta-1,4-galactosyltransferase 1 (*Hs*B4GalT1), human beta-1,4-galactosyltransferase 2 (*Hs*B4GalT2), human beta-1,4-galactosyltransferase 3 (*Hs*B4GalT3), human beta-1,4-galactosyltransferase 4 (*Hs*B4GalT4), human beta-1,4-galactosyltransferase 5 (*Hs*B4GalT5), human beta-1,4-galactosyltransferase 6 (*Hs*B4GalT6), human histo-blood group ABO system transferase (*Hs*-B-group), human 2-hydroxyacylsphingosine 1-beta-galactosyltransferase (*Hs*UGT8), human glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 (*Hs*C1GLT), human C1GALT1-specific chaperone 1 (*Hs*COSMC), human chitobiosyldiphosphodolichol beta-mannosyltransferase (*Hs*Alg1), human alpha-1,3/1,6-mannosyltransferase (*Hs*Alg2), human Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase (*Hs*Alg3), human GDP-man:man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Hs*Alg11), human dol-p-man:man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase (*Hs*Alg12), human isoform 2 of putative UDP-N-acetylglucosamine transferase (*Hs*Alg13), human UDP-N-acetylglucosamine transferase subunit alg14 homolog (*Hs*Alg14), human dolichol-phosphate mannosyltransferase subunit 1 (*Hs*DPM1), human GPI mannosyltransferase 1 (*Hs*PIGM), human GPI mannosyltransferase 3 (*Hs*PIGB), human GPI mannosyltransferase 4 (*Hs*PIGZ), human dolichyl-phosphate beta-glucosyltransferase (*Hs*Alg5), human dolichyl pyrophosphate $man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg6), human probable dolichyl pyrophosphate $Glc1Man_9GlcNAc_2$ alpha-1,3-glucosyltransferase (*Hs*Alg8), human Dol-P-Glc:$Glc_2Man_9GlcNAc_2$-PP-Dol alpha-1,2-glucosyltransferase (*Hs*Alg10), human ceramide glucosyltransferase (*Hs*UGCG), human beta-1,3-glucosyltransferase (*Hs*B3GLCT), human glycogenin-1 (*Hs*GLYG), human protein O-glucosyltransferase 1 (*Hs*POGLUT1), human alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (*Hs*GnTIII/MGAT3), human alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase a (*Hs*GnTIV/MGAT4), human beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (*Hs*GCNT1), human N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase (*Hs*GCNT2), human N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2 (*Hs*B3GNT2), human

acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase (*Hs*B3GNT6), human phosphatidylinositol N-acetylglucosaminyltransferase subunit A (*Hs*PIGA), human xyloside xylosyltransferase 1 (*Hs*XXLT1), human UDP-glucuronosyltransferase 1-1 (*Hs*UGT1A1), human beta-1,4-glucuronyltransferase 1 (*Hs*B4GAT1), human UDP-glucuronosyltransferase 1-3 (*Hs*UGT1A3), *Campylobacter jejuni* CsTII (*Cj*CstII), *Neisseria meningitidis* polysialic acid O-acetyltransferase (*Nm*Pst), *Campylobacter jejuni* beta-1,3-galactosyltransferase (*Cj*CgtB), *Helicobacter pylori* (strain 51) beta-4-galactosyltransferase (*Hp*LgtB), *Neisseria meningitidis* serogroup B (strain MC58) lacto-N-neotetraose biosynthesis glycosyltransferase LgtB (*Nm*LgtB), *Neisseria gonorrhoeae* lacto-N-neotetraose biosynthesis glycosyltransferase (*Ng*LgtB), *E.coli* galactoside 2-alpha-L-fucosyltransferase WbgL (*Ec*WbgL), *E.coli* undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (*Ec*WecA), *Legionella pneumophila* subsp. Pneumophila Subversion of eukaryotic traffic protein A (*Lp*SetA), *Neisseria meningitidis* alpha-2,9-polysialyltransferase (*Nm*SynE), yeast beta-1,4-mannosyltransferase OS=*saccharomyces cerevisiae* (*Sc*Alg1), yeast GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase (*Sc*Alg11), *Nicotiana tabacum* alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Nt*GnTI), *Nicotiana tabacum* alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like (*Nt*GnTII), *Bos taurus* n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Bt*GGTA1), mouse n-acetyllactosaminide alpha-1,3-galactosyltransferase (*Mm*GGTA1), rat n-acetyllactosaminide alpha-1,3-galactosyltransferase (RnGGTA1), and *Bos taurus* beta-1,4-galactosyltransferase 1 (BtB4GalT1).

50. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 49, wherein the glycosyltransferase is selected from the group consisting of *Campylobacter jejuni* CsTII (*Cj*CstII), human alpha-(1,3)-fucosyltransferase 7 (*Hs*FUT7), human CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1 (*Hs*ST3Gal1), human alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTI/MGAT1), human alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (*Hs*GnTII/MGAT2), human beta-1,4-galactosyltransferase 1 (*Hs*β4GalT1), human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1), and human alpha-(1,6)-fucosyltransferase (*Hs*FUT8).

51. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 50, wherein the water soluble expression decoy protein is selected from the group consisting of outer surface protein (OspA) lacking its native export signal peptide, DnaB lacking

its native export signal peptide, and maltose-binding protein (MBP) lacking its N-terminal signal peptide.

52. The tripartite glycosyltransferase fusion protein according to any one of claims 38 to 51, wherein the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide.

53. The tripartite glycosyltransferase fusion protein according to claim 38, wherein the amphipathic shield domain protein is truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*) and the water soluble expression decoy protein is maltose-binding protein (MBP) lacking its N-terminal signal peptide.

54. A method of cell-free glycan remodeling, said method comprising:

providing a glycan primer;

providing one or more tripartite glycosyltransferase fusion protein(s) according to any one of claims 26, 31, 37, or claims 38 to 53; and

incubating the glycan primer with the one or more tripartite glycosyltransferase fusion protein(s) under conditions effective to transfer a glycosyl group to the glycan primer to produce a modified glycan structure.

55. The method according to claim 54, wherein the glycan primer is a monosaccharide.

56. The method according to claim 54, wherein the glycan primer is an oligosaccharide.

57. The method according to claim 56, wherein the glycan primer comprises $Man_3GlcNAc_2$ or $Man_5GlcNAc_2$.

58. The method according to any one of claims 54 to 57, wherein the glycan primer is attached to an amino acid residue.

59. The method according to claim 58, wherein the amino acid residue is an asparagine residue.

60. The method according to any one of claim 58 or claim 59, wherein the glycan primer is attached to a glycoprotein.

61. The method according to claim 60, wherein the glycoprotein is an antibody.

62. The method according to any one of claims claim 54 to 61, wherein the tripartite glycosyltransferase fusion protein is selected from the group consisting of Sx-Δ29*Hs*GnTI, Sx-Δ29*Hs*GnTII, Sx-Δ30*Hs*FucT8, Sx-Δ44*Hs*β4GalT1, Sx-Δ26*Hs*ST6Gal1, and combinations thereof.

63. The method according to any one of claims 54 to 62, wherein said incubating is carried out with a plurality of different tripartite glycosyltransferase fusion proteins, at least some of the different tripartite glycosyltransferase proteins being used sequentially during said incubating.

64. The method according to any one of claims 54 to 62, wherein said incubating is carried out with a plurality of different tripartite glycosyltransferase fusion proteins, at least some of the different tripartite glycosyltransferase proteins being used simultaneously during said incubating.

Bacterial Strain	**Genotype**	**Source**
DH5α	F- (Φ80ΔlacZΔM15,) Δ(*lacIZYA-argF*)U169 *recA1 endA1 hsdR17*(rk^-,mk^+) *phoA supE44* λ- *thi-1 gyrA96 relA1*	Lab stock
BL21 (DE3)	F- *ompT gal dcm lon hsdSB*(r_B^- m_B^-) λ(DE3)	Lab stock
SHuffle® T7 Express *lysY*	MiniF *lysY* (CamR) / *fhuA2 lacZ*::T7 gene1 [*lon*] *ompT ahpC gal* λatt::pNEB3-r1-cDsbC (SpecR, $lacI^q$) Δ*trxB sulA11* R(mcr-73::miniTn10–TetS)2 [*dcm*] R(*zgb-210*::Tn10–TetS) *endA1* Δ*gor* Δ(*mcrC-mrr*)114::IS10	NEB
Origami2(DE3) *gmd*::*kan* Δ*waaL*	*E. coli* Origami2 cells carrying marked deletion of *gmd* gene and unmarked deletion of *waaL* gene	Reference 3
Yeast Strain	**Genotype**	**Source**
SBY49	MATa *pep4*Δ ::LEU2 *prb1*Δ ::LEU2 *ura3-52 his3-200 leu2-3,112 lys2-801 suc2-9*	Scott Emr lab
Cell Line		**Source**
HEK293T		ATCC
FreeStyle™ 293-F		Thermo Fisher
Expi293F™ GnTI-		Thermo Fisher
Plasmid	**Description**	**Source**
pET28a(+)	T7 *lac* promoter; Kan^R	Novagen
pET28a-SIMPLEx	MCS flanked by N-terminal *E. coli* MBP lacking its signal peptide (Δ1-26; ΔspMBP) and C-terminal truncated human ApoAI (Δ1-43; ApoAI*) followed by 6xHis tag cloned in pET28a(+); Kan^R	Reference 4
pET28a-SIMPLEx-GFP	superfolder GFP (sfGFP) cloned in MCS of pET28a-SIMPLEx; Kan^R	Reference 4
pJL1-sfGFP	T7 *lac* promoter, cell-free expression vector encoding superfolder GFP (sfGFP) gene; Kan^R	Reference 5
pVITRO1-Trastuzumab-IgG1/κ	EM7 promoter, encoding HER2/neu receptor specific humanized IgG1/κ antibody isotype; $Hygro^R$	Reference 6
pcDNA3	CMV promoter; Amp^R	Laboratory stock
pYS338	CPS promoter, galactose inducible vector; Amp^R	Scott Emr lab
pConYCGmCB	Plasmid pMW07 encoding genes for Man3GlcNAc2 biosynthesis; Cm^R	Reference 2
pTrc-spDsbA-MBP-GCG^{DQNAT}	Plasmid pTrc99Y encoding spDsbA-MBP-GCG^{DQNAT}; Amp^R	Reference 2
pMAF10	Plasmid pMLBAD encoding HA-tagged PglB from Campylobacter jejuni, Tmp^R	Reference 7

FIG. 1A

Plasmid	Description
pET28a-SIMPLEx-Δ25*Hs*FUT1	truncated human FUT1 (Δ1-25) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*FUT2	truncated human FUT2 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ34*Hs*FUT3	truncated human FUT3 (Δ1-34) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ172*Hs*FUT4	truncated human FUT4 (Δ1-172) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ34*Hs*FUT5	truncated human FUT5 (Δ1-34) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ34*Hs*FUT6	truncated human FUT6 (Δ1-34) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ36*Hs*FUT7	truncated human FUT7 (Δ1-36) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ30*Hs*FUT8	truncated human FUT8 (Δ1-30) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ32*Hs*FUT9	truncated human FUT9 (Δ1-32) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ31*Hs*FUT10	truncated human FUT10 (Δ1-31) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ24*Hs*FUT11	truncated human FUT11 (Δ1-24) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ26*Hs*POFUT1	truncated human POFUT1 (Δ1-26) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ34*Hs*ST3Gal1	truncated human ST3Gal1 (Δ1-34) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*ST3Gal3	truncated human ST3Gal3 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ26*Hs*ST3Gal4	truncated human ST3Gal4 (Δ1-26) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ25*Hs*ST3Gal6	truncated human ST3Gal6 (Δ1-25) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*ST6Gal1	human ST6Gal1 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ35*Hs*ST6GalNAc1	truncated human ST3GalNAc1 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*ST6GalNAc2	truncated human ST3GalNAc2 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ27*Hs*ST6GalNAc4	truncated human ST3GalNAc4 (Δ1-27) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ48*Hs*ST8Sia1	truncated human ST8Sia1 (Δ1-48) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ23*Hs*ST8Sia2	truncated human ST8Sia2 (Δ1-23) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ33*Hs*ST8Sia3	truncated human ST8Sia3 (Δ1-33) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ20*Hs*ST8Sia4	truncated human ST8Sia4 (Δ1-20) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*ppGalNAcT1	truncated human ppGalNAcT1 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ24*Hs*ppGalNAcT2	truncated human ppGalNAcT2 (Δ1-24) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ37*Hs*ppGalNAcT3	truncated human ppGalNAcT3 (Δ1-37) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ35*Hs*ppGalNAcT4	truncated human ppGalNAcT4 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R

FIG. 1B

Plasmid	Description
pET28a-SIMPLEx-Δ35*Hs*ppGalNAcT5	truncated human ppGalNAcT5 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*ppGalNAcT6	truncated human ppGalNAcT6 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ29*Hs*ppGalNAcT7	truncated human ppGalNAcT7 (Î”1-29) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ29*Hs*ppGalNAcT8	truncated human ppGalNAcT8 (Δ1-29) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*ppGalNAcT9	truncated human ppGalNAcT9 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ31*Hs*ppGalNAcT10	truncated human ppGalNAcT10 (Δ1-31) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ43*Hs*B3GALNT1	truncated human B3GALNT1 (Δ1-43) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ25*Hs*B4GALNT1	truncated human B4GALNT1 (Δ1-25) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ53*Hs*-A-group	truncated human A-ABO (Δ1-53) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ43*Hs*A4GalT	truncated human A4GalT (Δ1-43) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ26*Hs*B3GalT1	truncated human B3GalT1 (Δ1-26) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ45*Hs*B3GalT2	truncated human B3GalT2 (Δ1-45) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ44*Hs*B4GalT1	truncated human B4GalT1 (Δ1-44) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ36*Hs*B4GalT2	truncated human B4GalT2 (Δ1-36) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ31*Hs*B4GalT3	truncated human B4GalT3 (Δ1-31) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ38*Hs*B4GalT4	truncated human B4GalT4 (Δ1-38) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ35*Hs*B4GalT5	truncated human B4GalT5 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ35*Hs*B4GalT6	truncated human B4GalT6 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ53*Hs*-B-group	truncated human B-ABO (Δ1-53) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ20*Hs*UGT8	human UGT8 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ29*Hs*C1GLT	human C1GLT cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*COSMC	human COSMC cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLExΔ23 *Hs*Alg1	truncated human Alg1 (Δ1-23) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg2	human Alg2 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg3	human Alg3 cloned in MCS of pET28a-SIMPLEx; KanR
pET28a-SIMPLEx-Δ40 *Hs*Alg11	truncated human Alg11 (Δ1-40) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg12	human Alg12 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg13	human Alg13 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ24 *Hs*Alg14	truncated human Alg14 (Δ1-24) cloned in MCS of pET28a-SIMPLEx; Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-SIMPLEx-*Hs*DPM1	human DPM1 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*PIGM	human PIGM cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*PIGB	human PIGB cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*PIGZ	human PIGZ cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28 HsAlg5	truncated human Alg5 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg6	human Alg6 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg8	human Alg8 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Alg10	human Alg10 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*UGCG	human UGCG cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ27 *Hs*B3GLCT	truncated human B3GLCT (Δ1-27) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*GLYG	human Glycogenin cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*POGLUT1	human POGLUT1 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ29*Hs*GnTI/MGAT1	truncated human GnTI (Δ1-29) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ29*Hs*GnTII/MGAT2	truncated human GnTII (Δ1-29) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ23*Hs*GnTIII/MGAT3	truncated human GnTIII (Δ1-23) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ27*Hs*GnTIV/MGAT4	truncated human GnTIV (Δ1-27) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ32*Hs*GCNT1	truncated human GCNT1 (Δ1-32) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ23*Hs*GCNT2	truncated human GCNT2 (Δ1-23) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*B3GNT2	truncated human B3GNT2 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ31*Hs*B3GNT6	truncated human B3GNT6 (Δ1-31) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*PIGA	human PIGA cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ25*Hs*UGT1A1	truncated human UGT1A1 (Δ1-25) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ36*Hs*UGT1A3	truncated human UGT1A3 (Δ1-36) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ36*Hs*B4GAT1	truncated human B4GAT1 (Δ1-36) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ28*Hs*XXLT1	truncated human XXLT1 (Δ1-28) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Cj*CstII	*C. jejuni* CstII cloned in MCS of pET28a-SIMPLEx; Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-SIMPLEx-*Nm*Pst	*N. meningitidis* PolysiaT cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Cj*CgtB	*C. jejuni* CgtB cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hp*LgtB	*H. pylori* GalT cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Nm*LgtB	*N. meningitidis* GalT cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Ng*LgtB	*N. gonorrhea* GalT cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Ec*WbgL	*E. coli* WbgL cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Ec*WecA	*E. coli* WecA cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Lp*SetA	*L. pneumophila* SetA cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Nm*SynE	*N. meningitidis* SynE cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ34*Sc*Alg1	truncated *S. cerevisae* Alg1 (Δ1-34) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Sc*Alg2	*S. cerevisae* Alg2 cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ45*Sc*Alg11	truncated *S. cerevisae* Alg11 (Δ1-45) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ30*Nt*GnTI	truncated *N. tabacum* GnTI (Δ1-30) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ35*Nt*GnTII	truncated *N. tabacum* GnTII (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ22*Bt*GGTA1	truncated *B. taurus* GGTA1 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ60*Mm*GGTA1	truncated *M. musculus* GGTA1 (Δ1-60) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ58*Rn*GGTA1	truncated *R. norvergicus* GGTA1 (Δ1-58) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-Δ44*Bt*B4GalT1	truncated *B. taurus* B4GalT1 (Δ1-35) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*CDK4	human cyclin-dependent kinase 4 (CDK4) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*CDKN2A	human cyclin-dependent kinase inhibitor 2A (CDKN2A) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*EGFR_TK	truncated human EGFR (P694-G1022) cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*FOS	human proto-oncogene c-Fos cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*GATA2	human endothelial transcription factor GATA cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*JUN	human transcription factor JUN cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*MMP1	human interstitial collagenase cloned in MCS of pET28a-SIMPLEx; Kan^R
pET28a-SIMPLEx-*Hs*Proinsulin	human proinsulin cloned in MCS of pET28a-SIMPLEx; Kan^R
pcDNA3-SIMPLEx-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) as SIMPLEx chimera cloned in pcDNA3; Amp^R
pYS338-SIMPLEx-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) as SIMPLEx chimera cloned in pYS338; Amp^R

FIG. 1B (continued)

Plasmid	Description
pJL1-SIMPLEx-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) as SIMPLEx chimera cloned in pJL1; Kan^R
pET28a-Δ25*Hs*FUT1	truncated human FUT1 (Δ1-25) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ28*Hs*FUT2	truncated human FUT2 (Δ1-28) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ34*Hs*FUT3	truncated human FUT3 (Δ1-34) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ172*Hs*FUT4	truncated human FUT4 (Δ1-172) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ34*Hs*FUT5	truncated human FUT5 (Δ1-34) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ34*Hs*FUT6	truncated human FUT6 (Δ1-34) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ36*Hs*FUT7	truncated human FUT7 (Δ1-36) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ30*Hs*FUT8	truncated human FUT8 (Δ1-30) with C-terminal 6xHis tag in pET28a(+);Kan^R
pET28a-Δ32*Hs*FUT9	truncated human FUT9 (Δ1-32) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ31*Hs*FUT10	truncated human FUT10 (Δ1-31) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ24*Hs*FUT11	truncated human FUT11 (Δ1-24) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ26*Hs*POFUT1	truncated human POFUT1 (Δ1-26) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ34*Hs*ST3Gal1	truncated human ST3Gal1 (Δ1-34) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*ST3Gal3	truncated human ST3Gal3 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ26*Hs*ST3Gal4	truncated human ST3Gal4 (Δ1-26) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ25*Hs*ST3Gal6	truncated human ST3Gal6 (Δ1-25) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*ST6Gal1	Human ST6Gal1 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-MBP-Δ26*Hs*ST6Gal1	truncated human ST6Gal1 (Δ1-26) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ26*Hs*ST6Gal1-ApoAI	truncated human ST6Gal1 (Δ1-26) with ApoAI*-6xHis in pET28a(+); Kan^R
pET28a-Δ35*Hs*ST6GalNAc1	truncated human ST6GalNAc1 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*ST6GalNAc2	truncated human ST6GalNAc2 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ27*Hs*ST6GalNAc4	truncated human ST6GalNAc4 (Δ1-27) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ48*Hs*ST8Sia1	truncated human ST8Sia1 (Δ1-48) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ23*Hs*ST8Sia2	truncated human ST8Sia2 (Δ1-23) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ33*Hs*ST8Sia3	truncated human ST8Sia3 (Δ1-33) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ20*Hs*ST8Sia4	truncated human ST8Sia4 (Δ1-20) with C-terminal 6xHis tag in pET28a(+); Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-Δ28*Hs*ppGalNAcT1	truncated human ppGalNAcT1 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ24*Hs*ppGalNAcT2	truncated human ppGalNAcT2 (Δ1-24) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ37*Hs*ppGalNAcT3	truncated human ppGalNAcT3 (Δ1-37) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ35*Hs*ppGalNAcT4	truncated human ppGalNAcT4 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ35*Hs*ppGalNAcT5	truncated human ppGalNAcT5 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*ppGalNAcT6	truncated human ppGalNAcT6 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ29*Hs*ppGalNAcT7	truncated human ppGalNAcT7 (Δ1-29) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ29*Hs*ppGalNAcT8	truncated human ppGalNAcT8 (Δ1-29) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*ppGalNAcT9	truncated human ppGalNAcT9 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ31*Hs*ppGalNAcT10	truncated human ppGalNAcT10 (Δ1-31) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ43*Hs*B3GALNT1	truncated human B3GALNT1 (Δ1-43) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ25*Hs*B4GALNT1	truncated human B4GALNT1 (Δ1-25) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*-A-group	human A-ABO with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ43*Hs*A4GALT	truncated human A4GALT (Δ1-43) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ26*Hs*B3GalT1	truncated human B3GalT1 (Δ1-26) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ45*Hs*B3GalT2	truncated human B3GalT2 (Δ1-45) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ44*Hs*B4GalT1	truncated human B4GalT1 (Δ1-44) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ36*Hs*B4GalT2	truncated human B4GalT2 (Δ1-36) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ31*Hs*B4GalT3	truncated human B4GalT3 (Δ1-31) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ38*Hs*B4GalT4	truncated human B4GalT4 (Δ1-38) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ35*Hs*B4GalT5	truncated human B4GalT5 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ35*Hs*B4GalT6	truncated human B4GalT6 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*-B-group	human B-ABO with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ20*Hs*UGT8	truncated human UGT8 (Δ1-20) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ29*Hs*C1GLT	truncated human C1GLT (Δ1-29) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*COSMC	human COSMC with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ23*Hs*Alg1	truncated human Alg1 (Δ1-23) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg2	human Alg2 with C-terminal 6xHis tag in pET28a(+); Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-*Hs*Alg3	human Alg3 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ40*Hs*Alg11	truncated human Alg11 (Δ1-40) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg12	human Alg12 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg13	human Alg13 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ24*Hs*Alg14	truncated human Alg14 (Δ1-24) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*DPM1	human DPM1 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*PIGM	human PIGM with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*PIGB	human PIGB with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*PIGZ	human PIGZ with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*Alg5	truncated human Alg5 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg6	human Alg6 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg8	human Alg8 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*Alg10	human Alg10 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*UGCG	human UGCG with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ27*Hs*B3GLCT	truncated human B3GLCT (Δ1-27) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*GLYG	human glycogenin with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*POGLUT1	human POGLUT1 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ29*Hs*GnTI/MGAT1	truncated human GnTI (Δ1-29) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ29*Hs*GnTII/MGAT2	truncated human GnTII (Δ1-29) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ23*Hs*GnTIII/MGAT3	truncated human GnTIII (Δ1-23) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ27*Hs*GnTIV/MGAT4	truncated human GnTIV (Δ1-27) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ32*Hs*GCNT1	truncated human GCNT1 (Δ1-32) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ23*Hs*GCNT2	truncated human GCNT2 (Δ1-23) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*B3GNT2	truncated human B3GnT2 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ31*Hs*B3GNT6	truncated human B3GnT6 (Δ1-31) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*PIGA	human PIGA with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ25*Hs*UGT1A1	truncated human UGT1A1 (Δ1-25) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ36*Hs*UGT1A3	truncated human UGT1A3 (Δ1-36) with C-terminal 6xHis tag in pET28a(+); Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-Δ36*Hs*B4GAT1	truncated human B4GAT1 (Δ1-36) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ28*Hs*XXLT1	truncated human XXLT1 (Δ1-28) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Cj*CstII	*C. jejuni* CstII with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Nm*Pst	*N. meningitidis* PolysiaT with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Cj*CgtB	*C. jejuni* CgtB with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hp*LgtB	*H. pylori* GalT with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Nm*LgtB	*N. meningitidis* GalT with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Ng*LgtB	*N. gonorrhea* GalT with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Ec*WbgL	*E. coli* WbgL with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Ec*WecA	*E. coli* WecA with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Lp*SetA	*L. pneumophila* SetA with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Nm*SynE	*N. meningitidis* SynE with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a- Δ34*Sc*Alg1	truncated *S. cerevisae* Alg1 (Δ1-34) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Sc*Alg2	*S. cerevisae* Alg2 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ45*Sc*Alg11	truncated S. cerevisae Alg11 (Δ1-45) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ30*Nt*GnTI	truncated N. tabacum GnTI (Δ1-30) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ35*Nt*GnTII	truncated *N. tabacum* GnTII (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ22*Bt*GGTA1	truncated *B. taurus* GGTA1 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ60*Mm*GGTA1	truncated *M. musculus* GGTA1 (Δ1-60) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ58*Rn*GGTA1	truncated *R. norvergicus* GGTA1 (Δ1-58) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-Δ44*Bt*B4GalT1	truncated *B. taurus* B4GalT1 (Δ1-35) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*CDK4	human CDK4 with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*CDKN2A	human CDKN2A with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*EGFR_TK	truncated human EGFR (P694-G1022) with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*FOS	human proto-oncogene c-Fos with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*GATA2	human endothelial transcription factor GATA with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*JUN	human transcription factor JUN with C-terminal 6xHis tag in pET28a(+); Kan^R
pET28a-*Hs*MMP1	human Interstitial collagenase with C-terminal 6xHis tag in pET28a(+); Kan^R

FIG. 1B (continued)

Plasmid	Description
pET28a-*Hs*Pronsulin	human proinsulin with C-terminal 6xHis tag in pET28a(+); Kan^R
pcDNA3-Δ26*Hs*ST6Gal1	truncated *Hs*ST6Gal1 with C-terminal 6xHis tag in mammalian expression vector pcDNA3; Amp^R
pYS338-Δ26*Hs*ST6Gal1	truncated *Hs*ST6Gal1 with C-terminal 6xHis tag in yeast expression vector pYS338; Amp^R
pJL1-Δ26*Hs*ST6Gal1	truncated *Hs*ST6Gal1 with C-terminal 6xHis tag in cell-free expression vector pJL1; Kan^R

FIG. 1B (continued)

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
HsFUT1	11	HUMAN Galactoside 2-alpha-L-fucosyltransferase 1	P19526	type II membrane protein	BL21(DE3)
HsFUT2	11	HUMAN Galactoside 2-alpha-L-fucosyltransferase 2	Q10981	type II membrane protein	BL21(DE3)
HsFUT3	10	HUMAN Galactoside 3(4)-L-fucosyltransferase	P21217	type II membrane protein	BL21(DE3)
HsFUT4	10	HUMAN Alpha-(1,3)-fucosyltransferase 4	P22083	type II membrane protein	BL21(DE3)
HsFUT5	10	HUMAN Alpha-(1,3)-fucosyltransferase 5	Q11128	type II membrane protein	BL21(DE3)
HsFUT6	10	HUMAN Alpha-(1,3)-fucosyltransferase 6	P51993	type II membrane protein	BL21(DE3)
HsFUT7	10	HUMAN Alpha-(1,3)-fucosyltransferase 7	Q11130	type II membrane protein	SHuffle T7 Express *lysY*
HsFUT8	23	HUMAN Alpha-(1,6)-fucosyltransferase	Q9BYC5	type II membrane protein	SHuffle T7 Express *lysY*
HsFUT9	10	HUMAN Alpha-(1,3)-fucosyltransferase 9	Q9Y231	type II membrane protein	BL21(DE3)
HsFUT10	10	HUMAN Alpha-(1,3)-fucosyltransferase 10	Q6P4F1	type II membrane protein	BL21(DE3)
HsFUT11	10	HUMAN Alpha-(1,3)-fucosyltransferase 11	Q495W5	type II membrane protein	SHuffle T7 Express *lysY*
HsPOFUT1	65	HUMAN GDP-fucose protein *O*-fucosyltransferase 1	Q9H488	secretory protein	SHuffle T7 Express *lysY*
HsST3Gal1	29	HUMAN CMP-*N*-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1	Q11201	type II membrane protein	SHuffle T7 Express *lysY*
HsST3Gal3	29	HUMAN CMP-*N*-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase	Q11203	type II membrane protein	SHuffle T7 Express *lysY*
HsST3Gal4	29	HUMAN CMP-*N*-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4	Q11206	type II membrane protein	SHuffle T7 Express *lysY*
HsST3Gal6	29	HUMAN Type 2 lactosamine alpha-2,3-sialyltransferase	Q9Y274	type II membrane protein	BL21(DE3)
HsST6Gal1	29	HUMAN Beta-galactoside alpha-2,6-sialyltransferase 1	P15907	type II membrane protein	SHuffle T7 Express *lysY*
HsST6GalNAc1	29	HUMAN Alpha-*N*-acetylgalactosaminide alpha-2,6-sialyltransferase 1	Q9NSC7	type II membrane protein	SHuffle T7 Express *lysY*
HsST6GalNAc2	29	HUMAN Alpha-*N*-acetylgalactosaminide alpha-2,6-sialyltransferase 2	Q9UJ37	type II membrane protein	SHuffle T7 Express *lysY*
HsST6GalNAc4	29	HUMAN Alpha-*N*-acetyl-neuraminyl-2,3-beta-galactosyl-1,3-N-acetyl-galactosaminide alpha-2,6-sialyltransferase	Q9H4F1	type II membrane protein	SHuffle T7 Express *lysY*

FIG. 2A

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
HsST8Sia1	29	HUMAN Alpha-*N*-acetylneuraminide alpha-2,8-sialyltransferase	Q92185	type II membrane protein	SHuffle T7 Express *lysY*
HsST8Sia2	29	HUMAN Alpha-2,8-sialyltransferase 8B	Q92186	type II membrane protein	SHuffle T7 Express *lysY*
HsST8Sia3	29	HUMAN Sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase	O43173	type II membrane protein	SHuffle T7 Express *lysY*
HsST8Sia4	29	HUMAN CMP-*N*-acetylneuraminate-poly-alpha-2,8-sialyltransferase	Q92187	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT1	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 1	Q10472	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT2	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 2	Q10471	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT3	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 3	Q14435	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT4	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 4	Q8N4A0	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT5	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 5	Q7Z7M9	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT6	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 6	Q8NCL4	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT7	27	HUMAN N-acetylgalactosaminyltransferase 7	Q86SF2	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT8	27	HUMAN Probable polypeptide *N*-acetylgalactosaminyltransferase 8	Q9NY28	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT9	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 9	Q9HCQ5	type II membrane protein	SHuffle T7 Express *lysY*
HsGalNAcT10	27	HUMAN Polypeptide *N*-acetylgalactosaminyltransferase 10	Q86SR1	type II membrane protein	SHuffle T7 Express *lysY*
HsB3GALNT1	31	HUMAN UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1	O75752	type II membrane protein	BL21(DE3)
HsB4GALNT1	12	HUMAN Beta-1,4 *N*-acetylgalactosaminyltransferase 1	Q00973	type II membrane protein	SHuffle T7 Express *lysY*
HsA-ABO	6	HUMAN Histo-blood group ABO system transferase	P16442	type II membrane protein	BL21(DE3)
HsA4GALT	32	HUMAN Lactosylceramide 4-alpha-galactosyltransferase	Q9NPC4	type II membrane protein	BL21(DE3)
HsB3GalT1	31	HUMAN Beta-1,3-galactosyltransferase 1	Q9Y5Z6	type II membrane protein	BL21(DE3)
HsB3GalT2	31	HUMAN Beta-1,3-galactosyltransferase 2	O43825	type II membrane protein	BL21(DE3)

FIG. 2A (continued)

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
HsB4GalT1	7	HUMAN Beta-1,4-galactosyltransferase 1	P15291	type II membrane protein	SHuffle T7 Express *lysY*
HsB4GalT2	7	HUMAN Beta-1,4-galactosyltransferase 2	O60909	type II membrane protein	SHuffle T7 Express *lysY*
HsB4GalT3	7	HUMAN Beta-1,4-galactosyltransferase 3	O60512	type II membrane protein	SHuffle T7 Express *lysY*
HsB4GalT4	7	HUMAN Beta-1,4-galactosyltransferase 4	O60513	type II membrane protein	SHuffle T7 Express *lysY*
HsB4GalT5	7	HUMAN Beta-1,4-galactosyltransferase 5	O43286	type II membrane protein	SHuffle T7 Express *lysY*
HsB4GalT6	7	HUMAN Beta-1,4-galactosyltransferase 6	Q9UBX8	type II membrane protein	SHuffle T7 Express *lysY*
HsB-ABO	6	HUMAN Histo-blood group ABO system transferase	P16442	type II membrane protein	BL21(DE3)
HsUGT8	1	HUMAN 2-hydroxyacylsphingosine 1-beta-galactosyltransferase	Q16880	type I membrane protein	BL21(DE3)
HsC1GLT	31	HUMAN Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1	Q9NS00	type II membrane protein	BL21(DE3)
HsCOSMC	31	HUMAN C1GALT1-specific chaperone 1	Q96EU7	type II membrane protein	BL21(DE3)
HsAlg1	33	HUMAN Chitobiosyldiphosphodolichol beta-mannosyltransferase	Q9BT22	type II membrane protein	BL21(DE3)
HsAlg2	4	HUMAN Alpha-1,3/1,6-mannosyltransferase	Q9H553	type II membrane protein	BL21(DE3)
HsAlg3	58	HUMAN Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase	Q92685	multi-pass membrane protein	BL21(DE3)
HsAlg11	4	HUMAN GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase	Q2TAA5	type II membrane protein	BL21(DE3)
HsAlg12	22	HUMAN Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase	Q9BV10	multi-pass membrane protein	BL21(DE3)
HsAlg13	1	HUMAN Isoform 2 of Putative UDP-N-acetylglucosamine transferase	Q9NP73	cytosolic protein	BL21(DE3)
HsAlg14	1	HUMAN UDP-N-acetylglucosamine transferase subunit ALG14 homolog	Q96F25	type II membrane protein	BL21(DE3)
HsDPM1	2	HUMAN Dolichol-phosphate mannosyltransferase subunit 1	O60762	cytosolic protein	BL21(DE3)

FIG. 2A (continued)

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
HsPIGM	50	HUMAN GPI mannosyltransferase 1	Q9H3S5	multi-pass membrane protein	BL21(DE3)
HsPIGB	22	HUMAN GPI mannosyltransferase 3	Q92521	multi-pass membrane protein	BL21(DE3)
HsPIGZ	22	HUMAN GPI mannosyltransferase 4	Q86VD9	multi-pass membrane protein	BL21(DE3)
HsAlg5	2	HUMAN Dolichyl-phosphate beta-glucosyltransferase	Q9Y673	type II membrane protein	BL21(DE3)
HsAlg6	57	HUMAN Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase	Q9Y672	multi-pass membrane protein	BL21(DE3)
HsAlg8	57	HUMAN Probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase	Q9BVK2	multi-pass membrane protein	BL21(DE3)
HsAlg10	59	HUMAN Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase	Q5I7T1	multi-pass membrane protein	BL21(DE3)
HsUGCG	21	HUMAN Ceramide glucosyltransferase	Q16739	multi-pass membrane protein	BL21(DE3)
HsB3GLCT	31	HUMAN Beta-1,3-glucosyltransferase	Q6Y288	type II membrane protein	BL21(DE3)
HsGLYG	8	HUMAN Glycogenin-1	P46976	cytosolic protein	BL21(DE3)
HsPOGLUT1	24	HUMAN Protein O-glucosyltransferase 1	Q8NBL1	secretory protein	SHuffle T7 Express *lysY*
HsGnT1	13	HUMAN Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	P26572	type II membrane protein	SHuffle T7 Express *lysY*
HsGnT2	16	HUMAN Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	Q10469	type II membrane protein	SHuffle T7 Express *lysY*
HsGnT3	17	HUMAN Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase	Q09327	type II membrane protein	BL21(DE3)
HsGnT4a	54	HUMAN Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	Q9UM21	type II membrane protein	BL21(DE3)
HsGCNT1	14	HUMAN Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase	Q02742	type II membrane protein	SHuffle T7 Express *lysY*

FIG. 2A (continued)

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
HsGCNT2	14	HUMAN N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase	Q8N0V5	type II membrane protein	BL21(DE3)
HsB3GNT2	31	HUMAN N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2	Q9NY97	type II membrane protein	BL21(DE3)
HsB3GNT6	31	HUMAN Acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase	Q6ZMB0	type II membrane protein	BL21(DE3)
HsPIGA	4	HUMAN Phosphatidylinositol N-acetylglucosaminyltransferase subunit A	P37287	type II membrane protein	BL21(DE3)
HsXXLT1	14	HUMAN Xyloside xylosyltransferase 1	Q8NBI6	type II membrane protein	SHuffle T7 Express *lysY*
HsUGT1A1	1	HUMAN UDP-glucuronosyltransferase 1-1	P22309	type I membrane protein	BL21(DE3)
HsB4GAT1	49	HUMAN Beta-1,4-glucuronyltransferase 1	O43505	type II membrane protein	BL21(DE3)
HsUGT1A3	1	HUMAN UDP-glucuronosyltransferase 1-3	P35503	type II membrane protein	BL21(DE3)
CjCsTII	42	*Campylobacter jejuni* CsTII	A3EYK2	cytosolic protein	BL21(DE3)
NmPolysiaT	38	*Neisseria meningitidis* Polysialic acid *O*-acetyltransferase	A1KR99	cytosolic protein	BL21(DE3)
CjCgtB	2	*Campylobacter jejuni* Beta-1,3-galactosyltransferase	Q0GH19	cytosolic protein	BL21(DE3)
HpLgtB	25	*Helicobacter pylori* (strain 51)Beta-4-galactosyltransferase	D0IS16	cytosolic protein	BL21(DE3)
NmLgtB	25	*Neisseria meningitidis* serogroup B (strain MC58) Lacto-N-neotetraose biosynthesis glycosyltransferase LgtB	Q51116	cytosolic protein	BL21(DE3)
NgLgtB	25	*Neisseria gonorrhoeae* Lacto-N-neotetraose biosynthesis glycosyltransferase	Q50947	cytosolic protein	BL21(DE3)
EcFUT	11	*E.coli* galactoside 2-alpha-L-fucosyltransferase WbgL	E2DNL9	cytosolic protein	BL21(DE3)
EcWecA	4	*E.coli* Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	P0AC78	multi-pass membrane protein	BL21(DE3)
LpSetA	32	*Legionella pneumophila* subsp. Pneumophila Subversion of eukaryotic traffic protein A	Q5ZU30	cytosolic protein	BL21(DE3)
NmSynE	38	*Neisseria meningitidis* Alpha-2,9-polysialyltransferase	O06435	cytosolic protein	BL21(DE3)

FIG. 2A (continued)

Name	GTs family	Description	Uniprot ID*	Protein structure / topology	Host *E. coli* strain
ScAlg1	33	YEAST Beta-1,4-mannosyltransferase OS=Saccharomyces cerevisiae	P16661	multi-pass membrane protein	BL21(DE3)
ScAlg11	4	YEAST GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase	P53954	multi-pass membrane protein	BL21(DE3)
NtGnTI	13	*Nicotiana tabacum* alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	Q8VXD3	type II membrane protein	SHuffle T7 Express *lysY*
NtGnTII	16	*Nicotiana tabacum* alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase-like	A0A1S3XZI5	type II membrane protein	SHuffle T7 Express *lysY*
BtGGTA1	6	*Bos taurus* N-acetyllactosaminide alpha-1,3-galactosyltransferase	P14769	type II membrane protein	BL21(DE3)
MmGGTA1	6	MOUSE N-acetyllactosaminide alpha-1,3-galactosyltransferase	P23336	type II membrane protein	BL21(DE3)
RnGGTA1	6	RAT N-acetyllactosaminide alpha-1,3-galactosyltransferase	Q3L7M0	type II membrane protein	BL21(DE3)
BtB4GalT1	7	*Bos taurus* Beta-1,4-galactosyltransferase 1	P08037	type II membrane protein	SHuffle T7 Express *lysY*

FIG. 2A (continued)

Name	3D structure availability (UniPROT)	SIMPLEx express-ion score	Unfused express-ion score	pI - full length (ExPASy)	MW - full length (ExPASy)	Solubility prediction - full length	pI – truncated (ExPASy)	MW - truncated (ExPASy)	Solubility prediction - truncated	MW - SIMPLEx	pI - SIMPLEx (ExPASy)	Solubility prediction - SIMPLEx
HsFUT1	N	2	2	6.92	41.25	-1.585633	6.49	38.39	-0.642547	103.853	5.82	-0.111598
HsFUT2	N	2	1	8.69	39.02	-1.908709	8.68	35.75	-0.738752	101.213	6.07	-0.114397
HsFUT3	N	1	1	9.16	42.12	-0.643519	9.11	38.34	0.103183	103.803	6.23	0.369262
HsFUT4	N	1	0	10.04	59.08	0.050703	9.01	40.89	-0.081963	106.353	6.17	0.231434
HsFUT5	N	3	1	8.54	43.01	-0.816561	8.16	39.2	-0.112841	104.663	5.93	0.23094
HsFUT6	N	3	1	8.92	41.86	-0.816902	8.98	38.03	0.085867	103.493	6.16	0.364175
HsFUT7	N	2	0	10.07	39.24	-0.698032	9.75	35.72	-0.229632	101.183	6.52	0.208909
HsFUT8	Y	1	0	7.36	66.52	0.035308	6.92	63.04	0.70488	128.503	6.08	0.633266
HsFUT9	N	1	0	7.6	42.07	-1.951914	6.67	38.66	-1.13295	104.123	5.86	-0.342657
HsFUT10	N	2	1	8.62	56.1	-0.479454	8.25	52.71	0.290943	118.173	6.09	0.413479
HsFUT11	N	1	1	5.59	55.82	0.853536	5.48	53.63	1.28867	119.093	5.49	1.034772
HsPOFUT1	Y	1	1	8.65	41.28	-0.395634	8.65	41.41	-0.083866	106.873	6.19	0.203762
HsST3Gal1	N	1	1	9.21	39.08	-0.969125	8.81	35.17	0.069292	100.633	6.15	0.387008
HsST3Gal3	N	2	1	9.1	42.17	-0.527365	9.07	39.07	0.563796	104.533	6.08	0.638151
HsST3Gal4	N	2	1	9.45	38.05	-0.380776	9.3	35.17	0.404008	100.633	6.34	0.49835
HsST3Gal6	N	3	1	9.13	38.21	-1.316155	9.12	35.41	-0.220679	100.873	6.35	0.214156
HsST6Gal1	Y	2	0	9.13	46.6	-0.016537	9.08	43.7	0.953037	109.163	6.38	0.841736
HsST6GalNAc1	N	2	0	9.93	68.56	0.335124	9.9	64.55	1.038552	130.013	8.88	0.861091
HsST6GalNAc2	Y	2	2	9.38	41.94	-1.039498	9.38	38.97	-0.275925	104.433	6.46	0.132963
HsST6GalNAc4	N	2	0	8.9	34.2	-1.014276	8.84	31.34	0.213271	96.803	6.18	0.477558
HsST8Sia1	N	1	2	9.41	40.52	-1.085403	9.07	35.3	-0.230408	100.763	6.22	0.188328
HsST8Sia2	N	1	1	9.48	42.43	-0.45256	9.42	39.87	0.344699	105.333	6.61	0.48239

FIG. 2B

Name	3D structure availability (UniPROT)	SIMPLEx express-ion score	Unfused express-ion score	pI - full length (ExPASy)	MW - full length (ExPASy)	Solubility prediction - full length	pI – truncated (ExPASy)	MW - truncated (ExPASy)	Solubility prediction - truncated	MW - SIMPLEx	pI - SIMPLEx (ExPASy)	Solubility prediction - SIMPLEx
HsST8Sia3	Y	1	1	9.56	43.97	-1.099985	9.48	40.55	-0.427266	106.013	6.8	0.029342
HsST8Sia4	Y	1	1	9.77	41.29	0.328265	9.65	38.93	0.731562	104.393	7.19	0.723531
HsGalNAcT1	N	2	0	7.83	64.22	-0.126092	7.2	60.97	0.563126	126.433	6.01	0.550162
HsGalNAcT2	Y	2	0	8.63	64.73	0.230014	8.15	61.82	0.794765	127.283	6.12	0.735135
HsGalNAcT3	N	2	0	8.2	72.61	-0.105005	6.52	68.18	0.54999	133.643	5.91	0.526992
HsGalNAcT4	Y	1	0	7.55	66.67	-0.081421	7.21	62.87	0.691811	128.333	6.02	0.648077
HsGalNAcT5	N	0	0	9.51	106.27	0.485134	9.41	102.3	0.947823	167.763	9.13	1.249963
HsGalNAcT6	N	3	0	8.47	71.16	-0.322353	7.58	67.9	0.096998	133.363	6.14	0.243376
HsGalNAcT7	N	2	0	6.67	75.39	0.27012	6.34	72.32	0.788025	137.783	5.87	0.666059
HsGalNAcT8	N	1	0	9.02	72.85	-0.294557	8.86	69.71	0.160827	135.173	6.58	0.262351
HsGalNAcT9	N	1	0	8.73	68.36	0.088059	8.33	65.21	0.864852	130.673	6.11	0.757396
HsGalNAcT10	Y	1	0		68.99	0.710177	8.53	65.62	1.060668	131.083	6.31	0.879597
HsB3GALNT1	N	3	0	7.67	39.51	-2.454287	6.63	34.59	-1.557329	100.053	5.83	-0.565565
HsB4GALNT1	N	2	2	8.93	58.88	-0.083675	8.8	56.34	0.328105	121.803	6.12	0.41605
HsA-ABO	Y	2	1	9.14	40.93	-0.914732	8.79	35.2	-0.160723	100.663	6.08	0.224711
HsA4GALT	N	1	0	9.16	40.5	-0.960789	8.63	35.54	0.147262	101.003	6.05	0.38114
HsB3GalT1	N	1	1	9.38	37.99	-0.586535	9.44	35.23	0.343966	100.693	6.66	0.524214
HsB3GalT2	N	2	0	9.5	49.21	-0.976967	9.25	43.79	-0.087818	109.253	6.62	0.202552
HsB4GalT1	Y	2	2	8.88	43.92	-0.537125	8.51	39.43	0.012298	104.893	5.95	0.289415
HsB4GalT2	N	1	2	9.47	41.97	-0.285653	9.57	38.07	0.70733	103.533	6.5	0.718469
HsB4GalT3	N	2	1	9.34	43.93	0.965237	9.23	40.62	1.370679	106.083	6.32	1.103855
HsB4GalT4	N	1	0	9.18	40.04	-0.782721	9.1	35.87	0.581062	101.333	6.31	0.698749

FIG. 2B (continued)

Name	3D structure availability (UniPROT)	SIMPLEx express-ion score	Unfused express-ion score	pI - full length (ExPASy)	MW - full length (ExPASy)	Solubility prediction - full length	pI – truncated (ExPASy)	MW - truncated (ExPASy)	Solubility prediction - truncated	MW - SIMPLEx	pI - SIMPLEx (ExPASy)	Solubility prediction - SIMPLEx
HsB4GalT5	N	2	0	8.21	45.12	-0.730384	6.35	41.18	-0.111817	106.643	5.78	0.209056
HsB4GalT6	N	1	0	8.59	44.91	-1.006197	7.22	40.91	-0.121076	106.373	5.95	0.202678
HsB-ABO	Y	2	2	9.14	40.93	-0.914732	8.58	35.17	-0.227259	100.633	6.03	0.186706
HsUGT8	N	1	0	9.51	59.2	-1.887486	9.51	59.33	-1.610519	124.793	7.73	-0.910457
HsC1GLT	N	2	0	6.17	42.2	0.168209	6.07	39.15	0.909628	104.613	5.71	0.829515
HsCOSMC	N	2	0	6.35	36.38	-0.0083	6.35	36.38	-0.0083	101.843	5.79	0.29904
HsAlg1	N	2	2	6.8	52.52	-2.089866	6.84	50.27	-1.543528	115.733	5.98	-0.7668
HsAlg2	N	2	1	6.61	47.09	-0.6567	6.61	47.09	-0.6567	112.553	5.83	-0.19643
HsAlg3	N	1	1	9.56	50.13	-6.091757	9.56	50.13	-6.091757	115.593	7.62	-3.70106
HsAlg11	N	2	1	8.78	55.65	-2.860086	8.78	51.23	-1.755625	116.693	6.28	-0.929947
HsAlg12	N	2	0	9.63	54.65	-5.057991	9.63	54.65	-5.057991	120.113	7.22	-3.168034
HsAlg13	N	2	3	6.07	18.26	0.256384	6.07	18.26	0.256384	83.723	5.59	0.642243
HsAlg14	N	2	0	9.13	24. 15	-2.004067	9.06	21.8	-0.693359	87.263	5.99	0.091865
HsDPM1	N	2	0	9.57	29.5	0.560607	9.57	29.5	0.560607	94.963	6.27	0.715614
HsPIGM	N	2	0	9.26	49.46	-6.855026	9.26	49.46	-6.855026	114.923	6.82	-4.126503
HsPIGB	N	0	0	9.39	65.06	-5.582152	9.39	65.06	-5.582152	130.523	8.13	-3.745828
HsPIGZ	N	0	0	8.66	63.47	-3.827019	8.66	63.47	-3.827019	128.933	6.26	-2.492427
HsAlg5	N	1	1	9.34	39.95	-0.863591	9.34	34.36	-0.109992	99.823	6.3	0.263148
HsAlg6	N	3	0	8.8	58.12	-7.710645	8.8	58.12	-7.710645	123.583	6.5	-4.991913
HsAlg8	N	1	0	9.28	60.09	-6.310223	9.28	60.09	-6.310223	125.553	7.38	-4.092762
HsAlg10	N	1	0	9.33	55.45	-8.277015	9.33	55.45	-8.277015	120.913	7.57	-5.244245
HsUGCG	N	1	1	7.94	44.85	-2.827262	7.94	44.85	-2.827262	110.313	5.91	-1.510313

FIG. 2B (continued)

Name	3D structure availability (UniPROT)	SIMPLEx express-ion score	Unfused express-ion score	pI - full length (ExPASy)	MW - full length (ExPASy)	Solubility prediction - full length	pI – truncated (ExPASy)	MW - truncated (ExPASy)	Solubility prediction - truncated	MW - SIMPLEx	pI - SIMPLEx (ExPASy)	Solubility prediction - SIMPLEx
HsB3GLCT	N	3	3	7.23	56.56	-0.503064	6.97	53.79	-0.103251	119.253	5.98	0.118059
HsGLYG	Y	3	2	5.27	39.25	-1.020556	5.27	39.25	-1.020556	104.713	5.42	-0.320201
HsPOGLUT1	Y	3	1	8.91	43.49	0.041721	8.91	43.49	0.041721	108.953	6.28	0.287746
HsGnT1	N	3	0	9.25	50.88	-0.175912	9.09	47.68	0.669434	113.143	6.2	0.658152
HsGnT2	Y	1	0	8.99	51.55	-1.099039	8.52	48.2	-0.614427	113.663	6.19	-0.17129
HsGnT3	N	1	1	8.54	61.31	0.077471	7.65	58.65	0.56133	124.113	6.05	0.561896
HsGnT4a	N	2	0	6.94	61.54	-1.004912	6.56	58.57	-0.441541	124.033	5.88	-0.113021
HsGCNT1	N	0	0	8.65	49.8	-0.779419	7.6	46.01	-0.097932	111.473	6	0.186179
HsGCNT2	N	3	0	6.73	45.87	-0.948682	6.54	43.43	-0.311447	108.893	5.85	0.077693
HsB3GNT2	N	2	0	8.75	46.02	-0.070494	8.31	42.88	0.439267	108.343	6.1	0.536668
HsB3GNT6	N	3	0	7.65	42.75	0.504054	6.71	39.41	1.133293	104.873	5.89	0.958508
HsPIGA	N	3	0	8.63	54.13	-1.558742	8.63	54.13	-1.558742	119.593	6.33	-0.839692
HsXXLT1	N	1	1	8.35	43.81	-0.438954	6.98	39.43	0.142702	104.893	5.97	0.367988
HsUGT1A1	N	1	0	8.19	59.59	-1.911451	8.22	57.23	-1.643418	122.693	6.22	-0.919632
HsB4GAT1	N	2	1	6.77	47.12	-0.401231	6.52	43.2	0.507867	108.663	5.78	0.564854
HsUGT1A3	Y	1	1	8.41	60.34	-2.033895	8.54	57.48	-1.78952	122.943	6.35	-1.011897
CjCsTII	Y	1	1	8.81	34.58	0.19536	8.81	34.58	0.19536	100.043	6.15	0.440865
NmPolysiaT	N	2	2	9.52	57.98	-1.049253	9.52	57.98	-1.049253	123.443	8.41	-0.517116
CjCgtB	N	1	0	9.09	35.63	-0.759507	9.09	35.63	-0.759507	101.093	6.38	-0.110674
HpLgtB	N	2	0	8.75	31.89	-0.265663	8.75	31.89	-0.265663	97.353	6.19	0.247406
NmLgtB	N	2	0	6.18	31.58	0.977834	6.18	31.58	0.977834	97.043	5.72	0.931291
NgLgtB	N	2	2	6.27	31.78	1.102552	6.27	31.78	1.102552	97.243	5.74	0.994255

FIG. 2B (continued)

Name	3D structure availability (UniPROT)	SIMPLEx express-ion score	Unfused express-ion score	pI - full length (ExPASy)	MW - full length (ExPASy)	Solubility prediction - full length	pI – truncated (ExPASy)	MW - truncated (ExPASy)	Solubility prediction - truncated	MW - SIMPLEx	pI - SIMPLEx (ExPASy)	Solubility prediction - SIMPLEx
EcFUT	N	2	0	6.86	35.03	-0.657981	6.86	35.03	-0.657981	100.493	5.87	-0.034211
Ec*Wec*A	Y	2	0	10.01	40.96	-5.359558	10.01	40.96	-5.359558	106.423	7.96	-2.975737
LpSetA	N	2	0	5.24	73.13	1.044979	5.24	73.13	1.044979	138.593	5.36	0.878118
NmSynE	N	3	1	9.39	57.55	-1.267268	9.39	57.55	-1.267268	123.013	7.41	-0.629544
ScAlg1	Y	1	1	8.93	51.87	-2.025257	8.95	48.07	-1.024253	113.533	6.36	-0.430616
ScAlg11	Y	1	0	9.08	63.14	-1.256221	8.95	57.92	-0.561013	123.383	6.37	-0.193683
NtGnTI	N	1	0	6.29	52.1	0.048512	5.93	48.75	0.571012	114.213	5.64	0.601668
NtGnTII	N	2	1	8.36	49.22	-0.180279	7.01	45.58	0.3069	111.043	6.02	0.435935
BtGGTA1	Y	2	0	6.09	43.25	-0.420071	5.87	40.93	0.374534	106.393	5.63	0.496392
MmGGTA1	Y	2	2	6.08	46.48	-0.851121	5.73	39.85	0.009131	105.313	5.68	0.27429
RnGGTA1	Y	0	2	6.02	43.54	-0.743716	5.69	37.08	0.053175	102.543	5.57	0.31468
BtB4GalT1	Y	2	2	9.38	44.84	-0.196036	9.31	40.37	0.396673	105.833	6.39	0.514391

FIG. 2B (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsFUT1	MWLRSHRQLCLAFLLVCVLSVIFFLHIHQDSFPHGLGLSILCPDRRLVTPPVAIFCLPGTAMGPNASSSCPQHPASLSGTWTVYPNG RFGNQMGQYATLLALAQLNGRRAFILPAMHAALAPVFRITLPVLAPEVDSRTPWRELQLHDWMSEEYADLRDPFLKLSGFPCSWTF FHHLREQIRREFTLHDHLREEAQSVLGQLRLGRTGDRPRTFVGVHVRRGDYLQVMPQRWKGVVGDSAYLRQAMDWFRARHEAP VFVVTSNGMEWCKENIDTSQGDVTFAGDGQEATPWKDFALLTQCNHTIMTIGTFGFWAAYLAGGDTVYLANFTLPDSEFLKIFKPE AAFLPEWVGINADLSPLWTLAKP	1
HsFUT2	MLVVQMPFSFPMAHFILFVFTVSTIFHVQQRLAKIQAMWELPVQIPVLASTSKALGPSQLRGMWTINAIGRLGNQMGEYATLYALAK MNGRPAFIPAQMHSTLAPIFRITLPVLHSATASRIPWQNYHLNDWMEEEYRHIPGEYVRFTGYPCSWTFYHHLRQEILQEFTLHDHV REEAQKFLRGLQVNGSRPGTFVGVHVRRGDYVHVMPKVWKGVVADRRYLQQALDWFRARYSSLIFVVTSNGMAWCRENIDTSH GDVVFAGDGIEGSPAKDFALLTQCNHTIMTIGTFGIWAAYLTGGDTIYLANYTLPDSPFLKIFKPEAAFLPEWTGIAADLSPLLKH	2
HsFUT3	MDPLGAAKPQWPWRRCLAALLFQLLVAVCFFSYLRVSRDDATGSPRAPSGSSRQDTTPTRPTLLILLWTWPFHIPVALSRCSEMVP GTADCHITADRKVYPQADTVIVHHWDIMSNPKSRLPPSPRPQGQRWIWFNLEPPPNCQHLEALDRYFNLTMSYRSDSDIFTPYGWL EPWSGQPAHPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHKPLPKGTMMETLSRYKFYLAFENSLHPDYIT EKLWRNALEAWAVPVVLGPSRSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALDFCKACW KLQQESRYQTVRSIAAWFT	3
HsFUT4	MRRLWGAARKPSGAGWEKEWAEAPQEAPGAWSGRLGPGRSGRKGRAVPGWASWPAHLALAARPARHLGGAGQGPRPLHSGT APFHSRASGERQRRLEPQLQHESRCRSSTPADAWRAEAALPVRAMGAPWGSPTAAAGGRRGWRRGRGLPWTVCVLAAAGLTC TALITYACWGQLPPLPWASPTPSRPVGVLLWWEPFGGRDSAPRPPPDCRLRFNISGCRLLTDRASYGEAQAVLFHHRDLVKGPPD WPPPWGIQAHTAEEVDLRVLDYEEAAAAAEALATSSPRPPGQRWVWMNFESPSHSPGLRSLASNLFNWTLSYRADSDVFVPYGY LYPRSHPGDPPSGLAPPLSRKQGLVAWVVSHWDERQARVRYYHQLSQHVTVDVFGRGGPGQPVPEIGLLHTVARYKFYLAFENS QHLDYITEKLWRNALLAGAVPVVLGPDRANYERFVPRGAFIHVDDFPSASSLASYLLFLDRNPAVYRRYFHWRRSYAVHITSFWDE PWCRVCQAVQRAGDRPKSIRNLASWFER	4
HsFUT5	MDPLGPAKPQWLWRRCLAGLLFQLLVAVCFFSYLRVSRDDATGSPRPGLMAVEPVTGAPNGSRCQDSMATPAHPTLLILLWTWPF NTPVALPRCSEMVPGAADCNITADSSVYPQADAVIVHHWDIMYNPSANLPPPTRPQGQRWIWFSMESPSNCRHLEALDGYFNLTM SYRSDSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHKPLPKGTMMETLSRY KFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLR PRSFSWALAFCKACWKLQQESRYQTVRSIAAWFT	5
HsFUT6	MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVYPNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPG TADCNITADRKVYPQADAVIVHHREVMYNPSAQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRSDSDIFTPYGW LEPWSGQPAHPPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKVDVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYI	6

FIG. 2C

Name	Sequence – full length (FASTA)	SEQ ID NO.
	TEKLWRNALEAWAVPVVLGPSRSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALAFCKAC WKLQEESRYQTRGIAAWFT	
HsFUT7	MNNAGHGPTRRLRGLGVLAGVALLAALWLLWLLGSAPRGTPAPQPTITILVWHWPFTDQPPELPSDTCTRYGIARCHLSANRSLLA SADAVVFHHRELQTRRSHLPLAQRPRGQPWVWASMESPSHTHGLSHLRGIFNWVLSYRRDSDIFVPYGRLEPHWGPSPPLPAKS RVAAWVVSNFQERQLRARLYRQLAPHLRVDVFGRANGRPLCASCLVPTVAQYRFYLSFENSQHRDYITEKFWRNALVAGTVPVVL GPPRATYEAFVPADAFVHVDDFGSARELAAFLTGMNESRYQRFFAWRDRLRVRLFTDWRERFCAICDRYPHLPRSQVYEDLEGW FQA	7
HsFUT8	MRPWTGSWRWIMLILFAWGTLLFYIGGHLVRDNDHPDHSSRELSKILAKLERLKQQNEDLRRMAESLRIPEGPIDQGPAIGRVRVLE EQLVKAKEQIENYKKQTRNGLGKDHEILRRRIENGAKELWFFLQSELKKLKNLEGNELQRHADEFLLDLGHHERSIMTDLYYLSQTD GAGDWREKEAKDLTELVQRRITYLQNPKDCSKAKKLVCNINKGCGYGCQLHHVVYCFMIAYGTQRTLILESQNWRYATGGWETVF RPVSETCTDRSGISTGHWSGEVKDKNVQVVELPIVDSLHPRPPYLPLAVPEDLADRLVRVHGDPAVWWVSQFVKYLIRPQPWLEK EIEEATKKLGFKHPVIGVHVRRTDKVGTEAAFHPIEEYMVHVEEHFQLLARRMQVDKKRVYLATDDPSLLKEAKTKYPNYEFISDNSI SWSAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLDDIYYFGGQNAHNQIAIYAHQPRTAD EIPMEPGDIIGVAGNHWDGYSKGVNRKLGRTGLYPSYKVREKIETVKYPTYPEAEK	8
HsFUT9	MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESASSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQG CHLTTDRSLYNKSHAVLIHHRDISWDLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRRDSDIQVPYGFLTVSTNP FVFEVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTYGQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSV PVVLGPSRENYENYIPADSFIHVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHACLACDHVKRHQEYKSVGN LEKWFWN	9
HsFUT10	MVRIQRRKLLASCLCVTATVFLLVTLQVMVELGKFERKEFKSSSLQDGHTKMEEAPTHLNSFLKKEGLTFNRKRKWELDSYPIMLW WSPLTGETGRLGQCGADACFFTINRTYLHHHMTKAFLFYGTDFNIDSLPLPRKAHHDWAVFHEESPKNNYKLFHKPVITLFNYTATF SRHSHLPLTTQYLESIEVLKSLRYLVPLQSKNKLRKRLAPLVYVQSDCDPPSDRDSYVRELMTYIEVDSYGECLRNKDLPQQLKNPA SMDADGFYRIIAQYKFILAFENAVCDDYITEKFWRPLKLGVVPVYYGSPSITDWLPSNKSAILVSEFSHPRELASYIRRLDSDDRLYEA YVEWKLKGEISNQRLLTALRERKWGVQDVNQDNYIDAFECMVCTKVWANIRLQEKGLPPKRWEAEDTHLSCPEPTVFAFSPLRTP PLSSLREMWISSFEQSKKEAQALRWLVDRNQNFSSQEFWGLVFKD	10
HsFUT11	MAAGPIRVVLVLLGVLSVCAASGHGSVAEREAGGEAEWAEPWDGAVFRPPSALGAVGVTRSSGTPRPGREEAGDLPVLLWWSP GLFPHFPGDSERIECARGACVASRNRRALRDSRTRALLFYGTDFRASAAPLPRLAHQSWALLHEESPLNNFLLSHGPGIRLFNLTST FSRHSDYPLSLQWLPGTAYLRRPVPPPMERAEWRRRGYAPLLYLQSHCDVPADRDRYVRELMRHIPVDSYGKCLQNRELPTARL QDTATATTEDPELLAFLSRYKFHLALENAICNDYMTEKLWRPMHLGAVPVYRGSPSVRDWMPNNHSVILIDDFESPQKLAEFIDFLD	11

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	KNDEEYMKYLAYKQPGGITNQFLLDSLKHREWGVNDPLLPNYLNGFECFVCDYELARLDAEKAHAASPGDSPVFEPHIAQPSHMD CPVPTPGFGNVEEIPENDSWKEMWLQDYWQGLDQGEALTAMIHNNETEQTKFWDYLHEIFMKRQHL	
HsPOFUT1	MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV SYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFS ASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGT AGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDL YILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF	12
HsST3Gal1	MVTLRKRTLKVLTFLVLFIFLTSFFLNYSHTMVATTWFPKQMVLELSENLKRLIKHRPCTCTHCIGQRKLSAWFDERFNQTMQPLLTA QNALLEDDTYRWWLRLQREKKPNNLNDTIKELFRVVPGNVDPMLEKRSVGCRRCAVVGNSGNLRESSYGPEIDSHDFVLRMNKAP TAGFEADVGTKTTHHLVYPESFRELGDNVSMILVPFKTIDLEWVVSAITTGTISHTYIPVPAKIRVKQDKILIYHPAFIKYVFDNWLQGH GRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHYWENNPSAGAFRKTGVHDADFESNVTATLASINKIRIFKGR	13
HsST3Gal3	MGLLVFVRNLLLALCLFLVLGFLYYSAWKLHLLQWEEDSNSVVLSFDSAGQTLGSEYDRLGFLLNLDSKLPAELATKYANFSEGACK PGYASALMTAIFPRFSKPAPMFLDDSFRKWARIREFVPPFGIKGQDNLIKAILSVTKEYRLTPALDSLRCRRCIIVGNGGVLANKSLGS RIDDYDIVVRLNSAPVKGFEKDVGSKTTLRITYPEGAMQRPEQYERDSLFVLAGFKWQDFKWLKYIVYKERVSASDGFWKSVATRV PKEPPEIRILNPYFIQEAAFTLIGLPFNNGLMGRGNIPTLGSVAVTMALHGCDEVAVAGFGYDMSTPNAPLHYYETVRMAAIKESWTH NIQREKEFLRKLVKARVITDLSSGI	14
HsST3Gal4	MVSKSRWKLLAMLALVLVVMVWYSISREDRYIELFYFPIPEKKEPCLQGEAESKASKLFGNYSRDQPIFLRLEDYFWVKTPSAYELP YGTKGSEDLLLRVLAITSSSIPKNIQSLRCRRCVVVGNGHRLRNSSLGDAINKYDVVIRLNNAPVAGYEGDVGSKTTMRLFYPESAHF DPKVENNPDTLLVLVAFKAMDFHWIETILSDKKRVRKGFWKQPPLIWDVNPKQIRILNPFFMEIAADKLLSLPMQQPRKIKQKPTTGL LAITLALHLCDLVHIAGFGYPDAYNKKQTIHYYEQITLKSMAGSGHNVSQEALAIKRMLEMGAIKNLTSF	15
HsST3Gal6	MRGYLVAIFLSAVFLYYVLHCILWGTNVYWVAPVEMKRRNKIQPCLSKPAFASLLRFHQFHPFLCAADFRKIASLYGSDKFDLPYGM RTSAEYFRLALSKLQSCDLFDEFDNIPCKKCVVVGNGGVLKNKTLGEKIDSYDVIIRMNNGPVLGHEEEVGRRTTFRLFYPESVFSD PIHNDPNTTVILTAFKPHDLRWLLELLMGDKINTNGFWKKPALNLIYKPYQIRILDPFIIRTAAYELLHFPKVFPKNQKPKHPTTGIIAITL AFYICHEVHLAGFKYNFSDLKSPLHYYGNATMSLMNKNAYHNVTAEQLFLKDIIEKNLVINLTQD	16
HsST6Gal1	MIHTNLKKKFSCCVLVFLLFAVICVWKEKKKGSYYDSFKLQTKEFQVLKSLGKLAMGSDSQSVSSSSTQDPHRGRQTLGSLRGLAK AKPEASFQVWNKDSSSKNLIPRLQKIWKNYLSMNKYKVSYKGPGPGIKFSAEALRCHLRDHVNVSMVEVTDFPFNTSEWEGYLPK ESIRTKAGPWGRCAVVSSAGSLKSSQLGREIDDHDAVLRFNGAPTANFQQDVGTKTTIRLMNSQLVTTEKRFLKDSLYNEGILIVWD PSVYHSDIPKWYQNPDYNFFNNYKTYRKLHPNQPFYILKPQMPWELWDILQEISPEEIQPNPPSSGMLGIIIMMTLCDQVDIYEFLPS KRKTDVCYYYQKFFDSACTMGAYHPLLYEKNLVKHLNQGTDEDIYLLGKATLPGFRTIHC	17

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsST6GalNAc1	MRSCLWRCRHLSQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTIYAEPVPENNALN TQTQPKAHTTGDRGKEANQAPPEEQDKVPHTAQRAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGNGG QTRKLTASRTVSEKHQGKAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIPPKEKKPQATPPPAPFQSPTTQRNQRLKAAN FKSEPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKLFLPNLTLFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLVQKVVTRF PPVPQQQLLLASLPAGSLRCITCAVVGNGGILNNSHMGQEIDSHDYVFRLSGALIKGYEQDVGTRTSFYGFTAFSLTQSLLILGNRGF KNVPLGKDVRYLHFLEGTRDYEWLEALLMNQTVMSKNLFWFRHRPQEAFREALHMDRYLLLHPDFLRYMKNRFLRSKTLDGAHW RIYRPTTGALLLLTALQLCDQVSAYGFITEGHERFSDHYYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN	18
HsST6GalNAc2	MGLPRGSFFWLLLLLTAACSGLLFALYFSAVQRYPGPAAGARDTTSFEAFFQSKASNSWTGKGQACRHLLHLAIQRHPHFRGLFNL SIPVLLWGDLFTPALWDRLSQHKAPYGWRGLSHQVIASTLSLLNGSESAKLFAPPRDTPPKCIRCAVVGNGGILNGSRQGPNIDAHD YVFRLNGAVIKGFERDVGTKTSFYGFTVNTMKNSLVSYWNLGFTSVPQGQDLQYIFIPSDIRDYVMLRSAILGVPVPEGLDKGDRPH AYFGPEASASKFKLLHPDFISYLTERFLKSKLINTHFGDLYMPSTGALMLLTALHTCDQVSAYGFITSNYWKFSDHYFERKMKPLIFYA NHDLSLEAALWRDLHKAGILQLYQR	19
HsST6GalNAc4	MKAPGRLVLIILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPTGSRPTVPGPLHFSGYSSVPDGKPLVREPCRSCAVVSSSGQMLG SGLGAEIDSAECVFRMNQAPTVGFEADVGQRSTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRVLGGRTYRTLLQ LTRMYPGLQVYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTMILALELCEEIVVYGMVSDSYCREKSHPSVPYHYFEKGR LDECQMYLAHEQAPRSAHRFITEKAVFSRWAKKRPIVFAHPSWRTE	20
HsST8Sia1	MSPCGRARRQTSRGAMAVLAWKFPRTRLPMGASALCVVVLCWLYIFPVYRLPNEKEIVQGVLQQGTAWRRNQTAARAFRKQMED CCDPAHLFAMTKMNSPMGKSMWYDGEFLYSFTIDNSTYSLFPQATPFQLPLKKCAVVGNGGILKKSGCGRQIDEANFVMRCNLPP LSSEYTKDVGSKSQLVTANPSIIRQRFQNLLWSRKTFVDNMKIYNHSYIYMPAFSMKTGTEPSLRVYYTLSDVGANQTVLFANPNFL RSIGKFWKSRGIHAKRLSTGLFLVSAALGLCEEVAIYGFWPFSVNMHEQPISHHYYDNVLPFSGFHAMPEEFLQLWYLHKIGALRMQ LDPCEDTSLQPTS	21
HsST8Sia2	MQLQFRSWMLAALTLLVVFLIFADISEIEEEIGNSGGRGTIRSAVNSLHSKSNRAEVVINGSSSPAVVDRSNESIKHNIQPASSKWRH NQTLSLRIRKQILKFLDAEKDISVLKGTLKPGDIIHYIFDRDSTMNVSQNLYELLPRTSPLKNKHFGTCAIVGNSGVLLNSGCGQEIDAH SFVIRCNLAPVQEYARDVGLKTDLVTMNPSVIQRAFEDLVNATWREKLLQRLHSLNGSILWIPAFMARGGKERVEWVNELILKHHVN VRTAYPSLRLLHAVRGYWLTNKVHIKRPTTGLLMYTLATRFCKQIYLYGFWPFPLDQNQNPVKYHYYDSLKYGYTSQASPHTMPLE FKALKSLHEQGALKLTVGQCDGAT	22
HsST8Sia3	MRNCKMARVASVLGLVMLSVALLILSLISYVSLKKENIFTTPKYASPGAPRMYMFHAGFRSQFALKFLDPSFVPITNSLTQELQEKPS KWKFNRTAFLHQRQEILQHVDVIKNFSLTKNSVRIGQLMHYDYSSHKYVFSISNNFRSLLPDVSPIMNKHYNICAVVGNSGILTGSQC GQEIDKSDFVFRCNFAPTEAFQRDVGRKTNLTTFNPSILEKYYNNLLTIQDRNNFFLSLKKLDGAILWIPAFFFHTSATVTRTLVDFFV	23

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	EHRGQLKVQLAWPGNIMQHVNRYWKNKHLSPKRLSTGILMYTLASAICEEIHLYGFWPFGFDPNTREDLPYHYYDKKGTKFTTKW QESHQLPAEFQLLYRMHGEGLTKLTLSHCA	
HsST8Sia4	MRSIRKRWTICTISLLLIFYKTKEIARTEEHQETQLIGDGELSLSRSLVNSSDKIIRKAGSSIFQHNVEGWKINSSLVLEIRKNILRFLDAE RDVSVVKSSFKPGDVIHYVLDRRRTLNISHDLHSLLPEVSPMKNRRFKTCAVVGNSGILLDSECGKEIDSHNFVIRCNLAPVVEFAAD VGTKSDFITMNPSVVQRAFGGFRNESDREKFVHRLSMLNDSVLWIPAFMVKGGEKHVEWVNALILKNKLKVRTAYPSLRLIHAVRG YWLTNKVPIKRPSTGLLMYTLATRFCDEIHLYGFWPFPKDLNGKAVKYHYYDDLKYRYFSNASPHRMPLEFKTLNVLHNRGALKLTT GKCVKQ	24
HsGalNAcT1	MRKFAYCKVVLATSLIWVLLDMFLLLYFSECNKCDEKKERGLPAGDVLEPVQKPHEGPGEMGKPVVIPKEDQEKMKEMFKINQFNL MASEMIALNRSLPDVRLEGCKTKVYPDNLPTTSVVIVFHNEAWSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKL KVPVHVIRMEQRSGLIRARLKGAAVSKGQVITFLDAHCECTVGWLEPLLARIKHDRRTVVCPIIDVISDDTFEYMAGSDMTYGGFNW KLNFRWYPVPQREMDRRKGDRTLPVRTPTMAGGLFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFR KATPYTFPGGTGQIINKNNRRLAEVWMDEFKNFFYIISPGVTKVDYGDISSRVGLRHKLQCKPFSWYLENIYPDSQIPRHYFSLGEIR NVETNQCLDNMARKENEKVGIFNCHGMGGNQVFSYTANKEIRTDDLCLDVSKLNGPVTMLKCHHLKGNQLWEYDPVKLTLQHVNS NQCLDKATEEDSQVPSIRDCNGSRSQQWLLRNVTLPEIF	25
HsGalNAcT2	MRRRSRMLLCFAFLWVLGIAYYMYSGGGSALAGGAGGGAGRKEDWNEIDPIKKKDLHHSNGEEKAQSMETLPPGKVRWPDFNQE AYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIK EIILVDDYSNDPEDGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVIN MDNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISF RVWQCGGSLEIIPCSRVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYYAAVPSARNVPYGNIQSRLELRKKLSCKP FKWYLENVYPELRVPDHQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVDRAPGSLIKL QGCRENDSRQKWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ	26
HsGalNAcT3	MAHLKRLVKLHIKRHYHKKFWKLGAVIFFFIIVLVLMQREVSVQYSKEESRMERNMKNKNKMLDLMLEAVNNIKDAMPKMQIGAPVR QNIDAGERPCLQGYYTAAELKPVLDRPPQDSNAPGASGKAFKTTNLSVEEQKEKERGEAKHCFNAFASDRISLHRDLGPDTRPPEC IEQKFKRCPPLPTTSVIIVFHNEAWSTLLRTVHSVLYSSPAILLKEIILVDDASVDEYLHDKLDEYVKQFSIVKIVRQRERKGLITARLLGA TVATAETLTFLDAHCECFYGWLEPLLARIAENYTAVVSPDIASIDLNTFEFNKPSPYGSNHNRGNFDWSLSFGWESLPDHEKQRRK DETYPIKTPTFAGGLFSISKEYFEYIGSYDEEMEIWGGENIEMSFRVWQCGGQLEIMPCSVVGHVFRSKSPHSFPKGTQVIARNQVR LAEVWMDEYKEIFYRRNTDAAKIVKQKAFGDLSKRFEIKHRLQCKNFTWYLNNIYPEVYVPDLNPVISGYIKSVGQPLCLDVGENNQ GGKPLIMYTCHGLGGNQYFEYSAQHEIRHNIQKELCLHAAQGLVQLKACTYKGHKTVVTGEQIWEIQKDQLLYNPFLKMCLSANGE HPSLVSCNPSDPLQKWILSQND	27

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsGalNAcT4	MAVRWTWAGKSCLLLAFLTVAYIFVELLVSTFHASAGAGRARELGSRRLSDLQKNTEDLSRPLYKKPPADSRALGEWGKASKLQLN EDELKQQEELIERYAINIYLSDRISLHRHIEDKRMYECKSQKFNYRTLPTTSVIIAFYNEAWSTLLRTIHSVLETSPAVLLKEIILVDDLSD RVYLKTQLETYISNLDRVRLIRTNKREGLVRARLIGATFATGDVLTFLDCHCECNSGWLEPLLERIGRDETAVVCPVIDTIDWNTFEFY MQIGEPMIGGFDWRLTFQWHSVPKQERDRRISRIDPIRSPTMAGGLFAVSKKYFQYLGTYDTGMEVWGGENLELSFRVWQCGGK LEIHPCSHVGHVFPKRAPYARPNFLQNTARAAEVWMDEYKEHFYNRNPPARKEAYGDISERKLLRERLRCKSFDWYLKNVFPNLH VPEDRPGWHGAIRSRGISSECLDYNSPDNNPTGANLSLFGCHGQGGNQFFEYTSNKEIRFNSVTELCAEVPEQKNYVGMQNCPK DGFPVPANIIWHFKEDGTIFHPHSGLCLSAYRTPEGRPDVQMRTCDALDKNQIWSFEK	28
HsGalNAcT5	MNRIRKFFRGSGRVLAFIFVASVIWLLFDMAALRLSFSEINTRVIKEDIVRRERIGFRVQPDQGKIFYSSIKEMKPPLRGHGKGAWGKE NVRKTEESVLKVEVDLDQTQRERKMQNALGRGKVVPLWHPAHLQTLPVTPNKQKTDGRGTKPEASSHQGTPKQTTAQGAPKTSFI AAKGTQVVKISVHMGRVSLKQEPRKSHSPSSDTSKLAAERDLNVTISLSTDRPKQRSQAVANERAHPASTAVPKSGEAMALNKTKT QSKEVNANKHKANTSLPFPKFTVNSNRLRKQSINETPLGSLSKDDGARGAHGKKLNFSESHLVIITKEEEQKADPKEVSNSKTKTIFP KVLGKSQSKHISRNRSEMSSSSLAPHRVPLSQTNHALTGGLEPAKINITAKAPSTEYNQSHIKALLPEDSGTHQVLRIDVTLSPRDPK APGQFGRPVVVPHGKEKEAERRWKEGNFNVYLSDLIPVDRAIEDTRPAGCAEQLVHNNLPTTSVIMCFVDEVWSTLLRSVHSVINR SPPHLIKEILLVDDFSTKDYLKDNLDKYMSQFPKVRILRLKERHGLIRARLAGAQNATGDVLTFLDSHVECNVGWLEPLLERVYLSRK KVACPVIEVINDKDMSYMTVDNFQRGIFVWPMNFGWRTIPPDVIAKNRIKETDTIRCPVMAGGLFSIDKSYFFELGTYDPGLDVWGG ENMELSFKVWMCGGEIEIIPCSRVGHIFRNDNPYSFPKDRMKTVERNLVRVAEVWLDEYKELFYGHGDHLIDQGLDVGNLTQQREL RKKLKCKSFKWYLENVFPDLRAPIVRASGVLINVALGKCISIENTTVILEDCDGSKELQQFNYTWLRLIKCGEWCIAPIPDKGAVRLHP CDNRNKGLKWLHKSTSVFHPELVNHIVFENNQQLLCLEGNFSQKILKVAACDPVKPYQKWKFEKYYEA	29
HsGalNAcT6	MRLLRRRHMPLRLAMVGCAFVLFLFLLHRDVSSREEATEKPWLKSLVSRKDHVLDLMLEAMNNLRDSMPKLQIRAPEAQQTLFSIN QSCLPGFYTPAELKPFWERPPQDPNAPGADGKAFQKSKWTPLETQEKEEGYKKHCFNAFASDRISLQRSLGPDTRPPECVDQKF RRCPPLATTSVIIVFHNEAWSTLLRTVYSVLHTTPAILLKEIILVDDASTEEHLKEKLEQYVKQLQVVRVVRQEERKGLITARLLGASVA QAEVLTFLDAHCECFHGWLEPLLARIAEDKTVVVSPDIVTIDLNTFEFAKPVQRGRVHSRGNFDWSLTFGWETLPPHEKQRRKDET YPIKSPTFAGGLFSISKSYFEHIGTYDNQMEIWGGENVEMSFRVWQCGGQLEIIPCSVVGHVFRTKSPHTFPKGTSVIARNQVRLAE VWMDSYKKIFYRRNLQAAKMAQEKSFGDISERLQLREQLHCHNFSWYLHNVYPEMFVPDLTPTFYGAIKNLGTNQCLDVGENNRG GKPLIMYSCHGLGGNQYFEYTTQRDLRHNIAKQLCLHVSKGALGLGSCHFTGKNSQVPKDEEWELAQDQLIRNSGSGTCLTSQDK KPAMAPCNPSDPHQLWLFV	30
HsGalNAcT7	MRLKIGFILRSLLVVGSFLGLVVLWSSLTPRPDDPSPLSRMREDRDVNDPMPNRGGNGLAPGEDRFKPVVPWPHVEGVEVDLESIR RINKAKNEQEHHAGGDSQKDIMQRQYLTFKPQTFTYHDPVLRPGILGNFEPKEPEPPGVVGGPGEKAKPLVLGPEFKQAIQASIKEF GFNMVASDMISLDRSVNDLRQEECKYWHYDENLLTSSVVIVFHNEGWSTLMRTVHSVIKRTPRKYLAEIVLIDDFSNKEHLKEKLDE YIKLWNGLVKVFRNERREGLIQARSIGAQKAKLGQVLIYLDAHCEVAVNWYAPLVAPISKDRTICTVPLIDVINGNTYEIIPQGGGDED	31

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	GYARGAWDWSMLWKRVPLTPQEKRLRKTKTEPYRSPAMAGGLFAIEREFFFELGLYDPGLQIWGGENFEISYKIWQCGGKLLFVP CSRVGHIYRLEGWQGNPPPIYVGSSPTLKNYVRVVEVWWDEYKDYFYASRPESQALPYGDISELKKFREDHNCKSFKWFMEEIAY DITSHYPLPPKNVDWGEIRGFETAYCIDSMGKTNGGFVELGPCHRMGGNQLFRINEANQLMQYDQCLTKGADGSKVMITHCNLNE FKEWQYFKNLHRFTHIPSGKCLDRSEVLHQVFISNCDSSKTTQKWEMNNIHSV	
HsGalNAcT8	MMFWRKLPKALFIGLTLAIAVNLLLVFSSKGTLQNLFTGGLHRELPLHLNKRYGAVIKRLSHLEVELQDLKESMKLALRQQENVNSTL KRAKDEVRPLLKAMETKVNETKKHKTQMKLFPHSQLFRQWGEDLSEAQQKAAQDLFRKFGYNAYLSNQLPLNRTIPDTRDYRCLR KTYPSQLPSLSVILIFVNEALSIIQRAITSIINRTPSRLLKEIILVDDFSSNGELKVHLDEKIKLYNQKYPGLLKIIRHPERKGLAQARNTGW EAATADVVAILDAHIEVNVGWAEPILARIQEDRTVIVSPVFDNIRFDTFKLDKYELAVDGFNWELWCRYDALPQAWIDLHDVTAPVKS PSIMGILAANRHFLGEIGSLDGGMLIYGGENVELSLRVWQCGGKVEILPCSRIAHLERHHKPYALDLTAALKRNALRVAEIWMDEHKH MVYLAWNIPLQNSGIDFGDVSSRMALREKLKCKTFDWYLKNVYPLLKPLHTIVGYGRMKNLLDENVCLDQGPVPGNTPIMYYCHEF SSQNVYYHLTGELYVGQLIAEASASDRCLTDPGKAEKPTLEPCSKAAKNRLHIYWDFKPGGAVINRDTKRCLEMKKDLLGSHVLVLQ TCSTQVWEIQHTVRDWGQTNSQ	32
HsGalNAcT9	MAVARKIRTLLTVNILVFVGIVLFSVYCRLQGRSQELVRIVSGDRRVRSRHAKVGTLGDREAILQRLDHLEEVVYNQLNGLAKPIGLVE GPGGLGQGGLAATLRDDGQEAEGKYEEYGYNAQLSDRISLDRSIPDYRPRKCRQMSYAQDLPQVSVVFIFVNEALSVILRSVHSVV NHTPSQLLKEVILVDDNSDNVELKFNLDQYVNKRYPGLVKIVRNSRREGLIRARLQGWKAATAPVVGFFDAHVEFNTGWAEPALSRI REDRRRIVLPAIDNIKYSTFEVQQYANAAHGYNWGLRCMYIIPPQDWLDRGDESAPIRTPAMIGCSFVVDREYFGDIGLLDPGMEVY GGENVELGMRVWQCGGSMEVLPCSRVAHIERTRKPYNNDIDYYAKRNALRAAEVWMDDFKSHVYMAWNIPMSNPGVDFGDVSE RLALRQRLKCRSFKWYLENVYPEMRVYNNTLTYGEVRNSKASAYCLDQGAEDGDRAILYPCHGMSSQLVRYSADGLLQLGPLGST AFLPDSKCLVDDGTGRMPTLKKCEDVARPTQRLWDFTQSGPIVSRATGRCLEVEMSKDANFGLRLVVQRCSGQKWMIRNWIKHA RH	33
HsGalNAcT10	MRRKEKRLLQAVALVLAALVLLPNVGLWALYRERQPDGTPGGSGAAVAPAAGQGSHSRQKKTFFLGDGQKLKDWHDKEAIRRDA QRVGNGEQGRPYPMTDAERVDQAYRENGFNIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFHNEGWSSLLRTVHSVLNR SPPELVAEIVLVDDFSDREHLKKPLEDYMALFPSVRILRTKKREGLIRTRMLGASVATGDVITFLDSHCEANVNWLPPLLDRIARNRKT IVCPMIDVIDHDDFRYETQAGDAMRGAFDWEMYYKRIPIPPELQKADPSDPFESPVMAGGLFAVDRKWFWELGGYDPGLEIWGGE QYEISFKVWMCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVWMDEYAEYIYQRRPEYRHLSAGDVAVQKKLRSSL NCKSFKWFMTKIAWDLPKFYPPVEPPAAAWGEIRNVGTGLCADTKHGALGSPLRLEGCVRGRGEAAWNNMQVFTFTWREDIRPG DPQHTKKFCFDAISHTSPVTLYDCHSMKGNQLWKYRKDKTLYHPVSGSCMDCSESDHRIFMNTCNPSSLTQQWLFEHTNSTVLEK FNRN	34
HsB3GALNT1	MASALWTVLPSRMSLRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSH PSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAK	35

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	YVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIYEMMGHVKP IKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVCQLRRVIAAHGFSSKEIITFWQVMLRNTTCHY	
HsB4GALNT 1	MWLGRRALCALVLLLACASLGLLYASTRDAPGLRLPLAPWAPPQSPRRPELPDLAPEPRYAHIPVRIKEQVVGLLAWNNCSCESSG GGLPLPFQKQVRAIDLTKAFDPAELRAASATREQEFQAFLSRSQSPADQLLIAPANSPLQYPLQGVEVQPLRSILVPGLSLQAASGQ EVYQVNLTASLGTWDVAGEVTGVTLTGEGQADLTLVSPGLDQLNRQLQLVTYSSRSYQTNTADTVRFSTEGHEAAFTIRIRHPPNP RLYPPGSLPQGAQYNISALVTIATKTFLRYDRLRALITSIRRFYPTVTVVIADDSDKPERVSGPYVEHYLMPFGKGWFAGRNLAVSQV TTKYVLWVDDDFVFTARTRLERLVDVLERTPLDLVGGAVREISGFATTYRQLLSVEPGAPGLGNCLRQRRGFHHELVGFPGCVVTD GVVNFFLARTDKVREVGFDPRLSRVAHLEFFLDGLGSLRVGSCSDVVVDHASKLKLPWTSRDAGAETYARYRYPGSLDESQMAKH RLLFFKHRLQCMTSQ	36
HsA-ABO	MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMPGSLERGFCMAVREPDHLQRVSLPRMVYPQPKVLTPCRKDVLV VTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEV RAYKRWQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGVEILTPLFGTLHPGFYGSSREAFTYERRPQSQAYIPKDE GDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAV PKNHQAVRNP	37
HsA4GALT	MSKPPDLLLRLLRGAPRQRVCTLFIIGFKFTFFVSIMIYWHVVGEPKEKGQLYNLPAEIPCPTLTPPTPPSHGPTPGNIFFLETSDRTN PNFLFMCSVESAARTHPESHVLVLMKGLPGGNASLPRHLGISLLSCFPNVQMLPLDLRELFRDTPLADWYAAVQGRWEPYLLPVLS DASRIALMWKFGGIYLDTDFIVLKNLRNLTNVLGTQSRYVLNGAFLAFERRHEFMALCMRDFVDHYNGWIWGHQGPQLLTRVFKK WCSIRSLAESRACRGVTTLPPEAFYPIPWQDWKKYFEDINPEELPRLLSATYAVHVWNKKSQGTRFEATSRALLAQLHARYCPTTH EAMKMYL	38
HsB3GalT1	MASKVSCLYVLTVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFLINEPNKCEKNIPFLVILISTTHKE FDARQAIRETWGDENNFKGIKIATLFLLGKNADPVLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDS DIFVNMDNLIYKLLKPSTKPRRRYFTGYVINGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLHLEDV YVGLCLRKLGIHPFQNSGFNHWKMAYSLCRYRRVITVHQISPEEMHRIWNDMSSKKHLRC	39
HsB3GalT2	MLQWRRRHCCFAKMTWNAKRSLFRTHLIGVLSLVFLFAMFLFFNHHDWLPGRAGFKENPVTYTFRGFRSTKSETNHSSLRNIWKE TVPQTLRPQTATNSNNTDLSPQGVTGLENTLSANGSIYNEKGTGHPNSYHFKYIINEPEKCQEKSPFLILLIAAEPGQIEARRAIRQTW GNESLAPGIQITRIFLLGLSIKLNGYLQRAILEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLIN KLLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPPDLYPSERYPVFCSGTGYVFSGDLAEKIFKVSLGIRRLHLEDVYVGICLAKL RIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQFQPSELIKYWNHLQQNKHNACANAAKEKAGRYRHRKLH	40
HsB4GalT1	MRLREPLLSGSAAMPGASLQRACRLLVAVCALHLGVTLVYYLAGRDLSRLPQLVGVSTPLQGGSNSAAAIGQSSGELRTGGARPP PPLGASSQPRPGGDSSPVVDSGPGPASNLTSVPVPHTTALSLPACPEESPLLVGPMLIEFNMPVDLELVAKQNPNVKMGGRYAPR	41

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	DCVSPHKVAIIIPFRNRQEHLKYWLYYLHPVLQRQQLDYGIYVINQAGDTIFNRAKLLNVGFQEALKDYDYTCFVFSDVDLIPMNDHN AYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLTINGFPNNYWGWGGEDDDIFNRLVFRGMSISRPNAVVGRCRMIRH SRDKKNEPNPQRFDRIAHTKETMLSDGLNSLTYQVLDVQRYPLYTQITVDIGTPS	
HsB4GalT2	MSRLLGGTLERVCKAVLLLCLLHFLVAVILYFDVYAQHLAFFSRFSARGPAHALHPAASSSSSSSNCSRPNATASSSGLPEVPSALP GPTAPTLPPCPDSPPGLVGRLLIEFTSPMPLERVQRENPGVLMGGRYTPPDCTPAQTVAVIIPFRHREHHLRYWLHYLHPILRRQRL RYGVYVINQHGEDTFNRAKLLNVGFLEALKEDAAYDCFIFSDVDLVPMDDRNLYRCGDQPRHFAIAMDKFGFRLPYAGYFGGVSGL SKAQFLRINGFPNEYWGWGGEDDDIFNRISLTGMKISRPDIRIGRYRMIKHDRDKHNEPNPQRFTKIQNTKLTMKRDGIGSVRYQVL EVSRQPLFTNITVDIGRPPSWPPRG	42
HsB4GalT3	MLRRLLERPCTLALLVGSQLAVMMYLSLGGFRSLSALFGRDQGPTFDYSHPRDVYSNLSHLPGAPGGPPAPQGLPYCPERSPLLV GPVSVSFSPVPSLAEIVERNPRVEPGGRYRPAGCEPRSRTAIIVPHRAREHHLRLLLYHLHPFLQRQQLAYGIYVIHQAGNGTFNRA KLLNVGVREALRDEEWDCLFLHDVDLLPENDHNLYVCDPRGPRHVAVAMNKFGYSLPYPQYFGGVSALTPDQYLKMNGFPNEYW GWGGEDDDIATRVRLAGMKISRPPTSVGHYKMVKHRGDKGNEENPHRFDLLVRTQNSWTQDGMNSLTYQLLARELGPLYTNITAD IGTDPRGPRAPSGPRYPPGSSQAFRQEMLQRRPPARPGPLSTANHTALRGSH	43
HsB4GalT4	MGFNLTFHLSYKFRLLLLLTLCLTVVGWATSNYFVGAIQEIPKAKEFMANFHKTLILGKGKTLTNEASTKKVELDNCPSVSPYLRGQS KLIFKPDLTLEEVQAENPKVSRGRYRPQECKALQRVAILVPHRNREKHLMYLLEHLHPFLQRQQLDYGIYVIHQAEGKKFNRAKLLN VGYLEALKEENWDCFIFHDVDLVPENDFNLYKCEEHPKHLVVGRNSTGYRLRYSGYFGGVTALSREQFFKVNGFSNNYWGWGGE DDDLRLRVELQRMKISRPLPEVGKYTMVFHTRDKGNEVNAERMKLLHQVSRVWRTDGLSSCSYKLVSVEHNPLYINITVDFWFGA	44
HsB4GalT5	MRARRGLLRLPRRSLLAALFFFSLSSSLLYFVYVAPGIVNTYLFMMQAQGILIRDNVRTIGAQVYEQVLRSAYAKRNSSVNDSDYPLD LNHSETFLQTTTFLPEDFTYFANHTCPERLPSMKGPIDINMSEIGMDYIHELFSKDPTIKLGGHWKPSDCMPRWKVAILIPFRNRHEH LPVLFRHLLPMLQRQRLQFAFYVVEQVGTQPFNRAMLFNVGFQEAMKDLDWDCLIFHDVDHIPESDRNYYGCGQMPRHFATKLDK YMYLLPYTEFFGGVSGLTVEQFRKINGFPNAFWGWGGEDDDLWNRVQNAGYSVSRPEGDTGKYKSIPHHHRGEVQFLGRYALLR KSKERQGLDGLNNLNYFANITYDALYKNITVNLTPELAQVNEY	45
HsB4GalT6	MSVLRRMMRVSNRSLLAFIFFFSLSSSCLYFIYVAPGIANTYLFMVQARGIMLRENVKTIGHMIRLYTNKNSTLNGTDYPEGNNSSDY LVQTTTYLPENFTYSPYLPCPEKLPYMRGFLNVNVSEVSFDEIHQLFSKDLDIEPGGHWRPKDCKPRWKVAVLIPFRNRHEHLPIFFL HLIPMLQKQRLEFAFYVIEQTGTQPFNRAMLFNVGFKEAMKDSVWDCVIFHDVDHLPENDRNYYGCGEMPRHFAAKLDKYMYILPY KEFFGGVSGLTVEQFRKINGFPNAFWGWGGEDDDLWNRVHYAGYNVTRPEGDLGKYKSIPHHHRGEVQFLGRYKLLRYSKERQY IDGLNNLIYRPKILVDRLYTNISVNLMPELAPIEDY	46
HsB-ABO	MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMPGSLERGFCMAVREPDHLQRVSLPRMVYPQPKVLTPCRKDVLV VTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEV RAYKRWQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGVEILTPLFGTLHPGFYGSSREAFTYERRPQSQAYIPKDE	47

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	GDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAV PKNHQAVRNP	
HsUGT8	MKSYTPYFILLWSAVGIAKAAKIIIVPPIMFESHMYIFKTLASALHERGHHTVFLLSEGRDIAPSNHYSLQRYPGIFNSTTSDAFLQSKM RNIFSGRLTAIELFDILDHYTKNCDLMVGNHALIQGLKKEKFDLLLVDPNDMCGFVIAHLLGVKYAVFSTGLWYPAEVGAPAPLAYVP EFNSLLTDRMNLLQRMKNTGVYLISRLGVSFLVLPKYERIMQKYNLLPEKSMYDLVHGSSLWMLCTDVALEFPRPTLPNVVYVGGIL TKPASPLPEDLQRWVNGANEHGFVLVSFGAGVKYLSEDIANKLAGALGRLPQKVIWRFSGPKPKNLGNNTKLIEWLPQNDLLGHSK IKAFLSHGGLNSIFETIYHGVPVVGIPLFGDHYDTMTRVQAKGMGILLEWKTVTEKELYEALVKVINNPSYRQRAQKLSEIHKDQPGH PVNRTIYWIDYIIRHNGAHHLRAAVHQISFCQYFLLDIAFVLLLGAALLYFLLSWVTKFIYRKIKSLWSRNKHSTVNGHYHNGILNGKYK RNGHIKHEKKVK	48
HsC1GLT	MASKSWLNFLTFLCGSAIGFLLCSQLFSILLGEKVDTQPNVLHNDPHARHSDDNGQNHLEGQMNFNADSSQHKDENTDIAENLYQK VRILCWVMTGPQNLEKKAKHVKATWAQRCNKVLFMSSEENKDFPAVGLKTKEGRDQLYWKTIKAFQYVHEHYLEDADWFLKADD DTYVILDNLRWLLSKYDPEEPIYFGRRFKPYVKQGYMSGGAGYVLSKEALKRFVDAFKTDKCTHSSSIEDLALGRCMEIMNVEAGDS RDTIGKETFHPFVPEHHLIKGYLPRTFWYWNYNYYPPVEGPGCCSDLAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPTLPERIL KEISQANKNEDTKVKLGNP	49
HsCOSMC	MLSESSSFLKGVMLGSIFCALITMLGHIRIGHGNRMHHHEHHHLQAPNKEDILKISEDERMELSKSFRVYCIILVKPKDVSLWAAVKET WTKHCDKAEFFSSENVKVFESINMDTNDMWLMMRKAYKYAFDKYRDQYNWFFLARPTTFAIIENLKYFLLKKDPSQPFYLGHTIKS GDLEYVGMEGGIVLSVESMKRLNSLLNIPEKCPEQGGMIWKISEDKQLAVCLKYAGVFAENAEDADGKDVFNTKSVGLSIKEAMTY HPNQVVEGCCSDMAVTFNGLTPNQMHVMMYGVYRLRAFGHIFNDALVFLPPNGSDND	50
HsAlg1	MAASCLVLLALCLLLPLLLLGGWKRWRRGRAARHVVAVVLGDVGRSPRMQYHALSLAMHGFSVTLLGFCNSKPHDELLQNNRIQIV GLTELQSLAVGPRVFQYGVKVVLQAMYLLWKLMWREPGAYIFLQNPPGLPSIAVCWFVGCLCGSKLVIDWHNYGYSIMGLVHGPN HPLVLLAKWYEKFFGRLSHLNLCVTNAMREDLADNWHIRAVTVYDKPASFFKETPLDLQHRLFMKLGSMHSPFRARSEPEDPVTER SAFTERDAGSGLVTRLRERPALLVSSTSWTEDEDFSILLAALEKFEQLTLDGHNLPSLVCVITGKGPLREYYSRLIHQKHFQHIQVCT PWLEAEDYPLLLGSADLGVCLHTSSSGLDLPMKVVDMFGCCLPVCAVNFKCLHELVKHEENGLVFEDSEELAAQLQMLFSNFPDP AGKLNQFRKNLRESQQLRWDESWVQTVLPLVMDT	51
HsAlg2	MAEEQGRERDSVPKPSVLFLHPDLGVGGAERLVLDAALALQARGCSVKIWTAHYDPGHCFAESRELPVRCAGDWLPRGLGWGG RGAAVCAYVRMVFLALYVLFLADEEFDVVVCDQVSACIPVFRLARRRKKILFYCHFPDLLLTKRDSFLKRLYRAPIDWIEEYTTGMAD CILVNSQFTAAVFKETFKSLSHIDPDVLYPSLNVTSFDSVVPEKLDDLVPKGKKFLLLSINRYERKKNLTLALEALVQLRGRLTSQDWE RVHLIVAGGYDERVLENVEHYQELKKMVQQSDLGQYVTFLRSFSDKQKISLLHSCTCVLYTPSNEHFGIVPLEAMYMQCPVIAVNSG GPLESIDHSVTGFLCEPDPVHFSEAIEKFIREPSLKATMGLAGRARVKEKFSPEAFTEQLYRYVTKLLV	52

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsAlg3	MAAGLRKRGRSGSAAQAEGLCKQWLQRAWQERRLLLREPRYTLLVAACLCLAEVGITFWVIHRVAYTEIDWKAYMAEVEGVINGT YDYTQLQGDTGPLVYPAGFVYIFMGLYYATSRGTDIRMAQNIFAVLYLATLLLVFLIYHQTCKVPPFVFFFMCCASYRVHSIFVLRLFN DPVAMVLLFLSINLLLAQRWGWGCCFFSLAVSVKMNVLLFAPGLLFLLLTQFGFRGALPKLGICAGLQVVLGLPFLLENPSGYLSRSF DLGRQFLFHWTVNWRFLPEALFLHRAFHLALLTAHLTLLLLFALCRWHRTGESILSLLRDPSKRKVPPQPLTPNQIVSTLFTSNFIGIC FSRSLHYQFYVWYFHTLPYLLWAMPARWLTHLLRLLVLGLIELSWNTYPSTSCSSAALHICHAVILLQLWLGPQPFPKSTQHSKKAH	53
HsAlg11	MAAGERSWCLCKLLRFFYSLFFPGLIVCGTLCVCLVIVLWGIRLLLQRKKKLVSTSKNGKNQMVIAFFHPYCNAGGGGERVLWCALR ALQKKYPEAVYVVYTGDVNVNGQQILEGAFRRFNIRLIHPVQFVFLRKRYLVEDSLYPHFTLLGQSLGSIFLGWEALMQCVPDVYIDS MGYAFTLPLFKYIGGCQVGSYVHYPTISTDMLSVVKNQNIGFNNAAFITRNPFLSKVKLIYYYLFAFIYGLVGSCSDVVMVNSSWTLN HILSLWKVGNCTNIVYPPCDVQTFLDIPLHEKKMTPGHLLVSVGQFRPEKNHPLQIRAFAKLLNKKMVESPPSLKLVLIGGCRNKDDE LRVNQLRRLSEDLGVQEYVEFKINIPFDELKNYLSEATIGLHTMWNEHFGIGVVECMAAGTIILAHNSGGPKLDIVVPHEGDITGFLAE SEEDYAETIAHILSMSAEKRLQIRKSARASVSRFSDQEFEVTFLSSVEKLFK	54
HsAlg12	MAGKGSSGRRPLLLGLLVAVATVHLVICPYTKVEESFNLQATHDLLYHWQDLEQYDHLEFPGVVPRTFLGPVVIAVFSSPAVYVLSL LEMSKFYSQLIVRGVLGLGVIFGLWTLQKEVRRHFGAMVATMFCWVTAMQFHLMFYCTRTLPNVLALPVVLLALAAWLRHEWARFI WLSAFAIIVFRVELCLFLGLLLLLALGNRKVSVVRALRHAVPAGILCLGLTVAVDSYFWRQLTWPEGKVLWYNTVLNKSSNWGTSPL LWYFYSALPRGLGCSLLFIPLGLVDRRTHAPTVLALGFMALYSLLPHKELRFIIYAFPMLNITAARGCSYLLNNYKKSWLYKAGSLLVI GHLVVNAAYSATALYVSHFNYPGGVAMQRLHQLVPPQTDVLLHIDVAAAQTGVSRFLQVNSAWRYDKREDVQPGTGMLAYTHILM EAAPGLLALYRDTHRVLASVVGTTGVSLNLTQLPPFNVHLQTKLVLLERLPRPS	55
HsAlg13	MKCVFVTVGTTSFDDLIACVSAPDSLQKIESLGYNRLILQIGRGTVVPEPFSTESFTLDVYRYKDSLKEDIQKADLVISHAGAGSCLET LEKGKPLVVVINEKLMNNHQLELAKQLHKEGHLFYCTCSTLPGLLQSMDLSTLKCYPPGQPEKFSAFLDKVVGLQK	56
HsAlg14	MVCVLVLAAAAGAVAVFLILRIWVVLRSMDVTPRESLSILVVAGSGGHTTEILRLLGSLSNAYSPRHYVIADTDEMSANKINSFELDRA DRDPSNMYTKYYIHRIPRSREVQQSWPSTVFTTLHSMWLSFPLIHRVKPDLVLCNGPGTCVPICVSALLLGILGIKKVIIVYVESICRVE TLSMSGKILFHLSDYFIVQWPALKEKYPKSVYLGRIV	57
HsDPM1	MASLEVSRSPRRSRRELEVRSPRQNKYSVLLPTYNERENLPLIVWLLVKSFSESGINYEIIIIDDGSPDGTRDVAEQLEKIYGSDRILLR PREKKLGLGTAYIHGMKHATGNYIIIMDADLSHHPKFIPEFIRKQKEGNFDIVSGTRYKGNGGVYGWDLKRKIISRGANFLTQILLRPG ASDLTGSFRLYRKEVLEKLIEKCVSKGYVFQMEMIVRARQLNYTIGEVPISFVDRVYGESKLGGNEIVSFLKGLLTLFATT	58
HsPIGM	MGSTKHWGEWLLNLKVAPAGVFGVAFLARVALVFYGVFQDRTLHVRYTDIDYQVFTDAARFVTEGRSPYLRATYRYTPLLGWLLTP NIYLSELFGKFLFISCDLLTAFLLYRLLLLKGLGRRQACGYCVFWLLNPLPMAVSSRGNADSIVASLVLMVLYLIKKRLVACAAVFYGF AVHMKIYPVTYILPITLHLLPDRDNDKSLRQFRYTFQACLYELLKRLCNRAVLLFVAVAGLTFFALSFGFYYEYGWEFLEHTYFYHLTR RDIRHNFSPYFYMLYLTAESKWSFSLGIAAFLPQLILLSAVSFAYYRDLVFCCFLHTSIFVTFNKVCTSQYFLWYLCLLPLVMPLVRMP WKRAVVLLMLWFIGQAMWLAPAYVLEFQGKNTFLFIWLAGLFFLLINCSILIQIISHYKEEPLTERIKYD	59

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsPIGB	MRRPLSKCGMEPGGGDASLTLHGLQNRSHGKIKLRKRKSTLYFNTQEKSARRRGDLLGENIYLLLFTIALRILNCFLVQTSFVPDEY WQSLEVSHHMVFNYGYLTWEWTERLRSYTYPLIFASIYKILHLLGKDSVQLLIWIPRLAQALLSAVADVRLYSLMKQLENQEVARWV FFCQLCSWFTWYCCTRTLTNTMETVLTIIALFYYPLEGSKSMNSVKYSSLVALAFIIRPTAVILWTPLLFRHFCQEPRKLDLILHHFLPV GFVTLSLSLMIDRIFFGQWTLVQFNFLKFNVLQNWGTFYGSHPWHWYFSQGFPVILGTHLPFFIHGCYLAPKRYRILLVTVLWTLLVY SMLSHKEFRFIYPVLPFCMVFCGYSLTHLKTWKKPALSFLFLSNLFLALYTGLVHQRGTLDVMSHIQKVCYNNPNKSSASIFIMMPCH STPYYSHVHCPLPMRFLQCPPDLTGKSHYLDEADVFYLNPLNWLHREFHDDASLPTHLITFSILEEEISAFLISSNYKRTAVFFHTHLP EGRIGSHIYVYERKLKGKFNMKMKF	60
HsPIGZ	MQICGSSVASVAAGTSFQVLGPVCWQQLDLKMAVRVLWGGLSLLRVLWCLLPQTGYVHPDEFFQSPEVMAEDILGVQAARPWEF YPSSSCRSVLFPLLISGSTFWLLRLWEELGPWPGLVSGYALLVGPRLLLTALSFALDGAVYHLAPPMGADRWNALALLSGSYVTLVF YTRTFSNTIEGLLFTWLLVLVSSHVTWGPTRKEPAPGPRWRSWLLGGIVAAGFFNRPTFLAFAVVPLYLWGTRGATNPGLKSLTRE ALVLLPGAALTAAVFVATDSWYFSSPATSRNLVLTPVNFLHYNLNPQNLARHGTHARLTHLAVNGFLLFGVLHAQALQAAWQRLQV GLQASAQMGLLRALGARSLLSSPRSYLLLLYFMPLALLSAFSHQEARFLIPLLVPLVLLCSPQTQPVPWKGTVVLFNALGALLFGCLH QGGLVPGLEYLEQVVHAPVLPSTPTHYTLLFTHTYMPPRHLLHLPGLGAPVEVVDMGGTEDWALCQTLKSFTRQPACQVAGGPWL CRLFVVTPGTTRRAVEKCSFPFKNETLLFPHLTLEDPPALSSLLSGAWRDHLSLHIVELGEET	61
HsAlg5	MAPLLLQLAVLGAALAAAALVLISIVAFTTATKMPALHRHEEEKFFLNAKGQKETLPSIWDSPTKQLSVVVPSYNEEKRLPVMMDEAL SYLEKRQKRDPAFTYEVIVVDDGSKDQTSKVAFKYCQKYGSDKVRVITLVKNRGKGGAIRMGIFSSRGEKILMADADGATKFPDVEK LEKGLNDLQPWPNQMAIACGSRAHLEKESIAQRSYFRTLLMYGFHFLVWFLCVKGIRDTQCGFKLFTREAASRTFSSLHVERWAFD VELLYIAQFFKIPIAEIAVNWTEIEGSKLVPFWSWLQMGKDLLFIRLRYLTGAWRLEQTRKMN	62
HsAlg6	MEKWYLMTVVVLIGLTVRWTVSLNSYSGAGKPPMFGDYEAQRHWQEITFNLPVKQWYFNSSDNNLQYWGLDYPPLTAYHSLLCA YVAKFINPDWIALHTSRGYESQAHKLFMRTTVLIADLLIYIPAVVLYCCCLKEISTKKKIANALCILLYPGLILIDYGHFQYNSVSLGFALW GVLGISCDCDLLGSLAFCLAINYKQMELYHALPFFCFLLGKCFKKGLKGKGFVLLVKLACIVVASFVLCWLPFFTEREQTLQVLRRLF PVDRGLFEDKVANIWCSFNVFLKIKDILPRHIQLIMSFCSTFLSLLPACIKLILQPSSKGFKFTLVSCALSFFLFSFQVHEKSILLVSLPVC LVLSEIPFMSTWFLLVSTFSMLPLLLKDELLMPSVVTTMAFFIACVTSFSIFEKTSEEELQLKSFSISVRKYLPCFTFLSRIIQYLFLISVIT MVLLTLMTVTLDPPQKLPDLFSVLVCFVSCLNFLFFLVYFNIIIMWDSKSGRNQKKIS	63
HsAlg8	MAALTIATGTGNWFSALALGVTLLKCLLIPTYHSTDFEVHRNWLAITHSLPISQWYYEATSEWTLDYPPFFAWFEYILSHVAKYFDQE MLNVHNLNYSSSRTLLFQRFSVIFMDVLFVYAVRECCKCIDGKKVGKELTEKPKFILSVLLLWNFGLLIVDHIHFQYNGFLFGLMLLSIA RLFQKRHMEGAFLFAVLLHFKHIYLYVAPAYGVYLLRSYCFTANKPDGSIRWKSFSFVRVISLGLVVFLVSALSLGPFLALNQLPQVF SRLFPFKRGLCHAYWAPNFWALYNALDKVLSVIGLKLKFLDPNNIPKASMTSGLVQQFQHTVLPSVTPLATLICTLIAILPSIFCLWFKP QGPRGFLRCLTLCALSSFMFGWHVHEKAILLAILPMSLLSVGKAGDASIFLILTTTGHYSLFPLLFTAPELPIKILLMLLFTIYSISSLKTLF RKEKPLFNWMETFYLLGLGPLEVCCEFVFPFTSWKVKYPFIPLLLTSVYCAVGITYAWFKLYVSVLIDSAIGKTKKQ	64

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
HsAlg10	MAQLEGYCFSAALSCTFLVSCLLFSAFSRALREPYMDEIFHLPQAQRYCEGHFSLSQWDPMITTLPGLYLVSVGVVKPAIWIFAWSE HVVCSIGMLRFVNLLFSVGNFYLLYLLFHKVQPRNKAASSIQRVLSTLTLAVFPTLYFFNFLYYTEAGSMFFTLFAYLMCLYGNHKTS AFLGFCGFMFRQTNIIWAVFCAGNVIAQKLTEAWKTELQKKEDRLPPIKGPFAEFRKILQFLLAYSMSFKNLSMLFCLTWPYILLGFLF CAFVVVNGGIVIGDRSSHEACLHFPQLFYFFSFTLFFSFPHLLSPSKIKTFLSLVWKHGILFLVVTLVSVFLVWKFTYAHKYLLADNRH YTFYVWKRVFQRYAILKYLLVPAYIFAGWSIADSLKSKPIFWNLMFFICLFIVIVPQKLLEFRYFILPYVIYRLNITLPPTSRLVCELSCYAI VNFITFYIFLNKTFQWPNSQDIQRFMW	65
HsUGCG	MALLDLALEGMAVFGFVLFLVLWLMHFMAIIYTRLHLNKKATDKQPYSKLPGVSLLKPLKGVDPNLINNLETFFELDYPKYEVLLCVQ DHDDPAIDVCKKLLGKYPNVDARLFIGGKKVGINPKINNLMPGYEVAKYDLIWICDSGIRVIPDTLTDMVNQMTEKVGLVHGLPYVAD RQGFAATLEQVYFGTSHPRYYISANVTGFKCVTGMSCLMRKDVLDQAGGLIAFAQYIAEDYFMAKAIADRGWRFAMSTQVAMQNS GSYSISQFQSRMIRWTKLRINMLPATIICEPISECFVASLIIGWAAHHVFRWDIMVFFMCHCLAWFIFDYIQLRGVQGGTLCFSKLDYA VAWFIRESMTIYIFLSALWDPTISWRTGRYRLRCGGTAEEILDV	66
HsB3GLCT	MRPPACWWLLAPPALLALLTCSLAFGLASEDTKKEVKQSQDLEKSGISRKNDIDLKGIVFVIQSQSNSFHAKRAEQLKKSILKQAADL TQELPSVLLLHQLAKQEGAWTILPLLPHFSVTYSRNSSWIFFCEEETRIQIPKLLETLRRYDPSKEWFLGKALHDEEATIIHHYAFSEN PTVFKYPDFAAGWALSIPLVNKLTKRLKSESLKSDFTIDLKHEIALYIWDKGGGPPLTPVPEFCTNDVDFYCATTFHSFLPLCRKPVKK KDIFVAVKTCKKFHGDRIPIVKQTWESQASLIEYYSDYTENSIPTVDLGIPNTDRGHCGKTFAILERFLNRSQDKTAWLVIVDDDTLISI SRLQHLLSCYDSGEPVFLGERYGYGLGTGGYSYITGGGGMVFSREAVRRLLASKCRCYSNDAPDDMVLGMCFSGLGIPVTHSPLF HQARPVDYPKDYLSHQVPISFHKHWNIDPVKVYFTWLAPSDEDKARQETQKGFREEL	67
HsGLYG	MTDQAFVTLTTNDAYAKGALVLGSSLKQHRTTRRLVVLATPQVSDSMRKVLETVFDEVIMVDVLDSGDSAHLTLMKRPELGVTLTKL HCWSLTQYSKCVFMDADTLVLANIDDLFDREELSAAPDPGWPDCFNSGVFVYQPSVETYNQLLHLASEQGSFDGGDQGILNTFFS SWATTDIRKHLPFIYNLSSISIYSYLPAFKVFGASAKVVHFLGRVKPWNYTYDPKTKSVKSEAHDPNMTHPEFLILWWNIFTTNVLPLL QQFGLVKDTCSYVNVLSDLVYTLAFSCGFCRKEDVSGAISHLSLGEIPAMAQPFVSSEERKERWEQGQADYMGADSFDNIKRKLDT YLQ	68
HsPOGLUT1	MEWWASSPLRLWLLLFLLPSAQGRQKESGSKWKVFIDQINRSLENYEPCSSQNCSCYHGVIEEDLTPFRGGISRKMMAEVVRRKL GTHYQITKNRLYRENDCMFPSRCSGVEHFILEVIGRLPDMEMVINVRDYPQVPKWMEPAIPVFSFSKTSEYHDIMYPAWTFWEGGP AVWPIYPTGLGRWDLFREDLVRSAAQWPWKKKNSTAYFRGSRTSPERDPLILLSRKNPKLVDAEYTKNQAWKSMKDTLGKPAAKD VHLVDHCKYKYLFNFRGVAASFRFKHLFLCGSLVFHVGDEWLEFFYPQLKPWVHYIPVKTDLSNVQELLQFVKANDDVAQEIAERG SQFIRNHLQMDDITCYWENLLSEYSKFLSYNVTRRKGYDQIIPKMLKTEL	69
HsGnT1	MLKKQSAGLVLWGAILFVAWNALLLLFFWTRPAPGRPPSVSALDGDPASLTREVIRLAQDAEVELERQRGLLQQIGDALSSQRGRV PTAAPPAQPRVPVTPAPAVIPILVIACDRSTVRRCLDKLLHYRPSAELFPIIVSQDCGHEETAQAIASYGSAVTHIRQPDLSSIAVPPDH RKFQGYYKIARHYRWALGQVFRQFRFPAAVVVEDDLEVAPDFFEYFRATYPLLKADPSLWCVSAWNDNGKEQMVDASRPELLYRT	70

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	DFFPGLGWLLLAELWAELEPKWPKAFWDDWMRRPEQRQGRACIRPEISRTMTFGRKGVSHGQFFDQHLKFIKLNQQFVHFTQLD LSYLQREAYDRDFLARVYGAPQLQVEKVRTNDRKELGEVRVQYTGRDSFKAFAKALGVMDDLKSGVPRAGYRGIVTFQFRGRRVH LAPPLTWEGYDPSWN	
HsGnT2	MRFRIYKRKVLILTLVVAACGFVLWSSNGRQRKNEALAPPLLDAEPARGAGGRGGDHPSVAVGIRRVSNVSAASLVPAVPQPEADN LTLRYRSLVYQLNFDQTLRNVDKAGTWAPRELVLVVQVHNRPEYLRLLLDSLRKAQGIDNVLVIFSHDFWSTEINQLIAGVNFCPVLQ VFFPFSIQLYPNEFPGSDPRDCPRDLPKNAALKLGCINAEYPDSFGHYREAKFSQTKHHWWWKLHFVWERVKILRDYAGLILFLEE DHYLAPDFYHVFKKMWKLKQQECPECDVLSLGTYSASRSFYGMADKVDVKTWKSTEHNMGLALTRNAYQKLIECTDTFCTYDDYN WDWTLQYLTVSCLPKFWKVLVPQIPRIFHAGDCGMHHKKTCRPSTQSAQIESLLNNNKQYMFPETLTISEKFTVVAISPPRKNGGW GDIRDHELCKSYRRLQ	71
HsGnT3	MKMRRYKLFLMFCMAGLCLISFLHFFKTLSYVTFPRELASLSPNLVSSFFWNNAPVTPQASPEPGGPDLLRTPLYSHSPLLQPLPPS KAAEELHRVDLVLPEDTTEYFVRTKAGGVCFKPGTKMLERPPPGRPEEKPEGANGSSARRPPRYLLSARERTGGRGARRKWVEC VCLPGWHGPSCGVPTVVQYSNLPTKERLVPREVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVCESNFTAYGEPRPLKFREML TNGTFEYIRHKVLYVFLDHFPPGGRQDGWIADDYLRTFLTQDGVSRLRNLRPDDVFIIDDADEIPARDGVLFLKLYDGWTEPFAFHM RKSLYGFFWKQPGTLEVVSGCTVDMLQAVYGLDGIRLRRRQYYTMPNFRQYENRTGHILVQWSLGSPLHFAGWHCSWCFTPEGI YFKLVSAQNGDFPRWGDYEDKRDLNYIRGLIRTGGWFDGTQQEYPPADPSEHMYAPKYLLKNYDRFHYLLDNPYQEPRSTAAGG WRHRGPEGRPPARGKLDEAEV	72
HsGnT4a	MRLRNGTVATALAFITSFLTLSWYTTWQNGKEKLIAYQREFLALKERLRIAEHRISQRSSELNTIVQQFKRVGAETNGSKDALNKFSD NTLKLLKELTSKKSLQVPSIYYHLPHLLKNEGSLQPAVQIGNGRTGVSIVMGIPTVKREVKSYLIETLHSLIDNLYPEEKLDCVIVVFIGE TDIDYVHGVVANLEKEFSKEISSGLVEVISPPESYYPDLTNLKETFGDSKERVRWRTKQNLDYCFLMMYAQEKGIYYIQLEDDIIVKQ NYFNTIKNFALQLSSEEWMILEFSQLGFIGKMFQAPDLTLIVEFIFMFYKEKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFRP SLFQHVGLHSSLSGKIQKLTDKDYMKPLLLKIHVNPPAEVSTSLKVYQGHTLEKTYMGEDFFWAITPIAGDYILFKFDKPVNVESYLFH SGNQEHPGDILLNTTVEVLPFKSEGLEISKETKDKRLEDGYFRIGKFENGVAEGMVDPSLNPISAFRLSVIQNSAVWAILNEIHIKKAT N	73
HsGCNT1	MLRTLLRRRLFSYPTKYYFMVLVLSLITFSVLRIHQKPEFVSVRHLELAGENPSSDINCTKVLQGDVNEIQKVKLEILTVKFKKRPRWT PDDYINMTSDCSSFIKRRKYIVEPLSKEEAEFPIAYSIVVHHKIEMLDRLLRAIYMPQNFYCIHVDTKSEDSYLAAVMGIASCFSNVFVA SRLESVVYASWSRVQADLNCMKDLYAMSANWKYLINLCGMDFPIKTNLEIVRKLKLLMGENNLETERMPSHKEERWKKRYEVVNG KLTNTGTVKMLPPLETPLFSGSAYFVVSREYVGYVLQNEKIQKLMEWAQDTYSPDEYLWATIQRIPEVPGSLPASHKYDLSDMQAV ARFVKWQYFEGDVSKGAPYPPCDGVHVRSVCIFGAGDLNWMLRKHHLFANKFDVDVDLFAIQCLDEHLRHKALETLKH	74
HsGCNT2	MMGSWKHCLFSASLISALIFVFVYNTELWENKRFLRAALSNASLLAEACHQIFEGKVFYPTENALKTTLDEATCYEYMVRSHYVTETL SEEEAGFPLAYTVTIHKDFGTFERLFRAIYMPQNVYCVHLDQKATDAFKGAVKQLLSCFPNAFLASKKESVVYGGISRLQADLNCLE	75

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	DLVASEVPWKYVINTCGQDFPLKTNREIVQYLKGFKGKNITPGVLPPDHAVGRTKYVHQELLNHKNSYVIKTTKLKTPPPHDMVIYFG TAYVALTRDFANFVLQDQLALDLLSWSKDTYSPDEHFWVTLNRIPGVPGSMPNASWTGNLRAIKWSDMEDRHGGCHGHYVHGICI YGNGDLKWLVNSPSLFANKFELNTYPLTVECLELRHRERTLNQSETAIQPSWYF	
HsB3GNT2	MSVGRRRIKLLGILMMANVFIYFIMEVSKSSSQEKNGKGEVIIPKEKFWKISTPPEAYWNREQEKLNRQYNPILSMLTNQTGEAGRLS NISHLNYCEPDLRVTSVVTGFNNLPDRFKDFLLYLRCRNYSLLIDQPDKCAKKPFLLLAIKSLTPHFARRQAIRESWGQESNAGNQTV VRVFLLGQTPPEDNHPDLSDMLKFESEKHQDILMWNYRDTFFNLSLKEVLFLRWVSTSCPDTEFVFKGDDDVFVNTHHILNYLNSLS KTKAKDLFIGDVIHNAGPHRDKKLKYYIPEVVYSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLQKLGLVPEKHK GFRTFDIEEKNKNNICSYVDLMLVHSRKPQEMIDIWSQLQSAHLKC	76
HsB3GNT6	MAFPCRRSLTAKTLACLLVGVSFLALQQWFLQAPRSPREERSPQEETPEGPTDAPAADEPPSELVPGPPCVANASANATADFEQL PARIQDFLRYRHCRHFPLLWDAPAKCAGGRGVFLLLAVKSAPEHYERRELIRRTWGQERSYGGRPVRRLFLLGTPGPEDEARAER LAELVALEAREHGDVLQWAFADTFLNLTLKHLHLLDWLAARCPHARFLLSGDDDVFVHTANVVRFLQAQPPGRHLFSGQLMEGSV PIRDSWSKYFVPPQLFPGSAYPVYCSGGGFLLSGPTARALRAAARHTPLFPIDDAYMGMCLERAGLAPSGHEGIRPFGVQLPGAQ QSSFDPCMYRELLLVHRFAPYEMLLMWKALHSPALSCDRGHRVS	77
HsPIGA	MACRGGAGNGHRASATLSRVSPGSLYTCRTRTHNICMVSDFFYPNMGGVESHIYQLSQCLIERGHKVIIVTHAYGNRKGIRYLTSGL KVYYLPLKVMYNQSTATTLFHSLPLLRYIFVRERVTIIHSHSSFSAMAHDALFHAKTMGLQTVFTDHSLFGFADVSSVLTNKLLTVSLC DTNHIICVSYTSKENTVLRAALNPEIVSVIPNAVDPTDFTPDPFRRHDSITIVVVSRLVYRKGIDLLSGIIPELCQKYPDLNFIIGGEGPKR IILEEVRERYQLHDRVRLLGALEHKDVRNVLVQGHIFLNTSLTEAFCMAIVEAASCGLQVVSTRVGGIPEVLPENLIILCEPSVKSLCEG LEKAIFQLKSGTLPAPENIHNIVKTFYTWRNVAERTEKVYDRVSVEAVLPMDKRLDRLISHCGPVTGYIFALLAVFNFLFLIFLRWMTP DSIIDVAIDATGPRGAWTNNYSHSKRGGENNEISETR	78
HsXXLT1	MGLLRGGLPCARAMARLGAVRSHYCALLLAAALAVCAFYYLGSGRETFSSATKRLKEARAGAPAAPSPPALELARGSVAPAPGAKA KSLEGGGAGPVDYHLLMMFTKAEHNAALQAKARVALRSLLRLAKFEAHEVLNLHFVSEEASREVAKGLLRELLPPAAGFKCKVIFHD VAVLTDKLFPIVEAMQKHFSAGLGTYYSDSIFFLSVAMHQIMPKEILQIIQLDLDLKFKTNIRELFEEFDSFLPGAIIGIAREMQPVYRHT FWQFRHENPQTRVGGPPPEGLPGFNSGVMLLNLEAMRQSPLYSRLLEPAQVQQLADKYHFRGHLGDQDFFTMIGMEHPKLFHVL DCTWNRQLCTWWRDHGYSDVFEAYFRCEGHVKIYHGNCNTPIPED	79
HsUGT1A1	MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQRED VKESFVSLGHNVFENDSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALP CSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVK DYPRPIMPNMVFVGGINCLHQNPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTI LVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKS	80

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	YKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKG RVKKAHKSKTH	
HsB4GAT1	MQMSYAIRCAFYQLLLAALMLVAMLQLLYLSLLSGLHGQEEQDQYFEFFPPSPRSVDQVKAQLRTALASGGVLDASGDYRVYRGLL KTTMDPNDVILATHASVDNLLHLSGLLERWEGPLSVSVFAATKEEAQLATVLAYALSSHCPDMRARVAMHLVCPSRYEAAVPDPRE PGEFALLRSCQEVFDKLARVAQPGINYALGTNVSYPNNLLRNLAREGANYALVIDVDMVPSEGLWRGLREMLDQSNQWGGTALVV PAFEIRRARRMPMNKNELVQLYQVGEVRPFYYGLCTPCQAPTNYSRWVNLPEESLLRPAYVVPWQDPWEPFYVAGGKVPTFDER FRQYGFNRISQACELHVAGFDFEVLNEGFLVHKGFKEALKFHPQKEAENQHNKILYRQFKQELKAKYPNSPRRC	81
HsUGT1A3	MATGLQVPLPWLATGLLLLLSVQPWAESGKVLVVPIDGSHWLSMREVLRELHARGHQAVVLTPEVNMHIKEENFFTLTTYAISWTQ DEFDRHVLGHTQLYFETEHFLKKFFRSMAMLNNMSLVYHRSCVELLHNEALIRHLNATSFDVVLTDPVNLCAAVLAKYLSIPTVFFLR NIPCDLDFKGTQCPNPSSYIPRLLTTNSDHMTFMQRVKNMLYPLALSYICHAFSAPYASLASELFQREVSVVDILSHASVWLFRGDF VMDYPRPIMPNMVFIGGINCANRKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANN TILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDK SYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKK GRVKKAHKSKTH	82
CjCsTII	MKNNIIIAGNGSSLKEIDYSRLPNDFDVFRCNQFYFEDKYYLGKKCKAVFYNSTLFFEQYYTLKHLIQNQEYETELIMCSNFNQAHLE NENFLKNFYDYFPDAHLGYDFFKQLKDFNAYFKFHEIYFNQRITSGIYMCAVAIALGYKEIYLSGIDFYQNGSSYAFDTKQKNLLKLAP NFKNDNSHYIGHSKNTDIKALEFLEKTYKIKLYCLCPNSLLANFIELAPNLNSNFTIQEKNNYTKDILIPSKEAYKKFTKMTKTQEILKNNI YYKLTKDLLRLPSDIKHYFKGK	83
NmPolysiaT	MLKKIKKALFQPKKFFQDSMWLTTSPFYLTPPRNNLFVISNLGQLNQVQSLIKIQKLTNNLLVILYTSKNLKMPKLVHQSANKNLFESIY LFELPRSPNNITPKKLLYIYRSYKKILNIIQPAHLYMLSFTGHYSYLISIAKKKNITTHLIDEGTGTYAPLLESFSYHPTKLERYLIGNNLNI KGYIDHFDILHVPFPEYAKKIFNAKKYNRFFAHAGGISINNNIANLQKKYQISKNDYIFVSQRYPISDDLYYKSIVEILNSISLQIKGKIFIKL HPKEMGNNYVMSLFLNMVEINPRLVVINEPPFLIEPLIYLTNPKGIIGLASSSLIYTPLLSPSTQCLSIGELIINLIQKYSMVENTEMIQEHL EIIKKFNFINILNDLNGVISNPLFKTEETFETLLKSAEFAYKSKNYFQAIFYWQLASKNNITLLGHKALWYYNALYNVKQIYKMEYSDIFYI DNISVDFHSKDKLTWEKIKHYYYSADNRIGRDR	84
CjCgtB	MGFKISIILPTYNVEQYIARAIESCIKQTFKDIEIIVVDDCGNDNSINIAKEYSKKDKRIKIIHNEKNIGLLRARYEGVKVANSPYIMFLDPD DYLELNACEECIKILDEQDEVDLVFFNAIVESNVISYKKFDFNSGFYSKKEFVKKIIAEKNLYWTMWGKLIRKKLYLEAFASLRLEKDVK INMAEDVLLYYPMLSQAQKIAYMNCNLYHYVPNNNSICNTKNEVGVKNNIQDLQLVLNYLRQNYILNKYCSVLYVLIKYLLYIQIYKIKR TKLMVTLLAKINILTLKILFKYKKFLKQC	85
HpLgtB	MRVFIISLNQKVCDQFGLVFRDTTTLLHNINATHHKAQIFDAIYSKTFESELHPLVKKHLHPYFITQNIKDMGITTNLISRVSKFYYALKY HAKFMSFGELGCYASHYSLWEKCIELNEPICILEDDITLKEDFKEGLDFLEKHIQELGYARLMYLLYDANVKSEPLSHKNHEIQERVGII	86

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	KAYSHGVGTQGYVITPKIAKVFKKCSRKWVVPVDTIMDATFIHGVKNLVLQPFVIADDEQISTIARKEEPYSSKIALMRKLHFKYLKYW QFV	
NmLgtB	MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSGVEKACFMSHAVLWKQALDEGLPYITVF EDDVLLGEGAEKFLAEDAWLQERFDPDTAFIVRLETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALP PEGLHPVDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDSPANTFKHRLIRALTKISRER EKRRQRREQFIVPFQ	87
NgLgtB	MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVEKACFMSHAVLWEQALDEGVPYIAVF EDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVRLETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVL PPERLHPVDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDSPANTFKHRLIRALTKIGR EREKRRQRREQLIGKIIVPFQ	88
EcFUT	MSIIRLQGGLGNQLFQFSFGYALSKINGTPLYFDISHYAENDDHGGYRLNNLQIPEEYLQYYTPKINNIYKFLVRGSRLYPEIFLFLGFC NEFHAYGYDFEYIAQKWKSKKYIGYWQSEHFFHKHILDLKEFFIPKNVSEQANLLAAKILESQSSLSIHIRRGDYIKNKTATLTHGVCS LEYYKKALNKIRDLAMIRDVFIFSDDIFWCKENIETLLSKKYNIYYSEDLSQEEDLWLMSLANHHIIANSSFSWWGAYLGTSASQIVIYP TPWYDITPKNTYIPIVNHWINVDKHSSC	89
EcWecA	MNLLTVSTDLISIFLFTTLFLFFARKVAKKVGLVDKPNFRKRHQGLIPLVGGISVYAGICFTFGIVDYYIPHASLYLACAGVLVFIGALDD RFDISVKIRATIQAAVGIVMMVFGKLYLSSLGYIFGSWEMVLGPFGYFLTLFAVWAAINAFNMVDGIDGLLGGLSCVSFAAIGMILWFD GQTSLAIWCFAMIAAILPYIMLNLGILGRRYKVFMGDAGSTLIGFTVIWILLETTQGKTHPISPVTALWIIAIPLMDMVAIMYRRLRKGMS PFSPDRQHIHHLIMRAGFTSRQAFVLITLAAALLASIGVLAEYSHFVPEWVMLVLFLLAFFLYGYCIKRAWKVARFIKRVKRRLRRNRG GSPNLTK	90
LpSetA	MYKIYSYLGWRIDMKTENLPQAGQEAQIDKKIHFIWVGHIMPQKNIQVVSEWAEKNPGYETIIWVDKKIAPAKELDLFILDMKSKGITV KDINEEGVCRDSIRHELDQESPNYGMVSDMLRLNILAAEGGIYLDSDILCSAPFPDEIYAPFGFLLSPWSQGANNTLCNDIILCSKGN QIIQQLADAIEQSYIARDSFEFTHEYASMKETKGERIAKTLGVTGPGFLFHQLKKMGILNDKSEMEAIHWELQDQRYLIDGSVKEPDY FYVPQNNTNDASWVPSIKRPGIENMSFQERLENAVQLIAFDIQKTGLFNLDHYANELKVKQNSWCIAAETSPELKPDSYLLIRPRDKT GEWTLYYVDEDKKLNPVTLPVIKGAIKLSEVSDPLRKFHTLLSQVSDPVNPTAHELKQIGRALIELKPRQDEWHCKNKWSGAEEIAQ ELWQRITSNETLRAQIKQCFTQFESLKPRVAELGLTRASGAGTEVEAHESTVKEQEIISQNTVGEEGTKEKNSVQLASENSSDEKIKT AHDLIDEIIQDVIQLDGKLGLLGGNTRQLEDGRVINIPNGAAMIFDDYKKYKQGELTAESALESMIKIAKLSNQLNRHTFFNQRQPETG QFYKKVAAIDLQTTIAAEYDNNHGLRI	91
NmSynE	MLQKIRKALFHPKKFFQDSQWFATPLFSSFAPKSNLFIISTFAQLNQAHSLTKMQKLKNNLLVILYTTQNMKMPKLIQKSVDKELFSVT YMFELPRKPGIVSPKKFLYIQRGYKKLLKTIQPAHLYVMSFAGHYSSLLSLAKKMNITTHLVEEGTATYAPLLESFTYKPTKFEQRFVG NNLHQKGYFDKFDILHVAFPEYAKKIFNANEYHRFFAHSGGISTSQSIAKIQDKYRISQNDYIFVSQRYPVSDEVYYKTIVETLNQMSL	92

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	RIEGKIFIKLHPKEMENKNIMSLFLNMVTINPRLVVINEPPFLIDPLIYLTTPKGIIGLTSTSIVYTPLLSPTTQCLSIGQIVIDSIHHTAQQEN TALIEEHLEIVKQFDFIKILSSIEDGIDTNSFKTEETLEMLLKSAEYAYKNKNFYQAIFYWQLASNNDLSVLGYKSLWYYNALNKVKQNY KMKYLEINYIERISLYFNDKDKMIWQNIKNDFFKYSLCNQ	
ScAlg1	MFLEIPRWLLALIILYLSIPLVVYYVIPYLFYGNKSTKKRIIIFVLGDVGHSPRICYHAISFSKLGWQVELCGYVEDTLPKIISSDPNITVHH MSNLKRKGGGTSVIFMVKKVLFQVLSIFKLLWELRGSDYILVQNPPSIPILPIAVLYKLTGCKLIIDWHNLAYSILQLKFKGNFYHPLVLIS YMVEMIFSKFADYNLTVTEAMRKYLIQSFHLNPKRCAVLYDRPASQFQPLAGDISRQKALTTKAFIKDYIRDDFDTEKGDKIIVTSTSF TPDEDIGILLGALKIYENSYVKFDSSLPKILCFITGKGPLKEKYMKQVEEYDWKRCQIEFVWLSAEDYPKLLQLCDYGVSLHTSSSGLD LPMKILDMFGSGLPVIAMNYPVLDELVQHNVNGLKFVDRRELHESLIFAMKDADLYQKLKKNVTQEAENRWQSNWERTMRGLKLIH	93
ScAlg11	MGSAWTNYNFEEVKSHFGFKKYVVSSLVLVYGLIKVLTWIFRQWVYSSLNPFSKKSSLLNRAVASCGEKNVKVFGFFHPYCNAGG GGEKVLWKAVDITLRKDAKNVIVIYSGDFVNGENVTPENILNNVKAKFDYDLDSDRIFFISLKLRYLVDSSTWKHFTLIGQAIGSMILAF ESIIQCPPDIWIDTMGYPFSYPIIARFLRRIPIVTYTHYPIMSKDMLNKLFKMPKKGIKVYGKILYWKVFMLIYQSIGSKIDIVITNSTWTNN HIKQIWQSNTCKIIYPPCSTEKLVDWKQKFGTAKGERLNQAIVLAQFRPEKRHKLIIESFATFLKNLPDSVSPIKLIMAGSTRSKQDENY VKSLQDWSENVLKIPKHLISFEKNLPFDKIEILLNKSTFGVNAMWNEHFGIAVVEYMASGLIPIVHASAGPLLDIVTPWDANGNIGKAP PQWELQKKYFAKLEDDGETTGFFFKEPSDPDYNTTKDPLRYPNLSDLFLQITKLDYDCLRVMGARNQQYSLYKFSDLKFDKDWENF VLNPICKLLEEEERG	94
NtGnTI	MRGNKFCCDFRYLLILAAVAFIYTQMRLFATQSEYADRLAAAIEAENHCTSQTRLLIDQISLQQGRIVALEEQMKRQDQECRQLRALV QDLESKGIKKLIGNVQMPVAAVVVMACNRADYLEKTIKSILKYQISVASKYPLFISQDGSHPDVRKLALSYDQLTYMQHLDFEPVHTE RPGELIAYYKIARHYKWALDQLFYKHNFSRVIILEDDMEIAPDFFDFFEAGATLLDRDKSIMAISSWNDNGQMQFVQDPYALYRSDFF PGLGWMLSKSTWDELSPKWPKAYWDDWLRLKENHRGRQFIRPEVCRTYNFGEHGSSLGQFFKQYLEPIKLNDVQVDWKSMDLS YLLEDNYVKHFGDLVKKAKPIHGADAVLKAFNIDGDVRIQYRDQLDFENIARQFGIFEEWKDGVPRAAYKGIVVFRYQTSRRVFLVG HDSLQQLGIEDT	95
NtGnTII	MASCKKARIKDAAFRRLVSLALITVSGVVLLLILLRTNTVSNLINSYTNESSESFESIEEAKVSNTHLTRIPKLPLQNELSILLSKRNQMP PRNLNLYSKLAKDHVVIVLYVHNRPQYLQIVIDSLSRVEGISETLLIVSHDGYFEEMNKIVESIKFCQVKQIFAPYSPHIFSDSFPGVSPK DCKEKDDPVEKHCEGMPDQYGNHRSPKIVSLKHHWWWMMNTVWDGLEVTRHHSGHILFIEEDHFIYPNAYRNMQLLAELKSKKC PDCYAANLAPSEVKLRGEGWECLVAERMGNVGYAFNRTVWRKIHKKAAEFCSFDDYNWDITMWSTVYPSFGSPVYTLRGSRTSA VHFGKCGLHQGHNRNVACMDNGGVNIVVDEIDKVANIKPEWEVRKYEHQAGYKAGFKGWGGWGDVRDRELCMEFAKMYI	96
BtGGTA1	MNVKGKVILSMLVVSTVIVVFWEYIHSPEGSLFWINPSRNPEVGGSSIQKGWWLPRWFNNGYHEEDGDINEEKEQRNEDESKLKLS DWFNPFKRPEVVTMTKWKAPVVWEGTYNRAVLDNYYAKQKITVGLTVFAVGRYIEHYLEEFLTSANKHFMVGHPVIFYIMVDDVSR MPLIELGPLRSFKVFKIKPEKRWQDISMMRMKTIGEHIVAHIQHEVDFLFCMDVDQVFQDKFGVETLGESVAQLQAWWYKADPNDF	97

FIG. 2C (continued)

Name	Sequence – full length (FASTA)	SEQ ID NO.
	TYERRKESAAYIPFGEGDFYYHAAIFGGTPTQVLNITQECFKGILKDKKNDIEAQWHDESHLNKYFLLNKPTKILSPEYCWDYHIGLPA DIKLVKMSWQTKEYNVVRNNV	
MmGGTA1	MITMLQDLHVNKISMSRSKSETSLPSSRSGSQEKIMNVKGKVILLMLIVSTVVVVFWEYVNRIPEVGENRWQKDWWFPSWFKNGTH SYQEDNVEGRREKGRNGDRIEEPQLWDWFNPKNRPDVLTVTPWKAPIVWEGTYDTALLEKYYATQKLTVGLTVFAVGKYIEHYLE DFLESADMYFMVGHRVIFYVMIDDTSRMPVVHLNPLHSLQVFEIRSEKRWQDISMMRMKTIGEHILAHIQHEVDFLFCMDVDQVFQD NFGVETLGQLVAQLQAWWYKASPEKFTYERRELSAAYIPFGEGDFYYHAAIFGGTPTHILNLTRECFKGILQDKKHDIEAQWHDESH LNKYFLFNKPTKILSPEYCWDYQIGLPSDIKSVKVAWQTKEYNLVRNNV	98
RnGGTA1	MVTMPQDLHVKVSMSRSKSETSLLSSRSGSQEKIMNVKGKIILSVLMVSTVLVVFWEYVNRTHSYQEEDIERAREKGRNGDSIVEP QLWDWFNPKNRPEVLTVTPWKAPIVWEGTYDTALLEKYYARQKITVGLTVFAVGKYIEHYLEDFLESANKYFMVGHRVIFYVMMDD TSRMPAVHLSPLHSLQVFEIRSEKRWQDISMMRMKTIGEHILDHIQHEVDFLFCMDVDQVFQDNFGVETLGQLVAQLQAWWYKAS PDEFTYERRELSAAYIPFGEGDFYYHAAVFGGTPVHILNLTRECFKGILQDKKHDIEAQWHDESHLNKYFLFNKPTKILSPEYCWDYH IGLPSDIKNVKIAWQTKEYNLVRSNV	99
BtB4GalT1	MKFREPLLGGSAAMPGASLQRACRLLVAVCALHLGVTLVYYLAGRDLRRLPQLVGVHPPLQGSSHGAAAIGQPSGELRLRGVAPP PPLQNSSKPRSRAPSNLDAYSHPGPGPGSNLTSAPVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLGGRYT PMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYGIYVINQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMND HNTYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGEDDDIYNRLAFRGMSVSRPNAVIGKCRMI RHSRDKKNEPNPQRFDRIAHTKETMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS	100

FIG. 2C (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsFUT1	MHIHQDSFPHGLGLSILCPDRRLVTPPVAIFCLPGTAMGPNASSSCPQHPASLSGTWTVYPNGRFGNQMGQYATLLALAQLNG RRAFILPAMHAALAPVFRITLPVLAPEVDSRTPWRELQLHDWMSEEYADLRDPFLKLSGFPCSWTFFHHLREQIRREFTLHDHLR EEAQSVLGQLRLGRTGDRPRTFVGVHVRRGDYLQVMPQRWKGVVGDSAYLRQAMDWFRARHEAPVFVVTSNGMEWCKENI DTSQGDVTFAGDGQEATPWKDFALLTQCNHTIMTIGTFGFWAAYLAGGDTVYLANFTLPDSEFLKIFKPEAAFLPEWVGINADLS PLWTLAKP	101
HsFUT2	MQRLAKIQAMWELPVQIPVLASTSKALGPSQLRGMWTINAIGRLGNQMGEYATLYALAKMNGRPAFIPAQMHSTLAPIFRITLPVL HSATASRIPWQNYHLNDWMEEEYRHIPGEYVRFTGYPCSWTFYHHLRQEILQEFTLHDHVREEAQKFLRGLQVNGSRPGTFVG VHVRRGDYVHVMPKVWKGVVADRRYLQQALDWFRARYSSLIFVVTSNGMAWCRENIDTSHGDVVFAGDGIEGSPAKDFALLT QCNHTIMTIGTFGIWAAYLTGGDTIYLANYTLPDSPFLKIFKPEAAFLPEWTGIAADLSPLLKH	102
HsFUT3	MRVSRDDATGSPRAPSGSSRQDTTPTRPTLLILLWTWPFHIPVALSRCSEMVPGTADCHITADRKVYPQADTVIVHHWDIMSNP KSRLPPSPRPQGQRWIWFNLEPPPNCQHLEALDRYFNLTMSYRSDSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSN WKPDSARVRYYQSLQAHLKVDVYGRSHKPLPKGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYE RFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALDFCKACWKLQQESRYQTVRSIAAWFT	103
HsFUT4	MGQLPPLPWASPTPSRPVGVLLWWEPFGGRDSAPRPPPDCRLRFNISGCRLLTDRASYGEAQAVLFHHRDLVKGPPDWPPP WGIQAHTAEEVDLRVLDYEEAAAAAEALATSSPRPPGQRWVWMNFESPSHSPGLRSLASNLFNWTLSYRADSDVFVPYGYLYP RSHPGDPPSGLAPPLSRKQGLVAWVVSHWDERQARVRYYHQLSQHVTVDVFGRGGPGQPVPEIGLLHTVARYKFYLAFENSQ HLDYITEKLWRNALLAGAVPVVLGPDRANYERFVPRGAFIHVDDFPSASSLASYLLFLDRNPAVYRRYFHWRRSYAVHITSFWDE PWCRVCQAVQRAGDRPKSIRNLASWFER	104
HsFUT5	MRVSRDDATGSPRPGLMAVEPVTGAPNGSRCQDSMATPAHPTLLILLWTWPFNTPVALPRCSEMVPGAADCNITADSSVYPQA DAVIVHHWDIMYNPSANLPPPTRPQGQRWIWFSMESPSNCRHLEALDGYFNLTMSYRSDSDIFTPYGWLEPWSGQPAHPPLNL SAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHKPLPKGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWA VPVVLGPSRSNYERFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALAFCKACWKLQQESRYQT VRSIAAWFT	105
HsFUT6	MRVSQDDPTVYPNGSRFPDSTGTPAHSIPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQADAVIVHHREVMYNPS AQLPRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRSDSDIFTPYGWLEPWSGQPAHPPLNLSAKTELVAWAVSN WGPNSARVRYYQSLQAHLKVDVYGRSHKPLPQGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYE RFLPPDAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALAFCKACWKLQEESRYQTRGIAAWFT	106
HsFUT7	MAPRGTPAPQPTITILVWHWPFTDQPPELPSDTCTRYGIARCHLSANRSLLASADAVVFHHRELQTRRSHLPLAQRPRGQPWV WASMESPSHTHGLSHLRGIFNWVLSYRRDSDIFVPYGRLEPHWGPSPPLPAKSRVAAWVVSNFQERQLRARLYRQLAPHLRV	107

FIG. 2D

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	DVFGRANGRPLCASCLVPTVAQYRFYLSFENSQHRDYITEKFWRNALVAGTVPVVLGPPRATYEAFVPADAFVHVDDFGSAREL AAFLTGMNESRYQRFFAWRDRLRVRLFTDWRERFCAICDRYPHLPRSQVYEDLEGWFQA	
HsFUT8	MRDNDHPDHSSRELSKILAKLERLKQQNEDLRRMAESLRIPEGPIDQGPAIGRVRVLEEQLVKAKEQIENYKKQTRNGLGKDHEI LRRRIENGAKELWFFLQSELKKLKNLEGNELQRHADEFLLDLGHHERSIMTDLYYLSQTDGAGDWREKEAKDLTELVQRRITYLQ NPKDCSKAKKLVCNINKGCGYGCQLHHVVYCFMIAYGTQRTLILESQNWRYATGGWETVFRPVSETCTDRSGISTGHWSGEVK DKNVQVVELPIVDSLHPRPPYLPLAVPEDLADRLVRVHGDPAVWWVSQFVKYLIRPQPWLEKEIEEATKKLGFKHPVIGVHVRRT DKVGTEAAFHPIEEYMVHVEEHFQLLARRMQVDKKRVYLATDDPSLLKEAKTKYPNYEFISDNSISWSAGLHNRYTENSLRGVIL DIHFLSQADFLVCTFSSQVCRVAYEIMQTLHPDASANFHSLDDIYYFGGQNAHNQIAIYAHQPRTADEIPMEPGDIIGVAGNHWD GYSKGVNRKLGRTGLYPSYKVREKIETVKYPTYPEAEK	108
HsFUT9	MTNSWIFSPMESASSVLKMKNFFSTKTDYFNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNKSHAVLIHHRDISW DLTNLPQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRRDSDIQVPYGFLTVSTNPFVFEVPSKEKLVCWVVSNWNPE HARVKYYNELSKSIEIHTYGQAFGEYVNDKNLIPTISTCKFYLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSRENYENYIPADSFI HVEDYNSPSELAKYLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHACLACDHVKRHQEYKSVGNLEKWFWN	109
HsFUT10	MLGKFERKEFKSSSLQDGHTKMEEAPTHLNSFLKKEGLTFNRKRKWELDSYPIMLWWSPLTGETGRLGQCGADACFFTINRTY LHHHMTKAFLFYGTDFNIDSLPLPRKAHHDWAVFHEESPKNNYKLFHKPVITLFNYTATFSRHSHLPLTTQYLESIEVLKSLRYLVP LQSKNKLRKRLAPLVYVQSDCDPPSDRDSYVRELMTYIEVDSYGECLRNKDLPQQLKNPASMDADGFYRIIAQYKFILAFENAVC DDYITEKFWRPLKLGVVPVYYGSPSITDWLPSNKSAILVSEFSHPRELASYIRRLDSDDRLYEAYVEWKLKGEISNQRLLTALRER KWGVQDVNQDNYIDAFECMVCTKVWANIRLQEKGLPPKRWEAEDTHLSCPEPTVFAFSPLRTPPLSSLREMWISSFEQSKKEA QALRWLVDRNQNFSSQEFWGLVFKD	110
HsFUT11	MGSVAEREAGGEAEWAEPWDGAVFRPPSALGAVGVTRSSGTPRPGREEAGDLPVLLWWSPGLFPHFPGDSERIECARGACV ASRNRRALRDSRTRALLFYGTDFRASAAPLPRLAHQSWALLHEESPLNNFLLSHGPGIRLFNLTSTFSRHSDYPLSLQWLPGTAY LRRPVPPPMERAEWRRRGYAPLLYLQSHCDVPADRDRYVRELMRHIPVDSYGKCLQNRELPTARLQDTATATTEDPELLAFLS RYKFHLALENAICNDYMTEKLWRPMHLGAVPVYRGSPSVRDWMPNNHSVILIDDFESPQKLAEFIDFLDKNDEEYMKYLAYKQP GGITNQFLLDSLKHREWGVNDPLLPNYLNGFECFVCDYELARLDAEKAHAASPGDSPVFEPHIAQPSHMDCPVPTPGFGNVEEI PENDSWKEMWLQDYWQGLDQGEALTAMIHNNETEQTKFWDYLHEIFMKRQHL	111
HsPOFUT1	MGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDF MEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVL ALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSR STAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSS FTAFVKRERDLQGRPSSFFGMDRPPKLRDEF	112

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsST3Gal1	MTTWFPKQMVLELSENLKRLIKHRPCTCTHCIGQRKLSAWFDERFNQTMQPLLTAQNALLEDDTYRWWLRLQREKKPNNLNDT IKELFRVVPGNVDPMLEKRSVGCRRCAVVGNSGNLRESSYGPEIDSHDFVLRMNKAPTAGFEADVGTKTTHHLVYPESFRELG DNVSMILVPFKTIDLEWVVSAITTGTISHTYIPVPAKIRVKQDKILIYHPAFIKYVFDNWLQGHGRYPSTGILSVIFSMHVCDEVDLYG FGADSKGNWHHYWENNPSAGAFRKTGVHDADFESNVTATLASINKIRIFKGR	113
HsST3Gal3	MKLHLLQWEEDSNSVVLSFDSAGQTLGSEYDRLGFLLNLDSKLPAELATKYANFSEGACKPGYASALMTAIFPRFSKPAPMFLD DSFRKWARIREFVPPFGIKGQDNLIKAILSVTKEYRLTPALDSLRCRRCIIVGNGGVLANKSLGSRIDDYDIVVRLNSAPVKGFEKD VGSKTTLRITYPEGAMQRPEQYERDSLFVLAGFKWQDFKWLKYIVYKERVSASDGFWKSVATRVPKEPPEIRILNPYFIQEAAFT LIGLPFNNGLMGRGNIPTLGSVAVTMALHGCDEVAVAGFGYDMSTPNAPLHYYETVRMAAIKESWTHNIQREKEFLRKLVKARVI TDLSSGI	114
HsST3Gal4	MREDRYIELFYFPIPEKKEPCLQGEAESKASKLFGNYSRDQPIFLRLEDYFWVKTPSAYELPYGTKGSEDLLLRVLAITSSSIPKNI QSLRCRRCVVVGNGHRLRNSSLGDAINKYDVVIRLNNAPVAGYEGDVGSKTTMRLFYPESAHFDPKVENNPDTLLVLVAFKAMD FHWIETILSDKKRVRKGFWKQPPLIWDVNPKQIRILNPFFMEIAADKLLSLPMQQPRKIKQKPTTGLLAITLALHLCDLVHIAGFGYP DAYNKKQTIHYYEQITLKSMAGSGHNVSQEALAIKRMLEMGAIKNLTSF	115
HsST3Gal6	MTNVYWVAPVEMKRRNKIQPCLSKPAFASLLRFHQFHPFLCAADFRKIASLYGSDKFDLPYGMRTSAEYFRLALSKLQSCDLFD EFDNIPCKKCVVVGNGGVLKNKTLGEKIDSYDVIIRMNNGPVLGHEEEVGRRTTFRLFYPESVFSDPIHNDPNTTVILTAFKPHDL RWLLELLMGDKINTNGFWKKPALNLIYKPYQIRILDPFIIRTAAYELLHFPKVFPKNQKPKHPTTGIIAITLAFYICHEVHLAGFKYNF SDLKSPLHYYGNATMSLMNKNAYHNVTAEQLFLKDIIEKNLVINLTQD	116
HsST6Gal1	MKEKKKGSYYDSFKLQTKEFQVLKSLGKLAMGSDSQSVSSSSTQDPHRGRQTLGSLRGLAKAKPEASFQVWNKDSSSKNLIPR LQKIWKNYLSMNKYKVSYKGPGPGIKFSAEALRCHLRDHVNVSMVEVTDFPFNTSEWEGYLPKESIRTKAGPWGRCAVVSSAG SLKSSQLGREIDDHDAVLRFNGAPTANFQQDVGTKTTIRLMNSQLVTTEKRFLKDSLYNEGILIVWDPSVYHSDIPKWYQNPDYN FFNNYKTYRKLHPNQPFYILKPQMPWELWDILQEISPEEIQPNPPSSGMLGIIIMMTLCDQVDIYEFLPSKRKTDVCYYYQKFFDS ACTMGAYHPLLYEKNLVKHLNQGTDEDIYLLGKATLPGFRTIHC	117
HsST6GalNAc1	MKEPQTKPSRHQRTENIKERSLQSLAKPKSQAPTRARRTTIYAEPVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDKVPH TAQRAAWKSPEKEKTVVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGNGGQTRKLTASRTVSEKHQGKAATTAKTLIPKS QHRMLVPTGAVSTRTRQKGVTTAVIPPKEKKPQATPPPAPFQSPTTQRNQRLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSV KIKASKSLWLQKLFLPNLTLFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLVQKVVTRFPPVPQQQLLLASLPAGSLRCITCAVV GNGGILNNSHIGQEIDSHDYVFRLSGALIKGYEQDVGTRTSFYGFTAFSLTQSLLILGNRGFKNVPLGKDVRYLHFLEGTRDYEW LEALLMNQTVMSKNLFWFRHRPQEAFREALHMDRYLLLHPDFLRYMKNRFLRSKTLDGAHWRIYRPTTGALLLLTALQLCDQVS AYGFITEGHERFSDHYYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN	118

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsST6GalNAc2	MSAVQRYPGPAAGARDTTSFEAFFQSKASNSWTGKGQACRHLLHLAIQRHPHFRGLFNLSIPVLLWGDLFTPALWDRLSQHKA PYGWRGLSHQVIASTLSLLNGSESAKLFAPPRDTPPKCIRCAVVGNGGILNGSRQGPNIDAHDYVFRLNGAVIKGFERDVGTKTS FYGFTVNTMKNSLVSYWNLGFTSVPQGQDLQYIFIPSDIRDYVMLRSAILGVPVPEGLDKGDRPHAYFGPEASASKFKLLHPDFI SYLTERFLKSKLINTHFGDLYMPSTGALMLLTALHTCDQVSAYGFITSNYWKFSDHYFERKMKPLIFYANHDLSLEAALWRDLHK AGILQLYQR	119
HsST6GalNAc4	MAGLPLCLATCLDHHFPTGSRPTVPGPLHFSGYSSVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFRMNQAPTV GFEADVGQRSTLRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRVLGGRTYRTLLQLTRMYPGLQVYTFTERMMAY CDQIFQDETGKNRRQSGSFLSTGWFTMILALELCEEIVVYGMVSDSYCREKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAH RFITEKAVFSRWAKKRPIVFAHPSWRTE	120
HsST8Sia1	MYRLPNEKEIVQGVLQQGTAWRRNQTAARAFRKQMEDCCDPAHLFAMTKMNSPMGKSMWYDGEFLYSFTIDNSTYSLFPQAT PFQLPLKKCAVVGNGGILKKSGCGRQIDEANFVMRCNLPPLSSEYTKDVGSKSQLVTANPSIIRQRFQNLLWSRKTFVDNMKIYN HSYIYMPAFSMKTGTEPSLRVYYTLSDVGANQTVLFANPNFLRSIGKFWKSRGIHAKRLSTGLFLVSAALGLCEEVAIYGFWPFS VNMHEQPISHHYYDNVLPFSGFHAMPEEFLQLWYLHKIGALRMQLDPCEDTSLQPTS	121
HsST8Sia2	MDISEIEEEIGNSGGRGTIRSAVNSLHSKSNRAEVVINGSSSPAVVDRSNESIKHNIQPASSKWRHNQTLSLRIRKQILKFLDAEKD ISVLKGTLKPGDIIHYIFDRDSTMNVSQNLYELLPRTSPLKNKHFGTCAIVGNSGVLLNSGCGQEIDAHSFVIRCNLAPVQEYARDV GLKTDLVTMNPSVIQRAFEDLVNATWREKLLQRLHSLNGSILWIPAFMARGGKERVEWVNELILKHHVNVRTAYPSLRLLHAVRG YWLTNKVHIKRPTTGLLMYTLATRFCKQIYLYGFWPFPLDQNQNPVKYHYYDSLKYGYTSQASPHTMPLEFKALKSLHEQGALK LTVGQCDGAT	122
HsST8Sia3	MKKENIFTTPKYASPGAPRMYMFHAGFRSQFALKFLDPSFVPITNSLTQELQEKPSKWKFNRTAFLHQRQEILQHVDVIKNFSLT KNSVRIGQLMHYDYSSHKYVFSISNNFRSLLPDVSPIMNKHYNICAVVGNSGILTGSQCGQEIDKSDFVFRCNFAPTEAFQRDVG RKTNLTTFNPSILEKYYNNLLTIQDRNNFFLSLKKLDGAILWIPAFFFHTSATVTRTLVDFFVEHRGQLKVQLAWPGNIMQHVNRY WKNKHLSPKRLSTGILMYTLASAICEEIHLYGFWPFGFDPNTREDLPYHYYDKKGTKFTTKWQESHQLPAEFQLLYRMHGEGLT KLTLSHCA	123
HsST8Sia4	MKTKEIARTEEHQETQLIGDGELSLSRSLVNSSDKIIRKAGSSIFQHNVEGWKINSSLVLEIRKNILRFLDAERDVSVVKSSFKPGD VIHYVLDRRRTLNISHDLHSLLPEVSPMKNRRFKTCAVVGNSGILLDSECGKEIDSHNFVIRCNLAPVVEFAADVGTKSDFITMNP SVVQRAFGGFRNESDREKFVHRLSMLNDSVLWIPAFMVKGGEKHVEWVNALILKNKLKVRTAYPSLRLIHAVRGYWLTNKVPIK RPSTGLLMYTLATRFCDEIHLYGFWPFPKDLNGKAVKYHYYDDLKYRYFSNASPHRMPLEFKTLNVLHNRGALKLTTGKCVKQ	124
HsGalNAcT1	MSECNKCDEKKERGLPAGDVLEPVQKPHEGPGEMGKPVVIPKEDQEKMKEMFKINQFNLMASEMIALNRSLPDVRLEGCKTKV YPDNLPTTSVVIVFHNEAWSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKLKVPVHVIRMEQRSGLIRARLKGA AVSKGQVITFLDAHCECTVGWLEPLLARIKHDRRTVVCPIIDVISDDTFEYMAGSDMTYGGFNWKLNFRWYPVPQREMDRRKGD	125

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	RTLPVRTPTMAGGLFSIDRDYFQEIGTYDAGMDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFRKATPYTFPGGTGQIINKNN RRLAEVWMDEFKNFFYIISPGVTKVDYGDISSRVGLRHKLQCKPFSWYLENIYPDSQIPRHYFSLGEIRNVETNQCLDNMARKEN EKVGIFNCHGMGGNQVFSYTANKEIRTDDLCLDVSKLNGPVTMLKCHHLKGNQLWEYDPVKLTLQHVNSNQCLDKATEEDSQV PSIRDCNGSRSQQWLLRNVTLPEIF	
HsGalNAcT2	MSGGGSALAGGAGGGAGRKEDWNEIDPIKKKDLHHSNGEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNK FNQVESDKLRMDRAIPDTRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSNDPEDGALLGK IEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLERVAEDRTRVVSPIIDVINMDNFQYVGASADLKGG FDWNLVFKWDYMTPEQRRSRQGNPVAPIKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIPCS RVGHVFRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYYAAVPSARNVPYGNIQSRLELRKKLSCKPFKWYLENVYPELR VPDHQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQEWALTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQ KWEQIEGNSKLRHVGSNLCLDSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ	126
HsGalNAcT3	MQREVSVQYSKEESRMERNMKNKNKMLDLMLEAVNNIKDAMPKMQIGAPVRQNIDAGERPCLQGYYTAAELKPVLDRPPQDS NAPGASGKAFKTTNLSVEEQKEKERGEAKHCFNAFASDRISLHRDLGPDTRPPECIEQKFKRCPPLPTTSVIIVFHNEAWSTLLRT VHSVLYSSPAILLKEIILVDDASVDEYLHDKLDEYVKQFSIVKIVRQRERKGLITARLLGATVATAETLTFLDAHCECFYGWLEPLLA RIAENYTAVVSPDIASIDLNTFEFNKPSPYGSNHNRGNFDWSLSFGWESLPDHEKQRRKDETYPIKTPTFAGGLFSISKEYFEYIG SYDEEMEIWGGENIEMSFRVWQCGGQLEIMPCSVVGHVFRSKSPHSFPKGTQVIARNQVRLAEVWMDEYKEIFYRRNTDAAKI VKQKAFGDLSKRFEIKHRLQCKNFTWYLNNIYPEVYVPDLNPVISGYIKSVGQPLCLDVGENNQGGKPLIMYTCHGLGGNQYFE YSAQHEIRHNIQKELCLHAAQGLVQLKACTYKGHKTVVTGEQIWEIQKDQLLYNPFLKMCLSANGEHPSLVSCNPSDPLQKWILS QND	127
HsGalNAcT4	MAGAGRARELGSRRLSDLQKNTEDLSRPLYKKPPADSRALGEWGKASKLQLNEDELKQQEELIERYAINIYLSDRISLHRHIEDK RMYECKSQKFNYRTLPTTSVIIAFYNEAWSTLLRTIHSVLETSPAVLLKEIILVDDLSDRVYLKTQLETYISNLDRVRLIRTNKREGLV RARLIGATFATGDVLTFLDCHCECNSGWLEPLLERIGRDETAVVCPVIDTIDWNTFEFYMQIGEPMIGGFDWRLTFQWHSVPKQE RDRRISRIDPIRSPTMAGGLFAVSKKYFQYLGTYDTGMEVWGGENLELSFRVWQCGGKLEIHPCSHVGHVFPKRAPYARPNFL QNTARAAEVWMDEYKEHFYNRNPPARKEAYGDISERKLLRERLRCKSFDWYLKNVFPNLHVPEDRPGWHGAIRSRGISSECLD YNSPDNNPTGANLSLFGCHGQGGNQFFEYTSNKEIRFNSVTELCAEVPEQKNYVGMQNCPKDGFPVPANIIWHFKEDGTIFHPH SGLCLSAYRTPEGRPDVQMRTCDALDKNQIWSFEK	128
HsGalNAcT5	MSFSEINTRVIKEDIVRRERIGFRVQPDQGKIFYSSIKEMKPPLRGHGKGAWGKENVRKTEESVLKVEVDLDQTQRERKMQNAL GRGKVVPLWHPAHLQTLPVTPNKQKTDGRGTKPEASSHQGTPKQTTAQGAPKTSFIAAKGTQVVKISVHMGRVSLKQEPRKSH SPSSDTSKLAAERDLNVTISLSTDRPKQRSQAVANERAHPASTAVPKSGEAMALNKTKTQSKEVNANKHKANTSLPFPKFTVNS NRLRKQSINETPLGSLSKDDGARGAHGKKLNFSESHLVIITKEEEQKADPKEVSNSKTKTIFPKVLGKSQSKHISRNRSEMSSSSL	129

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	APHRVPLSQTNHALTGGLEPAKINITAKAPSTEYNQSHIKALLPEDSGTHQVLRIDVTLSPRDPKAPGQFGRPVVVPHGKEKEAE RRWKEGNFNVYLSDLIPVDRAIEDTRPAGCAEQLVHNNLPTTSVIMCFVDEVWSTLLRSVHSVINRSPPHLIKEILLVDDFSTKDYL KDNLDKYMSQFPKVRILRLKERHGLIRARLAGAQNATGDVLTFLDSHVECNVGWLEPLLERVYLSRKKVACPVIEVINDKDMSYM TVDNFQRGIFVWPMNFGWRTIPPDVIAKNRIKETDTIRCPVMAGGLFSIDKSYFFELGTYDPGLDVWGGENMELSFKVWMCGGE IEIIPCSRVGHIFRNDNPYSFPKDRMKTVERNLVRVAEVWLDEYKELFYGHGDHLIDQGLDVGNLTQQRELRKKLKCKSFKWYLE NVFPDLRAPIVRASGVLINVALGKCISIENTTVILEDCDGSKELQQFNYTWLRLIKCGEWCIAPIPDKGAVRLHPCDNRNKGLKWL HKSTSVFHPELVNHIVFENNQQLLCLEGNFSQKILKVAACDPVKPYQKWKFEKYYEA	
HsGalNAcT6	MRDVSSREEATEKPWLKSLVSRKDHVLDLMLEAMNNLRDSMPKLQIRAPEAQQTLFSINQSCLPGFYTPAELKPFWERPPQDP NAPGADGKAFQKSKWTPLETQEKEEGYKKHCFNAFASDRISLQRSLGPDTRPPECVDQKFRRCPPLATTSVIIVFHNEAWSTLL RTVYSVLHTTPAILLKEIILVDDASTEEHLKEKLEQYVKQLQVVRVVRQEERKGLITARLLGASVAQAEVLTFLDAHCECFHGWLE PLLARIAEDKTVVVSPDIVTIDLNTFEFAKPVQRGRVHSRGNFDWSLTFGWETLPPHEKQRRKDETYPIKSPTFAGGLFSISKSYF EHIGTYDNQMEIWGGENVEMSFRVWQCGGQLEIIPCSVVGHVFRTKSPHTFPKGTSVIARNQVRLAEVWMDSYKKIFYRRNLQ AAKMAQEKSFGDISERLQLREQLHCHNFSWYLHNVYPEMFVPDLTPTFYGAIKNLGTNQCLDVGENNRGGKPLIMYSCHGLGG NQYFEYTTQRDLRHNIAKQLCLHVSKGALGLGSCHFTGKNSQVPKDEEWELAQDQLIRNSGSGTCLTSQDKKPAMAPCNPSDP HQLWLFV	130
HsGalNAcT7	MPRPDDPSPLSRMREDRDVNDPMPNRGGNGLAPGEDRFKPVVPWPHVEGVEVDLESIRRINKAKNEQEHHAGGDSQKDIMQ RQYLTFKPQTFTYHDPVLRPGILGNFEPKEPEPPGVVGGPGEKAKPLVLGPEFKQAIQASIKEFGFNMVASDMISLDRSVNDLRQ EECKYWHYDENLLTSSVVIVFHNEGWSTLMRTVHSVIKRTPRKYLAEIVLIDDFSNKEHLKEKLDEYIKLWNGLVKVFRNERREGL IQARSIGAQKAKLGQVLIYLDAHCEVAVNWYAPLVAPISKDRTICTVPLIDVINGNTYEIIPQGGGDEDGYARGAWDWSMLWKRV PLTPQEKRLRKTKTEPYRSPAMAGGLFAIEREFFFELGLYDPGLQIWGGENFEISYKIWQCGGKLLFVPCSRVGHIYRLEGWQG NPPPIYVGSSPTLKNYVRVVEVWWDEYKDYFYASRPESQALPYGDISELKKFREDHNCKSFKWFMEEIAYDITSHYPLPPKNVD WGEIRGFETAYCIDSMGKTNGGFVELGPCHRMGGNQLFRINEANQLMQYDQCLTKGADGSKVMITHCNLNEFKEWQYFKNLH RFTHIPSGKCLDRSEVLHQVFISNCDSSKTTQKWEMNNIHSV	131
HsGalNAcT8	MKGTLQNLFTGGLHRELPLHLNKRYGAVIKRLSHLEVELQDLKESMKLALRQQENVNSTLKRAKDEVRPLLKAMETKVNETKKH KTQMKLFPHSQLFRQWGEDLSEAQQKAAQDLFRKFGYNAYLSNQLPLNRTIPDTRDYRCLRKTYPSQLPSLSVILIFVNEALSIIQ RAITSIINRTPSRLLKEIILVDDFSSNGELKVHLDEKIKLYNQKYPGLLKIIRHPERKGLAQARNTGWEAATADVVAILDAHIEVNVG WAEPILARIQEDRTVIVSPVFDNIRFDTFKLDKYELAVDGFNWELWCRYDALPQAWIDLHDVTAPVKSPSIMGILAANRHFLGEIG SLDGGMLIYGGENVELSLRVWQCGGKVEILPCSRIAHLERHHKPYALDLTAALKRNALRVAEIWMDEHKHMVYLAWNIPLQNSGI DFGDVSSRMALREKLKCKTFDWYLKNVYPLLKPLHTIVGYGRMKNLLDENVCLDQGPVPGNTPIMYYCHEFSSQNVYYHLTGEL	132

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	YVGQLIAEASASDRCLTDPGKAEKPTLEPCSKAAKNRLHIYWDFKPGGAVINRDTKRCLEMKKDLLGSHVLVLQTCSTQVWEIQ HTVRDWGQTNSQ	
HsGalNAcT9	MLQGRSQELVRIVSGDRRVRSRHAKVGTLGDREAILQRLDHLEEVVYNQLNGLAKPIGLVEGPGGLGQGGLAATLRDDGQEAE GKYEEYGYNAQLSDRISLDRSIPDYRPRKCRQMSYAQDLPQVSVVFIFVNEALSVILRSVHSVVNHTPSQLLKEVILVDDNSDNVE LKFNLDQYVNKRYPGLVKIVRNSRREGLIRARLQGWKAATAPVVGFFDAHVEFNTGWAEPALSRIREDRRRIVLPAIDNIKYSTFE VQQYANAAHGYNWGLRCMYIIPPQDWLDRGDESAPIRTPAMIGCSFVVDREYFGDIGLLDPGMEVYGGENVELGMRVWQCGG SMEVLPCSRVAHIERTRKPYNNDIDYYAKRNALRAAEVWMDDFKSHVYMAWNIPMSNPGVDFGDVSERLALRQRLKCRSFKW YLENVYPEMRVYNNTLTYGEVRNSKASAYCLDQGAEDGDRAILYPCHGMSSQLVRYSADGLLQLGPLGSTAFLPDSKCLVDDG TGRMPTLKKCEDVARPTQRLWDFTQSGPIVSRATGRCLEVEMSKDANFGLRLVVQRCSGQKWMIRNWIKHARH	133
HsGalNAcT1 0	MRERQPDGTPGGSGAAVAPAAGQGSHSRQKKTFFLGDGQKLKDWHDKEAIRRDAQRVGNGEQGRPYPMTDAERVDQAYRE NGFNIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFHNEGWSSLLRTVHSVLNRSPPELVAEIVLVDDFSDREHLKKPLED YMALFPSVRILRTKKREGLIRTRMLGASVATGDVITFLDSHCEANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAM RGAFDWEMYYKRIPIPPELQKADPSDPFESPVMAGGLFAVDRKWFWELGGYDPGLEIWGGEQYEISFKVWMCGGRMEDIPCS RVGHIYRKYVPYKVPAGVSLARNLKRVAEVWMDEYAEYIYQRRPEYRHLSAGDVAVQKKLRSSLNCKSFKWFMTKIAWDLPKF YPPVEPPAAAWGEIRNVGTGLCADTKHGALGSPLRLEGCVRGRGEAAWNNMQVFTFTWREDIRPGDPQHTKKFCFDAISHTSP VTLYDCHSMKGNQLWKYRKDKTLYHPVSGSCMDCSESDHRIFMNTCNPSSLTQQWLFEHTNSTVLEKFNRN	134
HsB3GALNT 1	MIERVNWMYFYEYEPIYRQDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQEAEKED KMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTDTDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNY SYRGFYQKTHISYQEYPFKVFPPYCSGLGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVC QLRRVIAAHGFSSKEIITFWQVMLRNTTCHY	135
HsB4GALNT 1	MSTRDAPGLRLPLAPWAPPQSPRRPELPDLAPEPRYAHIPVRIKEQVVGLLAWNNCSCESSGGGLPLPFQKQVRAIDLTKAFDP AELRAASATREQEFQAFLSRSQSPADQLLIAPANSPLQYPLQGVEVQPLRSILVPGLSLQAASGQEVYQVNLTASLGTWDVAGE VTGVTLTGEGQADLTLVSPGLDQLNRQLQLVTYSSRSYQTNTADTVRFSTEGHEAAFTIRIRHPPNPRLYPPGSLPQGAQYNISA LVTIATKTFLRYDRLRALITSIRRFYPTVTVVIADDSDKPERVSGPYVEHYLMPFGKGWFAGRNLAVSQVTTKYVLWVDDDFVFTA RTRLERLVDVLERTPLDLVGGAVREISGFATTYRQLLSVEPGAPGLGNCLRQRRGFHHELVGFPGCVVTDGVVNFFLARTDKVR EVGFDPRLSRVAHLEFFLDGLGSLRVGSCSDVVVDHASKLKLPWTSRDAGAETYARYRYPGSLDESQMAKHRLLFFKHRLQCM TSQ	136

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsA-ABO	MAVREPDHLQRVSLPRMVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAE KHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEVRAYKRWQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHV GVEILTPLFGTLHPGFYGSSREAFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDE SHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP	137
HsA4GALT	MGEPKEKGQLYNLPAEIPCPTLTPPTPPSHGPTPGNIFFLETSDRTNPNFLFMCSVESAARTHPESHVLVLMKGLPGG NASLPRHLGISLLSCFPNVQMLPLDLRELFRDTPLADWYAAVQGRWEPYLLPVLSDASRIALMWKFGGIYLDTDFIVLKNLRNLTN VLGTQSRYVLNGAFLAFERRHEFMALCMRDFVDHYNGWIWGHQGPQLLTRVFKKWCSIRSLAESRACRGVTTLPPEAFYPIPW QDWKKYFEDINPEELPRLLSATYAVHVWNKKSQGTRFEATSRALLAQLHARYCPTTHEAMKMYL	138
HsB3GalT1	MRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINPHSFEFLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIA TLFLLGKNADPVLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDI FVNMDNLIYKLLKPSTKPRRRYFTGYVINGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTGYIFSADVAELIYKTSLHTRLLHLEDV YVGLCLRKLGIHPFQNSGFNHWKMAYSLCRYRRVITVHQISPEEMHRIWNDMSSKKHLRC	139
HsB3GalT2	MHDWLPGRAGFKENPVTYTFRGFRSTKSETNHSSLRNIWKETVPQTLRPQTATNSNNTDLSPQGVTGLENTLSANGSIYNEKG TGHPNSYHFKYIINEPEKCQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLAPGIQITRIFLLGLSIKLNGYLQRAILEESRQYHDIIQ QEYLDTYYNLTIKTLMGMNWVATYCPHIPYVMKTDSDMFVNTEYLINKLLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPP DLYPSERYPVFCSGTGYVFSGDLAEKIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPNEFVFNHWRVSYSSCKYSHLITSHQFQ PSELIKYWNHLQQNKHNACANAAKEKAGRYRHRKLH	140
HsB4GalT1	MRDLSRLPQLVGVSTPLQGGSNSAAAIGQSSGELRTGGARPPPPLGASSQPRPGGDSSPVVDSGPGPASNLTSVPVPHTTALS LPACPEESPLLVGPMLIEFNMPVDLELVAKQNPNVKMGGRYAPRDCVSPHKVAIIIPFRNRQEHLKYWLYYLHPVLQRQQLDYGI YVINQAGDTIFNRAKLLNVGFQEALKDYDYTCFVFSDVDLIPMNDHNAYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQ QFLTINGFPNNYWGWGGEDDDIFNRLVFRGMSISRPNAVVGRCRMIRHSRDKKNEPNPQRFDRIAHTKETMLSDGLNSLTYQV LDVQRYPLYTQITVDIGTPS	141
HsB4GalT2	MQHLAFFSRFSARGPAHALHPAASSSSSSSNCSRPNATASSSGLPEVPSALPGPTAPTLPPCPDSPPGLVGRLLIEFTSPMPLE RVQRENPGVLMGGRYTPPDCTPAQTVAVIIPFRHREHHLRYWLHYLHPILRRQRLRYGVYVINQHGEDTFNRAKLLNVGFLEAL KEDAAYDCFIFSDVDLVPMDDRNLYRCGDQPRHFAIAMDKFGFRLPYAGYFGGVSGLSKAQFLRINGFPNEYWGWGGEDDDIF NRISLTGMKISRPDIRIGRYRMIKHDRDKHNEPNPQRFTKIQNTKLTMKRDGIGSVRYQVLEVSRQPLFTNITVDIGRPPSWPPRG	142
HsB4GalT3	MRSLSALFGRDQGPTFDYSHPRDVYSNLSHLPGAPGGPPAPQGLPYCPERSPLLVGPVSVSFSPVPSLAEIVERNPRVEPGGR YRPAGCEPRSRTAIIVPHRAREHHLRLLLYHLHPFLQRQQLAYGIYVIHQAGNGTFNRAKLLNVGVREALRDEEWDCLFLHDVDL LPENDHNLYVCDPRGPRHVAVAMNKFGYSLPYPQYFGGVSALTPDQYLKMNGFPNEYWGWGGEDDDIATRVRLAGMKISRPP	143

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	TSVGHYKMVKHRGDKGNEENPHRFDLLVRTQNSWTQDGMNSLTYQLLARELGPLYTNITADIGTDPRGPRAPSGPRYPPGSS QAFRQEMLQRRPPARPGPLSTANHTALRGSH	
HsB4GalT4	MQEIPKAKEFMANFHKTLILGKGKTLTNEASTKKVELDNCPSVSPYLRGQSKLIFKPDLTLEEVQAENPKVSRGRYRPQECKALQ RVAILVPHRNREKHLMYLLEHLHPFLQRQQLDYGIYVIHQAEGKKFNRAKLLNVGYLEALKEENWDCFIFHDVDLVPENDFNLYK CEEHPKHLVVGRNSTGYRLRYSGYFGGVTALSREQFFKVNGFSNNYWGWGGEDDDLRLRVELQRMKISRPLPEVGKYTMVFH TRDKGNEVNAERMKLLHQVSRVWRTDGLSSCSYKLVSVEHNPLYINITVDFWFGA	144
HsB4GalT5	MPGIVNTYLFMMQAQGILIRDNVRTIGAQVYEQVLRSAYAKRNSSVNDSDYPLDLNHSETFLQTTTFLPEDFTYFANHTCPERLP SMKGPIDINMSEIGMDYIHELFSKDPTIKLGGHWKPSDCMPRWKVAILIPFRNRHEHLPVLFRHLLPMLQRQRLQFAFYVVEQVG TQPFNRAMLFNVGFQEAMKDLDWDCLIFHDVDHIPESDRNYYGCGQMPRHFATKLDKYMYLLPYTEFFGGVSGLTVEQFRKIN GFPNAFWGWGGEDDDLWNRVQNAGYSVSRPEGDTGKYKSIPHHHRGEVQFLGRYALLRKSKERQGLDGLNNLNYFANITYDA LYKNITVNLTPELAQVNEY	145
HsB4GalT6	MPGIANTYLFMVQARGIMLRENVKTIGHMIRLYTNKNSTLNGTDYPEGNNSSDYLVQTTTYLPENFTYSPYLPCPEKLPYMRGFL NVNVSEVSFDEIHQLFSKDLDIEPGGHWRPKDCKPRWKVAVLIPFRNRHEHLPIFFLHLIPMLQKQRLEFAFYVIEQTGTQPFNRA MLFNVGFKEAMKDSVWDCVIFHDVDHLPENDRNYYGCGEMPRHFAAKLDKYMYILPYKEFFGGVSGLTVEQFRKINGFPNAFW GWGGEDDDLWNRVHYAGYNVTRPEGDLGKYKSIPHHHRGEVQFLGRYKLLRYSKERQYIDGLNNLIYRPKILVDRLYTNISVNL MPELAPIEDY	146
HsB-ABO	MAVREPDHLQRVSLPRMVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTVFAIKKYVAFLKLFLETAE KHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEVGAYKRWQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHV GVEILTPLFGTLHPSFYGSSREAFTYERRPQSQAYIPKDEGDFYYMGAFFGGSVQEVQRLTRACHQAMMVDQANGIEAVWHDE SHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP	147
HsUGT8	MAKIIIVPPIMFESHMYIFKTLASALHERGHHTVFLLSEGRDIAPSNHYSLQRYPGIFNSTTSDAFLQSKMRNIFSGRLTAIELFDILD HYTKNCDLMVGNHALIQGLKKEKFDLLLVDPNDMCGFVIAHLLGVKYAVFSTGLWYPAEVGAPAPLAYVPEFNSLLTDRMNLLQ RMKNTGVYLISRLGVSFLVLPKYERIMQKYNLLPEKSMYDLVHGSSLWMLCTDVALEFPRPTLPNVVYVGGILTKPASPLPEDLQ RWVNGANEHGFVLVSFGAGVKYLSEDIANKLAGALGRLPQKVIWRFSGPKPKNLGNNTKLIEWLPQNDLLGHSKIKAFLSHGGL NSIFETIYHGVPVVGIPLFGDHYDTMTRVQAKGMGILLEWKTVTEKELYEALVKVINNPSYRQRAQKLSEIHKDQPGHPVNRTIYW IDYIIRHNGAHHLRAAVHQISFCQYFLLDIAFVLLLGAALLYFLLSWVTKFIYRKIKSLWSRNKHSTVNGHYHNGILNGKYKRNGHIK HEKKVK	148

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsC1GLT	MLLGEKVDTQPNVLHNDPHARHSDDNGQNHLEGQMNFNADSSQHKDENTDIAENLYQKVRILCWVMTGPQNLEKKAKHVKAT WAQRCNKVLFMSSEENKDFPAVGLKTKEGRDQLYWKTIKAFQYVHEHYLEDADWFLKADDDTYVILDNLRWLLSKYDPEEPIYF GRRFKPYVKQGYMSGGAGYVLSKEALKRFVDAFKTDKCTHSSSIEDLALGRCMEIMNVEAGDSRDTIGKETFHPFVPEHHLIKG YLPRTFWYWNYNYYPPVEGPGCCSDLAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPTLPERILKEISQANKNEDTKVKLGN P	149
HsAlg1	MKRWRRGRAARHVVAVVLGDVGRSPRMQYHALSLAMHGFSVTLLGFCNSKPHDELLQNNRIQIVGLTELQSLAVGPRVFQYGV KVVLQAMYLLWKLMWREPGAYIFLQNPPGLPSIAVCWFVGCLCGSKLVIDWHNYGYSIMGLVHGPNHPLVLLAKW YEKFFGRLSHLNLCVTNAMREDLADNWHIRAVTVYDKPASFFKETPLDLQHRLFMKLGSMHSPFRARSEPEDPVTERSAFTERD AGSGLVTRLRERPALLVSSTSWTEDEDFSILLAALEKFEQLTLDGHNLPSLVCVITGKGPLREYYSRLIHQKHFQHIQVCTPWLEA EDYPLLLGSADLGVCLHTSSSGLDLPMKVVDMFGCCLPVCAVNFKCLHELVKHEENGLVFEDSEELAAQLQMLFSNFPDPAGKL NQFRKNLRESQQLRWDESWVQTVLPLVMDT	150
HsAlg11	MGIRLLLQRKKKLVSTSKNGKNQMVIAFFHPYCNAGGGGERVLWCALRALQKKYPEAVYVVYTGDVNVNGQQILEGAFRRFNIR LIHPVQFVFLRKRYLVEDSLYPHFTLLGQSLGSIFLGWEALMQCVPDVYIDSMGYAFTLPLFKYIGGCQVGSYVHYPTISTDMLSV VKNQNIGFNNAAFITRNPFLSKVKLIYYYLFAFIYGLVGSCSDVVMVNSSWTLNHILSLWKVGNCTNIVYPPCDVQTFLDIPLHEKK MTPGHLLVSVGQFRPEKNHPLQIRAFAKLLNKKMVESPPSLKLVLIGGCRNKDDELRVNQLRRLSEDLGVQEYVEFKINIPFDELK NYLSEATIGLHTMWNEHFGIGVVECMAAGTIILAHNSGGPKLDIVVPHEGDITGFLAESEEDYAETIAHILSMSAEKRLQIRKSARA SVSRFSDQEFEVTFLSSVEKLFK	151
HsAlg14	MVLRSMDVTPRESLSILVVAGSGGHTTEILRLLGSLSNAYSPRHYVIADTDEMSANKINSFELDRADRDPSNMYTKYYIHRIPRSR EVQQSWPSTVFTTLHSMWLSFPLIHRVKPDLVLCNGPGTCVPICVSALLLGILGIKKVIIVYVESICRVETLSMSGKILFHLSDYFIV QWPALKEKYPKSVYLGRIV	152
HsAlg5	MTTATKMPALHRHEEEKFFLNAKGQKETLPSIWDSPTKQLSVVVPSYNEEKRLPVMMDEALSYLEKRQKRDPAFTYEVIVVDDG SKDQTSKVAFKYCQKYGSDKVRVITLVKNRGKGGAIRMGIFSSRGEKILMADADGATKFPDVEKLEKGLNDLQPWPNQMAIACG SRAHLEKESIAQRSYFRTLLMYGFHFLVWFLCVKGIRDTQCGFKLFTREAASRTFSSLHVERWAFDVELLYIAQFFKIPIAEIAVNW TEIEGSKLVPFWSWLQMGKDLLFIRLRYLTGAWRLEQTRKMN	153
HsB3GLCT	MASEDTKKEVKQSQDLEKSGISRKNDIDLKGIVFVIQSQSNSFHAKRAEQLKKSILKQAADLTQELPSVLLLHQLAKQEGAWTILP LLPHFSVTYSRNSSWIFFCEEETRIQIPKLLETLRRYDPSKEWFLGKALHDEEATIIHHYAFSENPTVFKYPDFAAGWALSIPLVNK LTKRLKSESLKSDFTIDLKHEIALYIWDKGGGPPLTPVPEFCTNDVDFYCATTFHSFLPLCRKPVKKKDIFVAVKTCKKFHGDRIPIV KQTWESQASLIEYYSDYTENSIPTVDLGIPNTDRGHCGKTFAILERFLNRSQDKTAWLVIVDDDTLISISRLQHLLSCYDSGEPVFL GERYGYGLGTGGYSYITGGGGMVFSREAVRRLLASKCRCYSNDAPDDMVLGMCFSGLGIPVTHSPLFHQARPVDYPKDYLSH QVPISFHKHWNIDPVKVYFTWLAPSDEDKARQETQKGFREEL	154

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsGnT1	MTRPAPGRPPSVSALDGDPASLTREVIRLAQDAEVELERQRGLLQQIGDALSSQRGRVPTAAPPAQPRVPVTPAPAVIPILVIAC DRSTVRRCLDKLLHYRPSAELFPIIVSQDCGHEETAQAIASYGSAVTHIRQPDLSSIAVPPDHRKFQGYYKIARHYRWALGQVFR QFRFPAAVVVEDDLEVAPDFFEYFRATYPLLKADPSLWCVSAWNDNGKEQMVDASRPELLYRTDFFPGLGWLLLAELWAELEP KWPKAFWDDWMRRPEQRQGRACIRPEISRTMTFGRKGVSHGQFFDQHLKFIKLNQQFVHFTQLDLSYLQREAYDRDFLARVY GAPQLQVEKVRTNDRKELGEVRVQYTGRDSFKAFAKALGVMDDLKSGVPRAGYRGIVTFQFRGRRVHLAPPLTWEGYDPSWN	155
HsGnT2	MQRKNEALAPPLLDAEPARGAGGRGGDHPSVAVGIRRVSNVSAASLVPAVPQPEADNLTLRYRSLVYQLNFDQTLRNVDKAGT WAPRELVLVVQVHNRPEYLRLLLDSLRKAQGIDNVLVIFSHDFWSTEINQLIAGVNFCPVLQVFFPFSIQLYPNEFPGSDPRDCPR DLPKNAALKLGCINAEYPDSFGHYREAKFSQTKHHWWWKLHFVWERVKILRDYAGLILFLEEDHYLAPDFYHVFKKMWKLKQQ ECPECDVLSLGTYSASRSFYGMADKVDVKTWKSTEHNMGLALTRNAYQKLIECTDTFCTYDDYNWDWTLQYLTVSCLPKFWKV LVPQIPRIFHAGDCGMHHKKTCRPSTQSAQIESLLNNNKQYMFPETLTISEKFTVVAISPPRKNGGWGDIRDHELCKSYRRLQ	156
HsGnT3	MHFFKTLSYVTFPRELASLSPNLVSSFFWNNAPVTPQASPEPGGPDLLRTPLYSHSPLLQPLPPSKAAEELHRVDLVLPEDTTEY FVRTKAGGVCFKPGTKMLERPPPGRPEEKPEGANGSSARRPPRYLLSARERTGGRGARRKWVECVCLPGWHGPSCGVPTVV QYSNLPTKERLVPREVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVCESNFTAYGEPRPLKFREMLTNGTFEYIRHKVLYVFL DHFPPGGRQDGWIADDYLRTFLTQDGVSRLRNLRPDDVFIIDDADEIPARDGVLFLKLYDGWTEPFAFHMRKSLYGFFWKQPGT LEVVSGCTVDMLQAVYGLDGIRLRRRQYYTMPNFRQYENRTGHILVQWSLGSPLHFAGWHCSWCFTPEGIYFKLVSAQNGDFP RWGDYEDKRDLNYIRGLIRTGGWFDGTQQEYPPADPSEHMYAPKYLLKNYDRFHYLLDNPYQEPRSTAAGGWRHRGPEGRP PARGKLDEAEV	157
HsGnT4a	MQNGKEKLIAYQREFLALKERLRIAEHRISQRSSELNTIVQQFKRVGAETNGSKDALNKFSDNTLKLLKELTSKKSLQVPSIYYHLP HLLKNEGSLQPAVQIGNGRTGVSIVMGIPTVKREVKSYLIETLHSLIDNLYPEEKLDCVIVVFIGETDIDYVHGVVANLEKEFSKEIS SGLVEVISPPESYYPDLTNLKETFGDSKERVRWRTKQNLDYCFLMMYAQEKGIYYIQLEDDIIVKQNYFNTIKNFALQLSSEEWMI LEFSQLGFIGKMFQAPDLTLIVEFIFMFYKEKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFRPSLFQHVGLHSSLSGKIQK LTDKDYMKPLLLKIHVNPPAEVSTSLKVYQGHTLEKTYMGEDFFWAITPIAGDYILFKFDKPVNVESYLFHSGNQEHPGDILLNTT VEVLPFKSEGLEISKETKDKRLEDGYFRIGKFENGVAEGMVDPSLNPISAFRLSVIQNSAVWAILNEIHIKKATN	158
HsGCNT1	MRIHQKPEFVSVRHLELAGENPSSDINCTKVLQGDVNEIQKVKLEILTVKFKKRPRWTPDDYINMTSDCSSFIKRRKYIVEPLSKEE AEFPIAYSIVVHHKIEMLDRLLRAIYMPQNFYCIHVDTKSEDSYLAAVMGIASCFSNVFVASRLESVVYASWSRVQADLNCMKDLY AMSANWKYLINLCGMDFPIKTNLEIVRKLKLLMGENNLETERMPSHKEERWKKRYEVVNGKLTNTGTVKMLPPLETPLFSGSAY FVVSREYVGYVLQNEKIQKLMEWAQDTYSPDEYLWATIQRIPEVPGSLPASHKYDLSDMQAVARFVKWQYFEGDVSKGAPYPP CDGVHVRSVCIFGAGDLNWMLRKHHLFANKFDVDVDLFAIQCLDEHLRHKALETLKH	159

FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
HsGCNT2	MYNTELWENKRFLRAALSNASLLAEACHQIFEGKVFYPTENALKTTLDEATCYEYMVRSHYVTETLSEEEAGFPLAYTVTIHKDF GTFERLFRAIYMPQNVYCVHLDQKATDAFKGAVKQLLSCFPNAFLASKKESVVYGGISRLQADLNCLEDLVASEVPWKYVINTCG QDFPLKTNREIVQYLKGFKGKNITPGVLPPDHAVGRTKYVHQELLNHKNSYVIKTTKLKTPPPHDMVIYFGTAYVALTRDFANFVL QDQLALDLLSWSKDTYSPDEHFWVTLNRIPGVPGSMPNASWTGNLRAIKWSDMEDRHGGCHGHYVHGICIYGNGDLKWLVNS PSLFANKFELNTYPLTVECLELRHRERTLNQSETAIQPSWYF	160
HsB3GNT2	MKSSSQEKNGKGEVIIPKEKFWKISTPPEAYWNREQEKLNRQYNPILSMLTNQTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNL PDRFKDFLLYLRCRNYSLLIDQPDKCAKKPFLLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVRVFLLGQTPPEDNHPDLSDM LKFESEKHQDILMWNYRDTFFNLSLKEVLFLRWVSTSCPDTEFVFKGDDDVFVNTHHILNYLNSLSKTKAKDLFIGDVIHNAGPHR DKKLKYYIPEVVYSGLYPPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLQKLGLVPEKHKGFRTFDIEEKNKNNICSY VDLMLVHSRKPQEMIDIWSQLQSAHLKC	161
HsB3GNT6	MQAPRSPREERSPQEETPEGPTDAPAADEPPSELVPGPPCVANASANATADFEQLPARIQDFLRYRHCRHFPLLWDAPAKCAG GRGVFLLLAVKSAPEHYERRELIRRTWGQERSYGGRPVRRLFLLGTPGPEDEARAERLAELVALEAREHGDVLQWAFADTFLN LTLKHLHLLDWLAARCPHARFLLSGDDDVFVHTANVVRFLQAQPPGRHLFSGQLMEGSVPIRDSWSKYFVPPQLFPGSAYPVY CSGGGFLLSGPTARALRAAARHTPLFPIDDAYMGMCLERAGLAPSGHEGIRPFGVQLPGAQQSSFDPCMYRELLLVHRFAPYE MLLMWKALHSPALSCDRGHRVS	162
HsXXLT1	MRETFSSATKRLKEARAGAPAAPSPPALELARGSVAPAPGAKAKSLEGGGAGPVDYHLLMMFTKAEHNAALQAKARVALRSLL RLAKFEAHEVLNLHFVSEEASREVAKGLLRELLPPAAGFKCKVIFHDVAVLTDKLFPIVEAMQKHFSAGLGTYYSDSIFFLSVAMH QIMPKEILQIIQLDLDLKFKTNIRELFEEFDSFLPGAIIGIAREMQPVYRHTFWQFRHENPQTRVGGPPPEGLPGFNSGVMLLNLEA MRQSPLYSRLLEPAQVQQLADKYHFRGHLGDQDFFTMIGMEHPKLFHVLDCTWNRQLCTWWRDHGYSDVFEAYFRCEGHVKI YHGNCNTPIPED	163
HsUGT1A1	MHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFENDSFLQRVI KTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRP LSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCL HQNPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDLLGHPMT RAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDR PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH	164
HsB4GAT1	MHGQEEQDQYFEFFPPSPRSVDQVKAQLRTALASGGVLDASGDYRVYRGLLKTTMDPNDVILATHASVDNLLHLSGLLERWEG PLSVSVFAATKEEAQLATVLAYALSSHCPDMRARVAMHLVCPSRYEAAVPDPREPGEFALLRSCQEVFDKLARVAQPGINYALG TNVSYPNNLLRNLAREGANYALVIDVDMVPSEGLWRGLREMLDQSNQWGGTALVVPAFEIRRARRMPMNKNELVQLYQVGEV	165

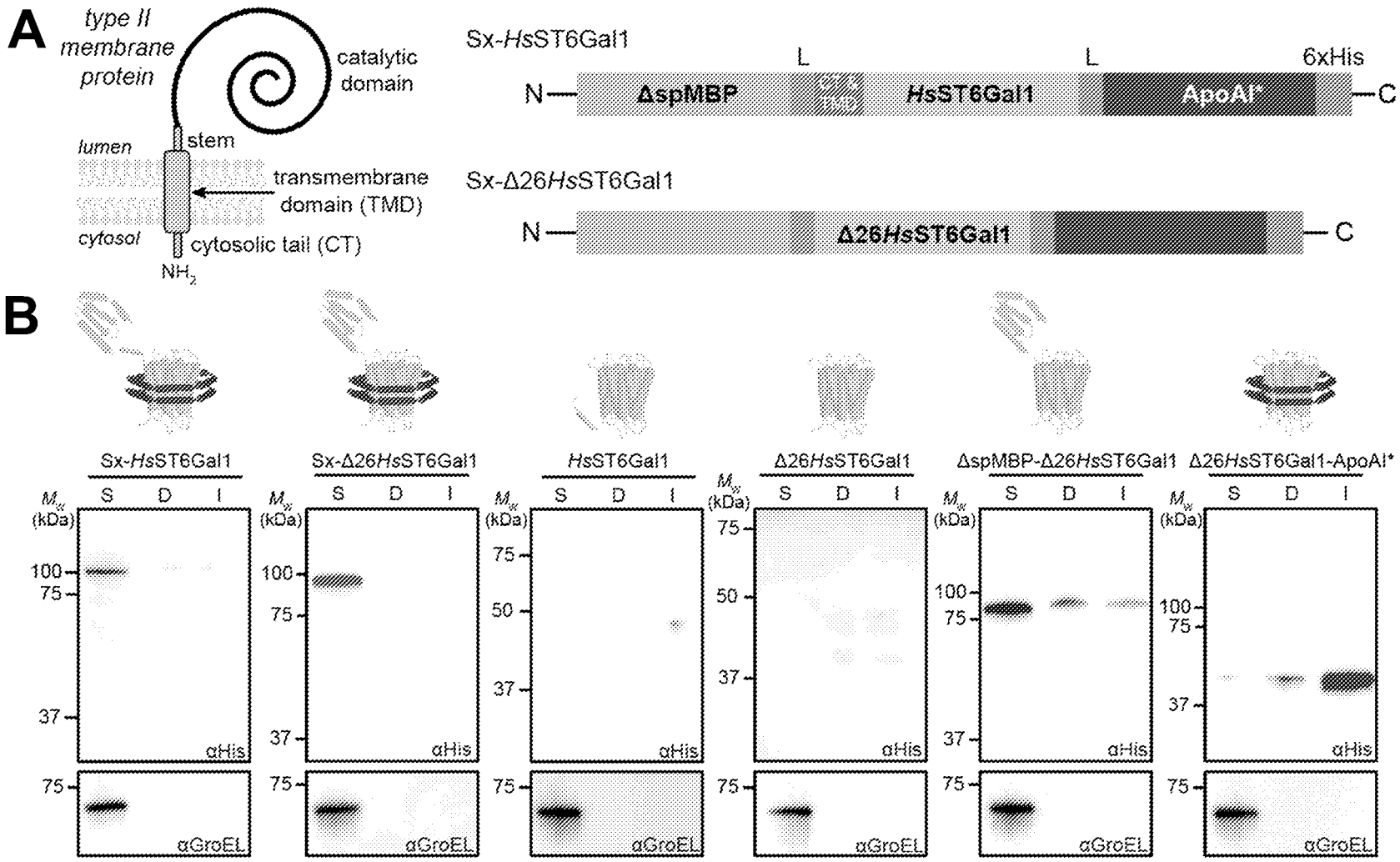
FIG. 2D (continued)

Name	Sequence – truncated (FASTA)	SEQ ID NO.
	RPFYYGLCTPCQAPTNYSRWVNLPEESLLRPAYVVPWQDPWEPFYVAGGKVPTFDERFRQYGFNRISQACELHVAGFDFEVL NEGFLVHKGFKEALKFHPQKEAENQHNKILYRQFKQELKAKYPNSPRRC	
HsUGT1A3	MGKVLVVPIDGSHWLSMREVLRELHARGHQAVVLTPEVNMHIKEENFFTLTTYAISWTQDEFDRHVLGHTQLYFETEHFLKKFF RSMAMLNNMSLVYHRSCVELLHNEALIRHLNATSFDVVLTDPVNLCAAVLAKYLSIPTVFFLRNIPCDLDFKGTQCPNPSSYIPRLL TTNSDHMTFMQRVKNMLYPLALSYICHAFSAPYASLASELFQREVSVVDILSHASVWLFRGDFVMDYPRPIMPNMVFIGGINCAN RKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRA FITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPV EPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH	166
ScAlg1	MGNKSTKKRIIIFVLGDVGHSPRICYHAISFSKLGWQVELCGYVEDTLPKIISSDPNITVHHMSNLKRKGGGTSVIFMVKKVLFQVL SIFKLLWELRGSDYILVQNPPSIPILPIAVLYKLTGCKLIIDWHNLAYSILQLKFKGNFYHPLVLISYMVEMIFSKFADYNLTVTEAMRK YLIQSFHLNPKRCAVLYDRPASQFQPLAGDISRQKALTTKAFIKDYIRDDFDTEKGDKIIVTSTSFTPDEDIGILLGALKIYENSYVKF DSSLPKILCFITGKGPLKEKYMKQVEEYDWKRCQIEFVWLSAEDYPKLLQLCDYGVSLHTSSSGLDLPMKILDMFGSGLPVIAMN YPVLDELVQHNVNGLKFVDRRELHESLIFAMKDADLYQKLKKNVTQEAENRWQSNWERTMRGLKLIH	167
ScAlg11	MYSSLNPFSKKSSLLNRAVASCGEKNVKVFGFFHPYCNAGGGGEKVLWKAVDITLRKDAKNVIVIYSGDFVNGENVTPENILNNV KAKFDYDLDSDRIFFISLKLRYLVDSSTWKHFTLIGQAIGSMILAFESIIQCPPDIWIDTMGYPFSYPIIARFLRRIPIVTYTHYPIMSKD MLNKLFKMPKKGIKVYGKILYWKVFMLIYQSIGSKIDIVITNSTWTNNHIKQIWQSNTCKIIYPPCSTEKLVDWKQKFGTAKGERLN QAIVLAQFRPEKRHKLIIESFATFLKNLPDSVSPIKLIMAGSTRSKQDENYVKSLQDWSENVLKIPKHLISFEKNLPFDKIEILLNKST FGVNAMWNEHFGIAVVEYMASGLIPIVHASAGPLLDIVTPWDANGNIGKAPPQWELQKKYFAKLEDDGETTGFFFKEPSDPDYN TTKDPLRYPNLSDLFLQITKLDYDCLRVMGARNQQYSLYKFSDLKFDKDWENFVLNPICKLLEEEERG	168
NtGnTI	ATQSEYADRLAAAIEAENHCTSQTRLLIDQISLQQGRIVALEEQMKRQDQECRQLRALVQDLESKGIKKLIGNVQMPVAAVVVMA CNRADYLEKTIKSILKYQISVASKYPLFISQDGSHPDVRKLALSYDQLTYMQHLDFEPVHTERPGELIAYYKIARHYKWALDQLFYK HNFSRVIILEDDMEIAPDFFDFFEAGATLLDRDKSIMAISSWNDNGQMQFVQDPYALYRSDFFPGLGWMLSKSTWDELSPKWPK AYWDDWLRLKENHRGRQFIRPEVCRTYNFGEHGSSLGQFFKQYLEPIKLNDVQVDWKSMDLSYLLEDNYVKHFGDLVKKAKPI HGADAVLKAFNIDGDVRIQYRDQLDFENIARQFGIFEEWKDGVPRAAYKGIVVFRYQTSRRVFLVGHDSLQQLGIEDT	169
NtGnTII	RTNTVSNLINSYTNESSESFESIEEAKVSNTHLTRIPKLPLQNELSILLSKRNQMPPRNLNLYSKLAKDHVVIVLYVHNRPQYLQIVI DSLSRVEGISETLLIVSHDGYFEEMNKIVESIKFCQVKQIFAPYSPHIFSDSFPGVSPKDCKEKDDPVEKHCEGMPDQYGNHRSP KIVSLKHHWWWMMNTVWDGLEVTRHHSGHILFIEEDHFIYPNAYRNMQLLAELKSKKCPDCYAANLAPSEVKLRGEGWECLVA ERMGNVGYAFNRTVWRKIHKKAAEFCSFDDYNWDITMWSTVYPSFGSPVYTLRGSRTSAVHFGKCGLHQGHNRNVACMDNG GVNIVVDEIDKVANIKPEWEVRKYEHQAGYKAGFKGWGGWGDVRDRELCMEFAKMYI	170

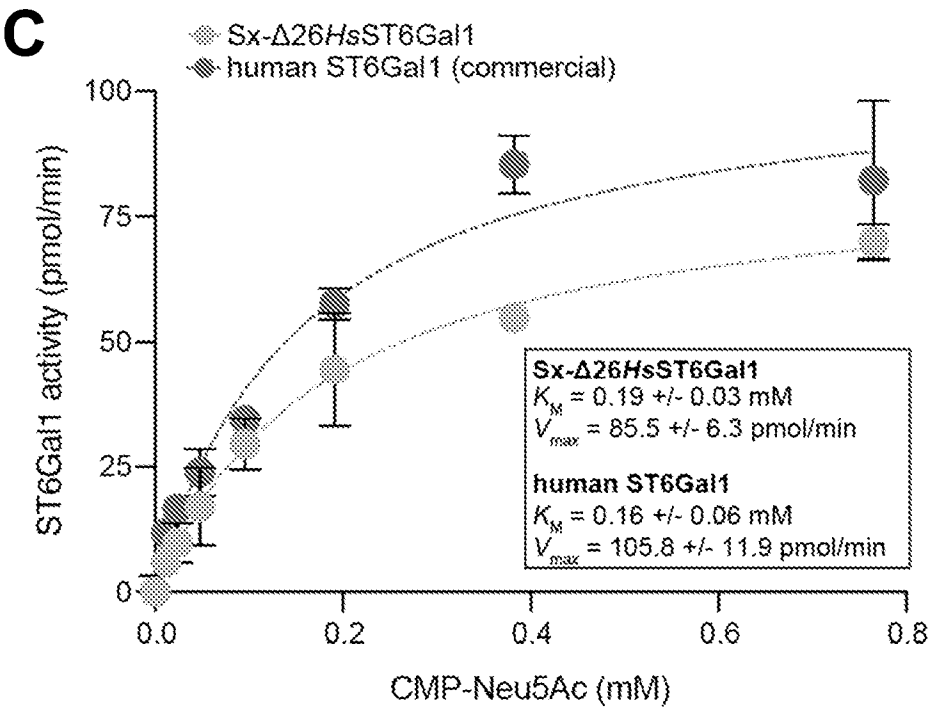
FIG. 2D (continued)

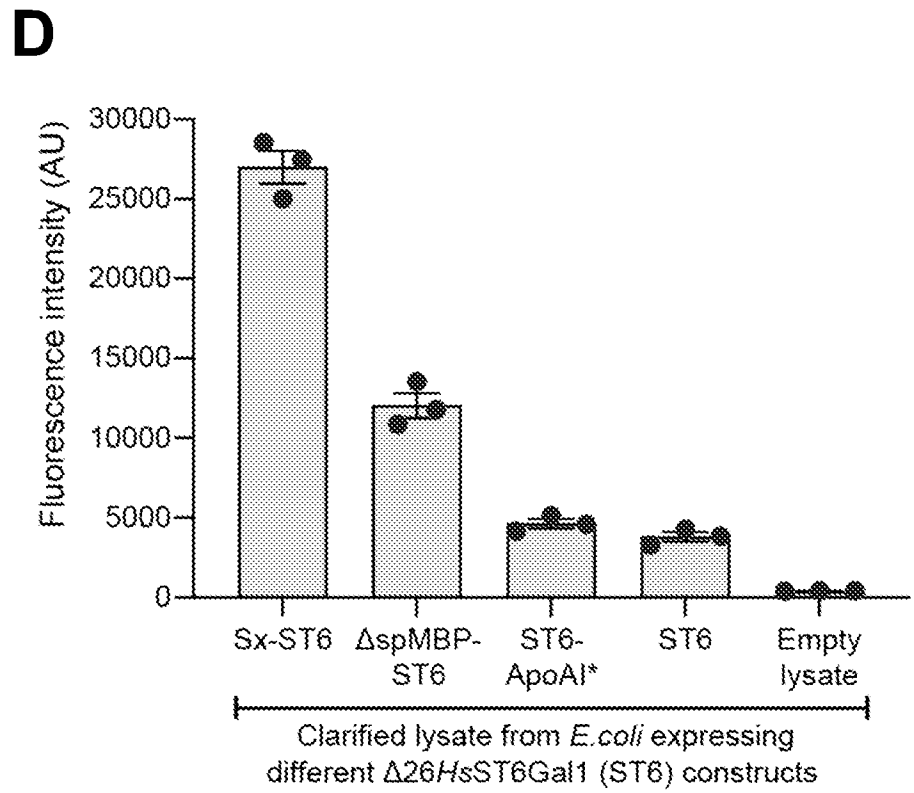
Name	Sequence – truncated (FASTA)	SEQ ID NO.
BtGGTA1	EYIHSPEGSLFWINPSRNPEVGGSSIQKGWWLPRWFNNGYHEEDGDINEEKEQRNEDESKLKLSDWFNPFKRPEVVTMTKWK APVVWEGTYNRAVLDNYYAKQKITVGLTVFAVGRYIEHYLEEFLTSANKHFMVGHPVIFYIMVDDVSRMPLIELGPLRSFKVFKIK PEKRWQDISMMRMKTIGEHIVAHIQHEVDFLFCMDVDQVFQDKFGVETLGESVAQLQAWWYKADPNDFTYERRKESAAYIPFG EGDFYYHAAIFGGTPTQVLNITQECFKGILKDKKNDIEAQWHDESHLNKYFLLNKPTKILSPEYCWDYHIGLPADIKLVKMSWQTK EYNVVRNNV	171
MmGGTA1	NRIPEVGENRWQKDWWFPSWFKNGTHSYQEDNVEGRREKGRNGDRIEEPQLWDWFNPKNRPDVLTVTPWKAPIVWEGTYD TALLEKYYATQKLTVGLTVFAVGKYIEHYLEDFLESADMYFMVGHRVIFYVMIDDTSRMPVVHLNPLHSLQVFEIRSEKRWQDISM MRMKTIGEHILAHIQHEVDFLFCMDVDQVFQDNFGVETLGQLVAQLQAWWYKASPEKFTYERRELSAAYIPFGEGDFYYHAAIF GGTPTHILNLTRECFKGILQDKKHDIEAQWHDESHLNKYFLFNKPTKILSPEYCWDYQIGLPSDIKSVKVAWQTKEYNLVRNNV	172
RnGGTA1	RTHSYQEEDIERAREKGRNGDSIVEPQLWDWFNPKNRPEVLTVTPWKAPIVWEGTYDTALLEKYYARQKITVGLTVFAVGKYIE HYLEDFLESANKYFMVGHRVIFYVMMDDTSRMPAVHLSPLHSLQVFEIRSEKRWQDISMMRMKTIGEHILDHIQHEVDFLFCMDV DQVFQDNFGVETLGQLVAQLQAWWYKASPDEFTYERRELSAAYIPFGEGDFYYHAAVFGGTPVHILNLTRECFKGILQDKKHDI EAQWHDESHLNKYFLFNKPTKILSPEYCWDYHIGLPSDIKNVKIAWQTKEYNLVRSNV	173
BtB4GalT1	RDLRRLPQLVGVHPPLQGSSHGAAAIGQPSGELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGSNLTSAPVPSTT TRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLGGRYTPMDCISPHKVAIIPFRNRQEHLKYWLYYLHPILQRQQLDYG IYVINQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSK QQFLSINGFPNNYWGWGGEDDDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRFDRIAHTKETMLSDGLNSLTYM VLEVQRYPLYTKITVDIGTPS	174

FIG. 2D (continued)

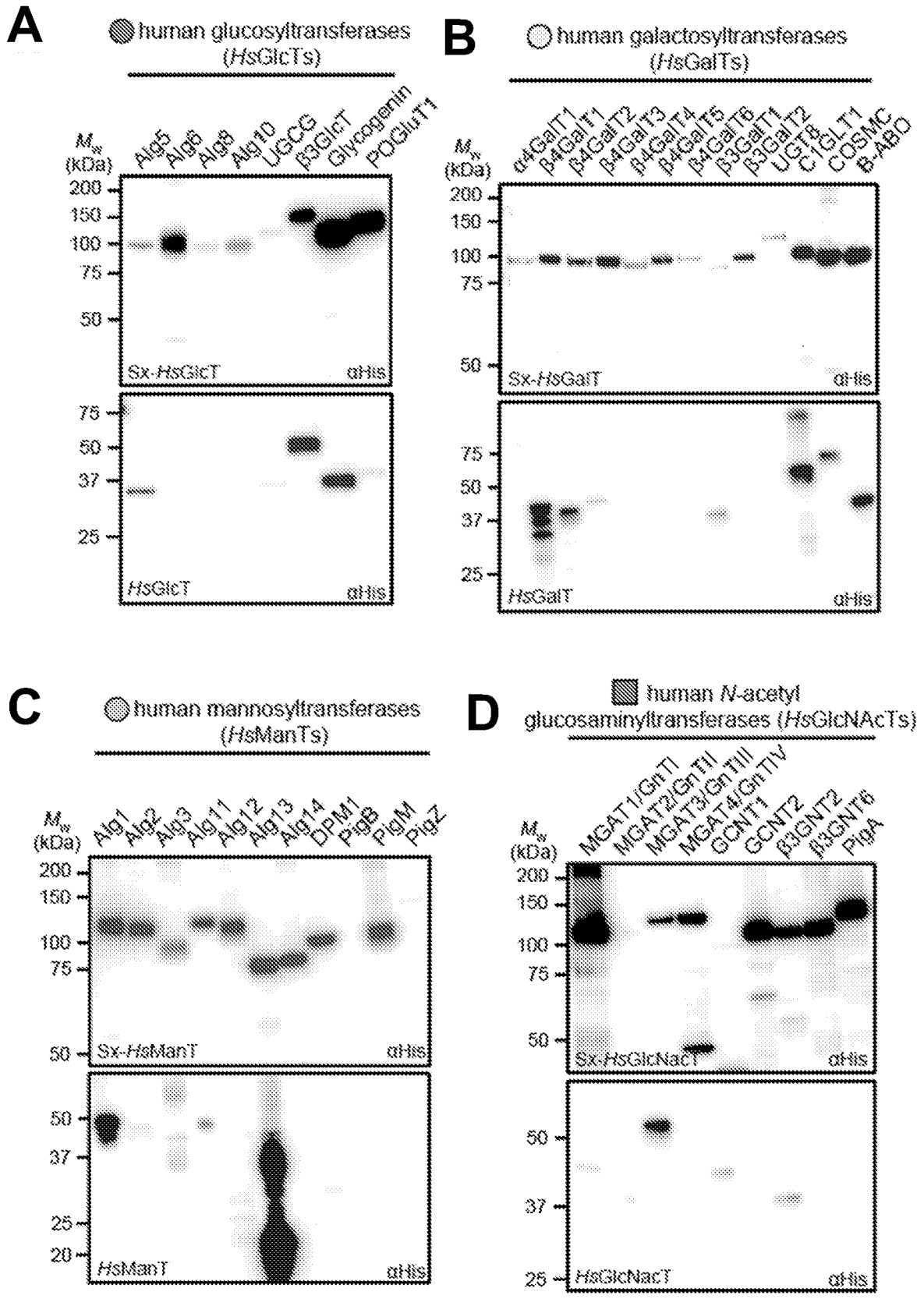

A
type II membrane protein
catalytic domain
stem
lumen
transmembrane domain (TMD)
cytosol
cytosolic tail (CT)
NH2
Sx-HsST6Gal1
N
ΔspMBP
L
HsST6Gal1
L
6xHis
ApoAI
C
Sx-Δ26HsST6Gal1
N
Δ26HsST6Gal1
C
B
Sx-HsST6Gal1
Sx-Δ26HsST6Gal1
HsST6Gal1
Δ26HsST6Gal1
ΔspMBP-Δ26HsST6Gal1
Δ26HsST6Gal1-ApoAI*
S D I
Mw (kDa)
100
75
50
37
αHis
αGroEL

FIGS. 3A–3B

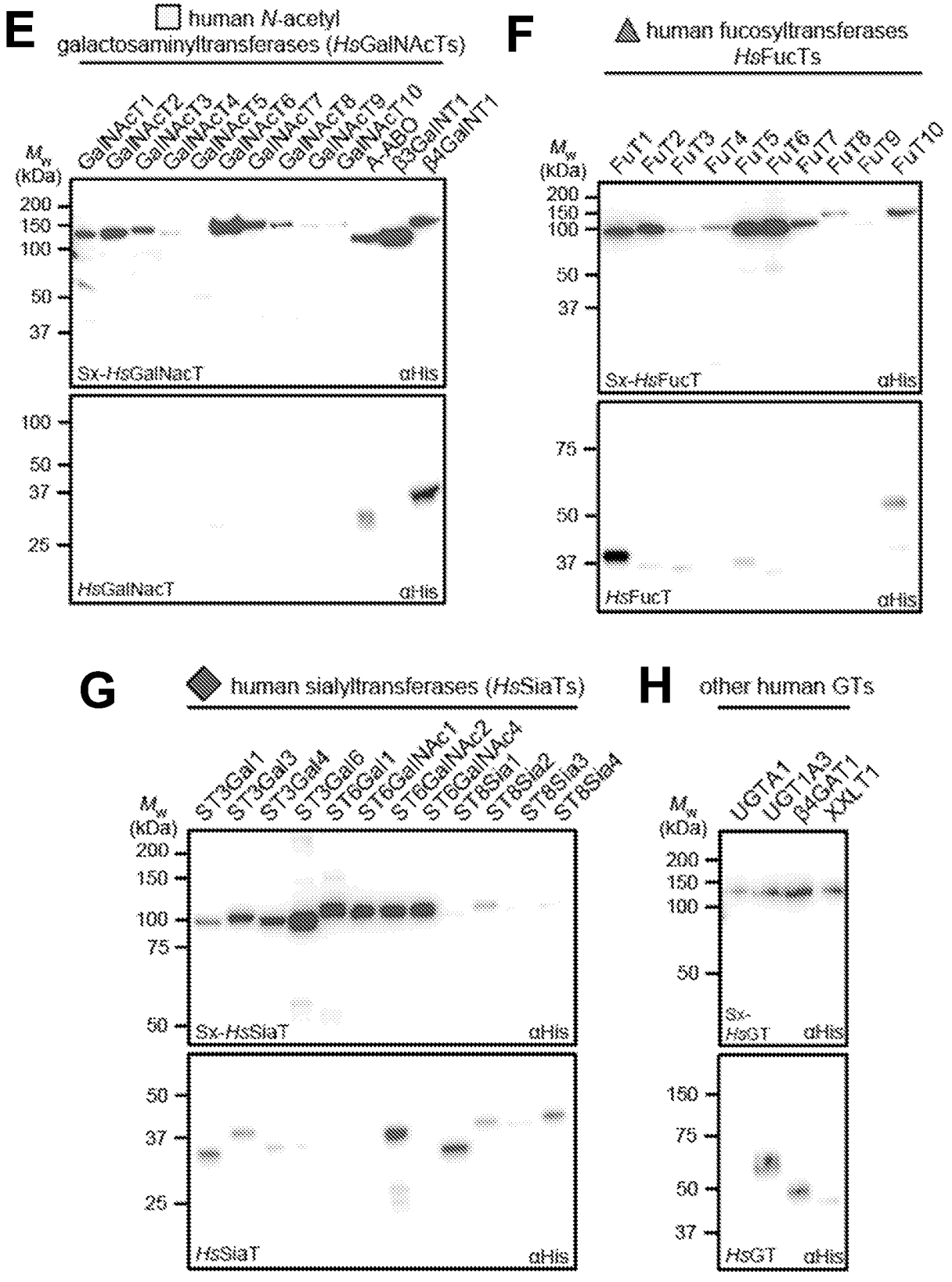

C
Sx-Δ26HsST6Gal1
human ST6Gal1 (commercial)
ST6Gal1 activity (pmol/min)
100
75
50
25
0
0.0
0.2
0.4
0.6
0.8
CMP-Neu5Ac (mM)
Sx-Δ26HsST6Gal1
KM = 0.19 +/- 0.03 mM
Vmax = 85.5 +/- 6.3 pmol/min
human ST6Gal1
KM = 0.16 +/- 0.06 mM
Vmax = 105.8 +/- 11.9 pmol/min


D
Fluorescence intensity (AU)
30000
25000
20000
15000
10000
5000
0
Sx-ST6
ΔspMBP-ST6
ST6-ApoAI*
ST6
Empty lysate
Clarified lysate from E.coli expressing
different Δ26HsST6Gal1 (ST6) constructs

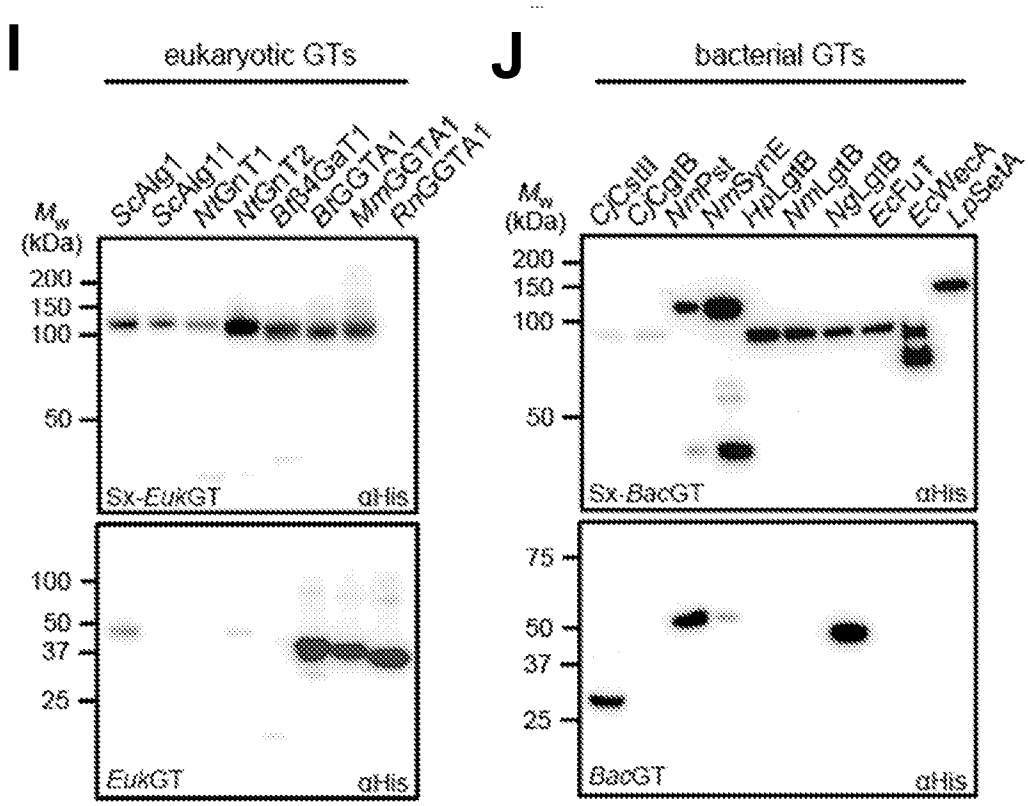
FIGS. 3C–3D


A
human glucosyltransferases (HsGlcTs)
Mw (kDa)
Alg5
Alg6
Alg8
Alg10
UGCG
β3GlcT
Glycogenin
POGLUT1
200
150
100
75
50
Sx-HsGlcT
αHis
75
50
37
25
HsGlcT
αHis
B
human galactosyltransferases (HsGalTs)
Mw (kDa)
α4GalT1
B4GalT1
B4GalT2
B4GalT3
B4GalT4
B4GalT5
B4GalT6
B3GalT1
B3GalT2
UGT8
C1GALT1
COSMC
B-ABO
200
150
100
75
50
Sx-HsGalT
αHis
75
50
37
25
HsGalT
αHis
C
human mannosyltransferases (HsManTs)
Mw (kDa)
Alg1
Alg2
Alg3
Alg11
Alg12
Alg13
Alg14
DPM1
PigB
PigM
PigZ
200
150
100
75
50
Sx-HsManT
αHis
50
37
25
20
HsManT
αHis
D
human N-acetyl glucosaminyltransferases (HsGlcNAcTs)
Mw (kDa)
MGAT1/GnTI
MGAT2/GnTII
MGAT3/GnTIII
MGAT4/GnTIV
GCNT1
GCNT2
β3GNT2
β3GNT6
PigA
200
150
100
75
50
Sx-HsGlcNacT
αHis
50
37
25
HsGlcNacT
αHis

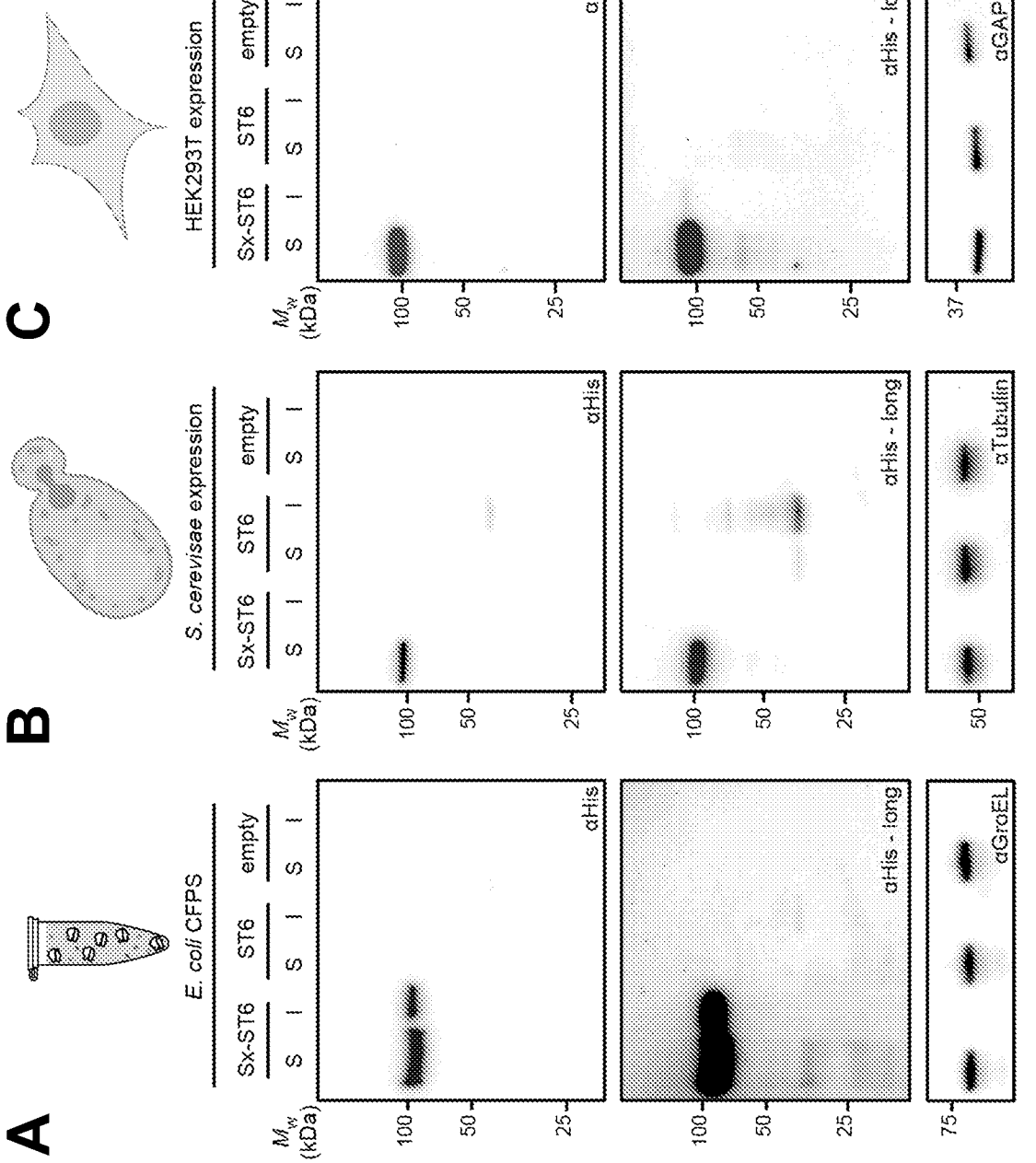
FIGS. 4A–4D


E
human N-acetyl galactosaminyltransferases (HsGalNAcTs)
GalNAcT1
GalNAcT2
GalNAcT3
GalNAcT4
GalNAcT5
GalNAcT6
GalNAcT7
GalNAcT8
GalNAcT9
GalNAcT10
A-ABO
β3Gal/NT1
β4Gal/NT1
Mw (kDa)
200
150
100
50
37
Sx-HsGalNacT
αHis
100
50
37
25
HsGalNacT
αHis
F
human fucosyltransferases HsFucTs
FuT1
FuT2
FuT3
FuT4
FuT5
FuT6
FuT7
FuT8
FuT9
FuT10
Mw (kDa)
200
150
100
50
37
Sx-HsFucT
αHis
75
50
37
HsFucT
αHis
G
human sialyltransferases (HsSiaTs)
ST3Gal1
ST3Gal3
ST3Gal4
ST3Gal6
ST6Gal1
ST6GalNAc1
ST6GalNAc2
ST6GalNAc4
ST8Sia1
ST8Sia2
ST8Sia3
ST8Sia4
Mw (kDa)
200
150
100
75
50
Sx-HsSiaT
αHis
50
37
25
HsSiaT
αHis
H
other human GTs
UGTA1
UGT1A3
β4GAT1
XXLT1
Mw (kDa)
200
150
100
50
Sx-HsGT
αHis
150
75
50
37
HsGT
αHis

FIGS. 4E–4H


I
eukaryotic GTs
ScAlg1
ScAlg11
NtGnT1
NtGnT2
BtB4GaT1
BtGGTA1
MmGGTA1
RnGGTA1
Mw (kDa)
200
150
100
50
Sx-EukGT
αHis
100
50
37
25
EukGT
αHis
J
bacterial GTs
CjCstII
CjCgtB
NmPgt
NmSynE
HpLgtB
NmLgtB
NgLgtB
EcFUT
EcWecA
LpSpsA
Mw (kDa)
200
150
100
50
Sx-BacGT
αHis
75
50
37
25
BacGT
αHis

FIGS. 4I–4J


A
B
C
E. coli CFPS
S. cerevisae expression
HEK293T expression
Sx-ST6
ST6
empty
S
I
S
I
S
I
M_w (kDa)
100
50
25
75
37
αHis
αHis - long
αGroEL
αTubulin
αGAPDH

FIGS. 5A–5C

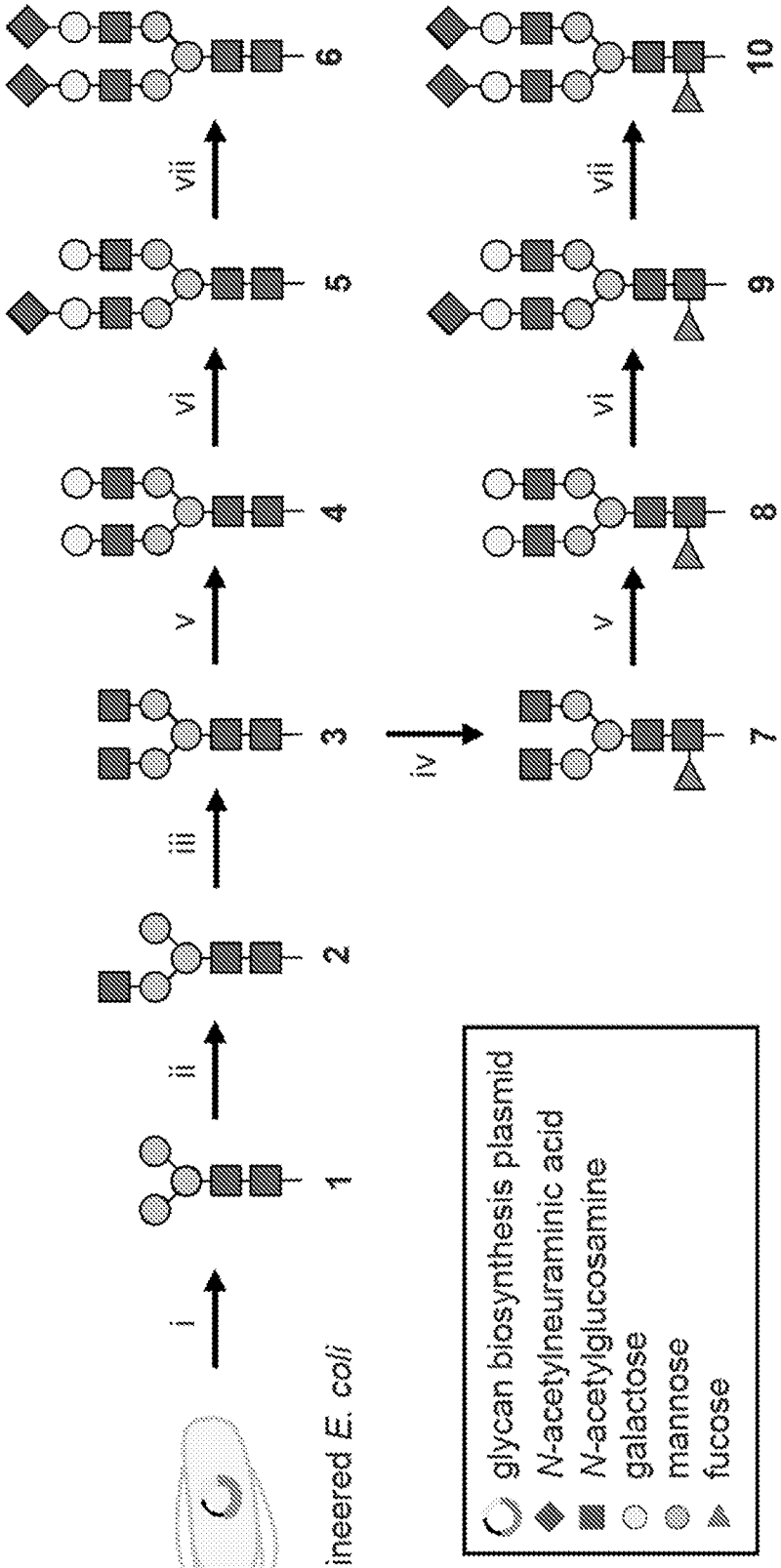

Glycoengineered E. coli
glycan biosynthesis plasmid
N-acetylneuraminic acid
N-acetylglucosamine
galactose
mannose
fucose
i
ii
iii
iv
v
vi
vii
1
2
3
4
5
6
7
8
9
10

FIG. 6A

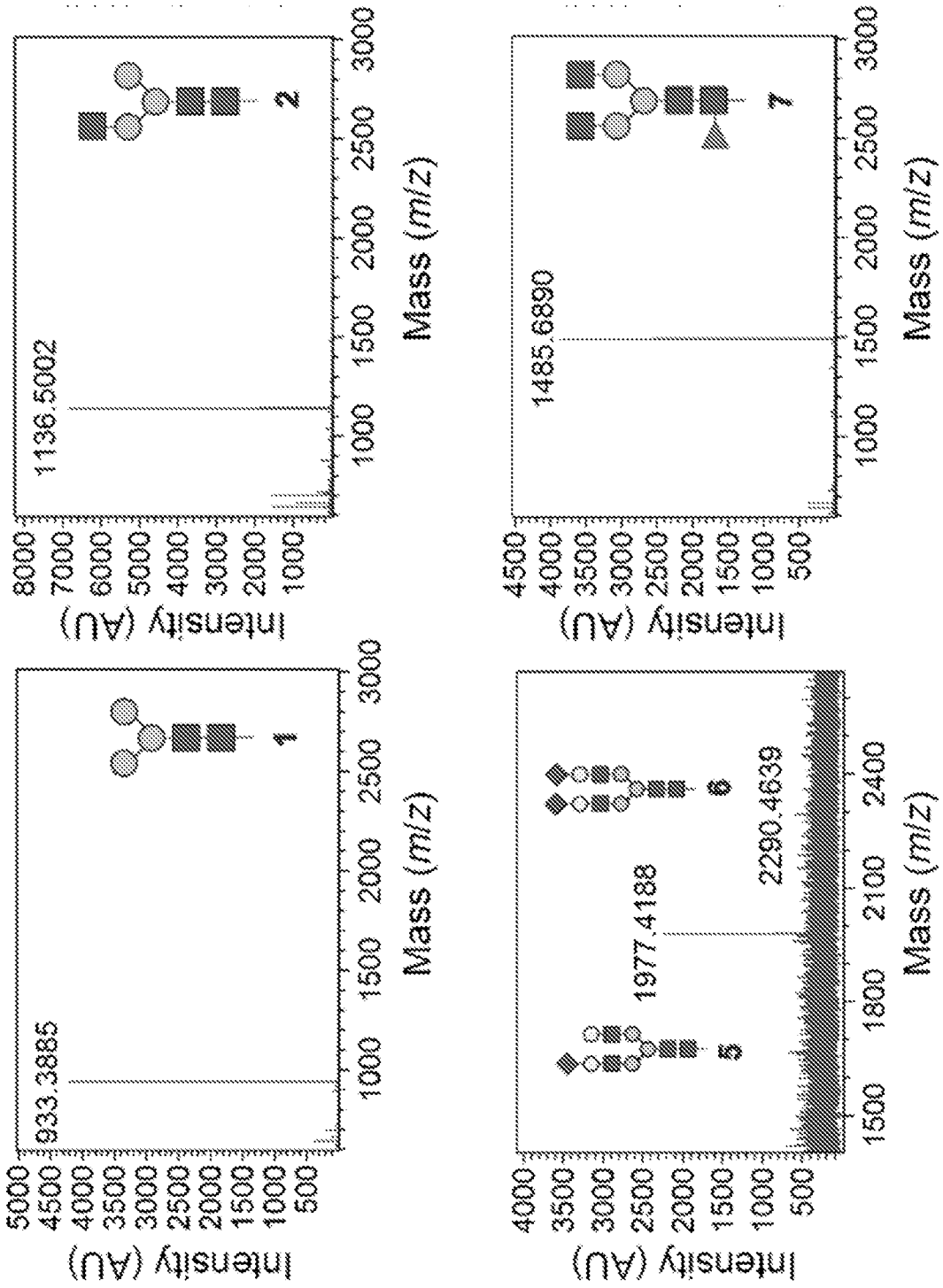

Intensity (AU)
Mass (m/z)
933.3885
1
1136.5002
2
1977.4188
5
6
2290.4639
1485.6890
7

FIG. 6B

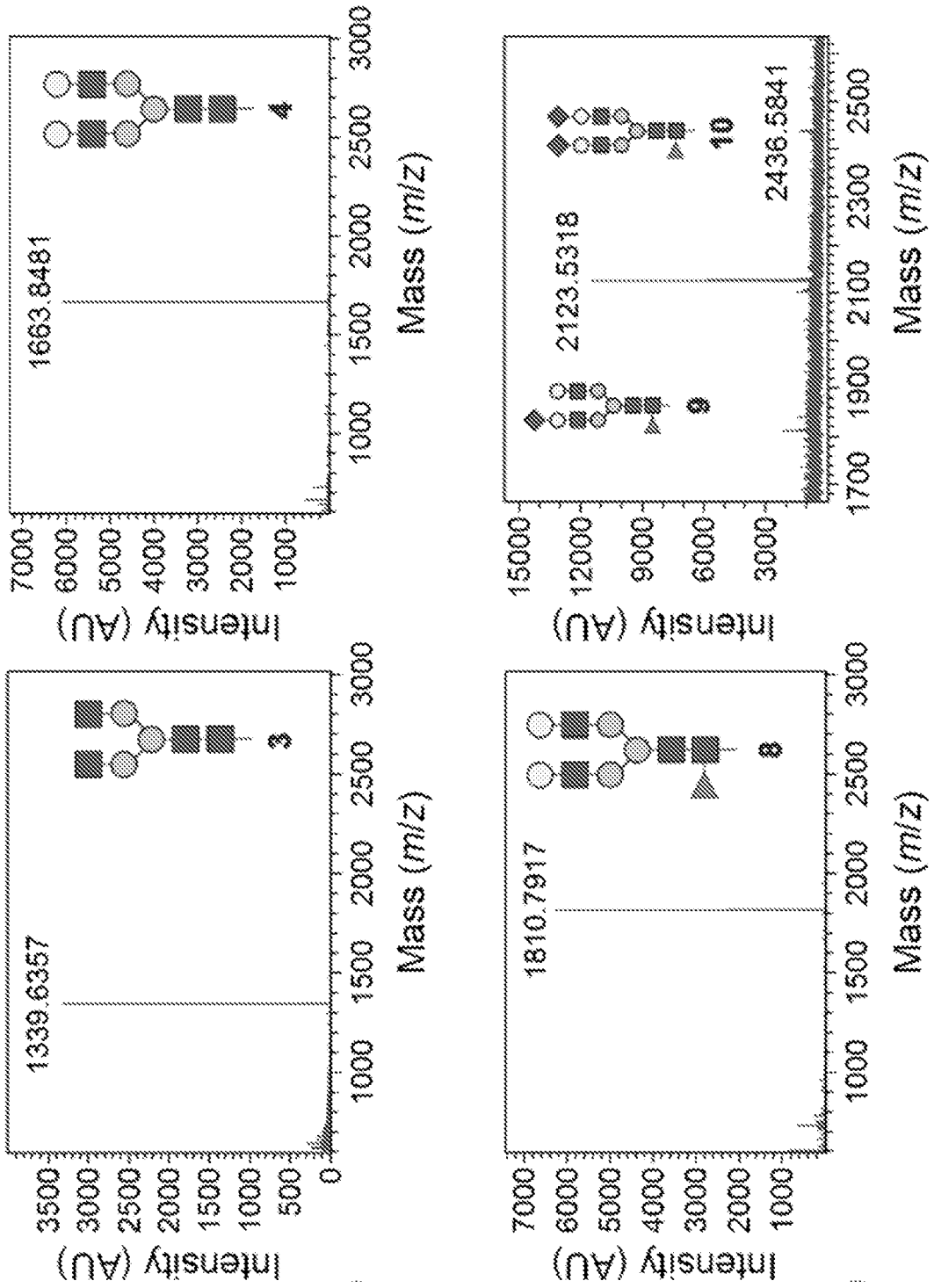


FIG. 6B (continued)

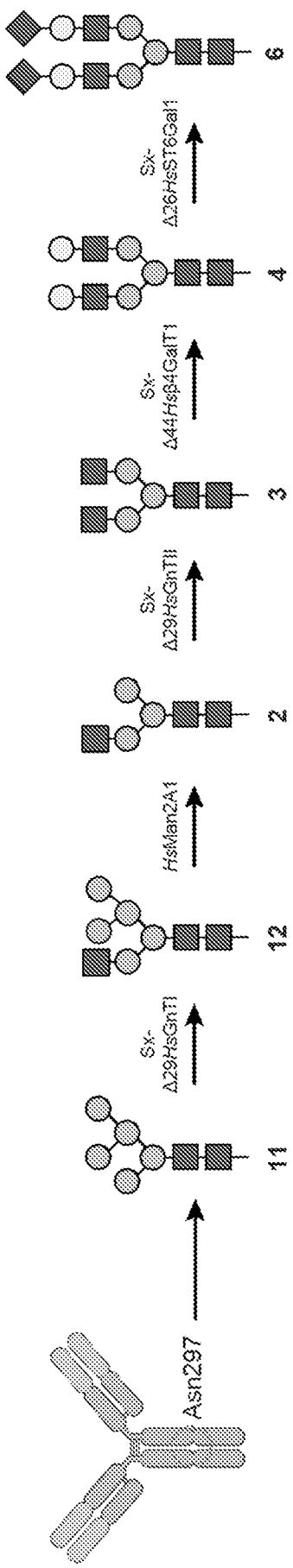

Asn297
11
Sx-
Δ29HsGnTII
12
HsMan2A1
2
Sx-
Δ29HsGnTII
3
Sx-
Δ44HsB4GalT1
4
Sx-
Δ26HsST6Gal1
6

FIG. 7A

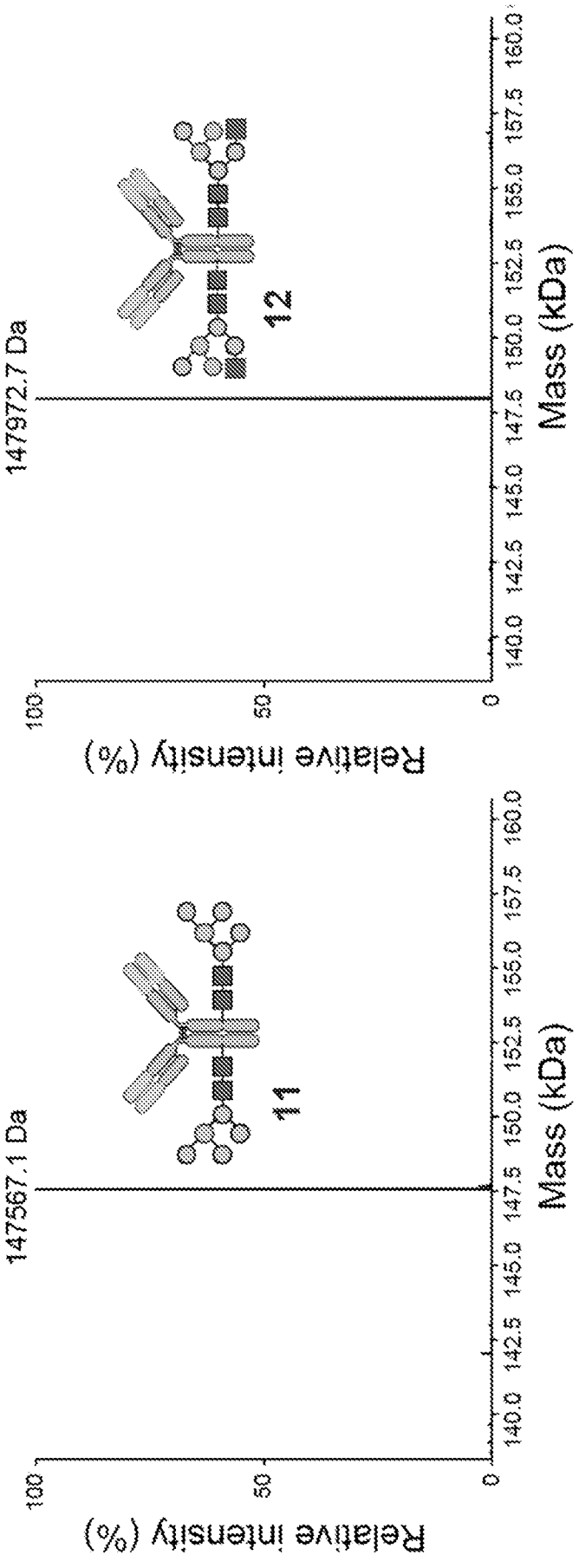

147567.1 Da
11
147972.7 Da
12
Relative intensity (%)
100
50
0
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)

FIG. 7B

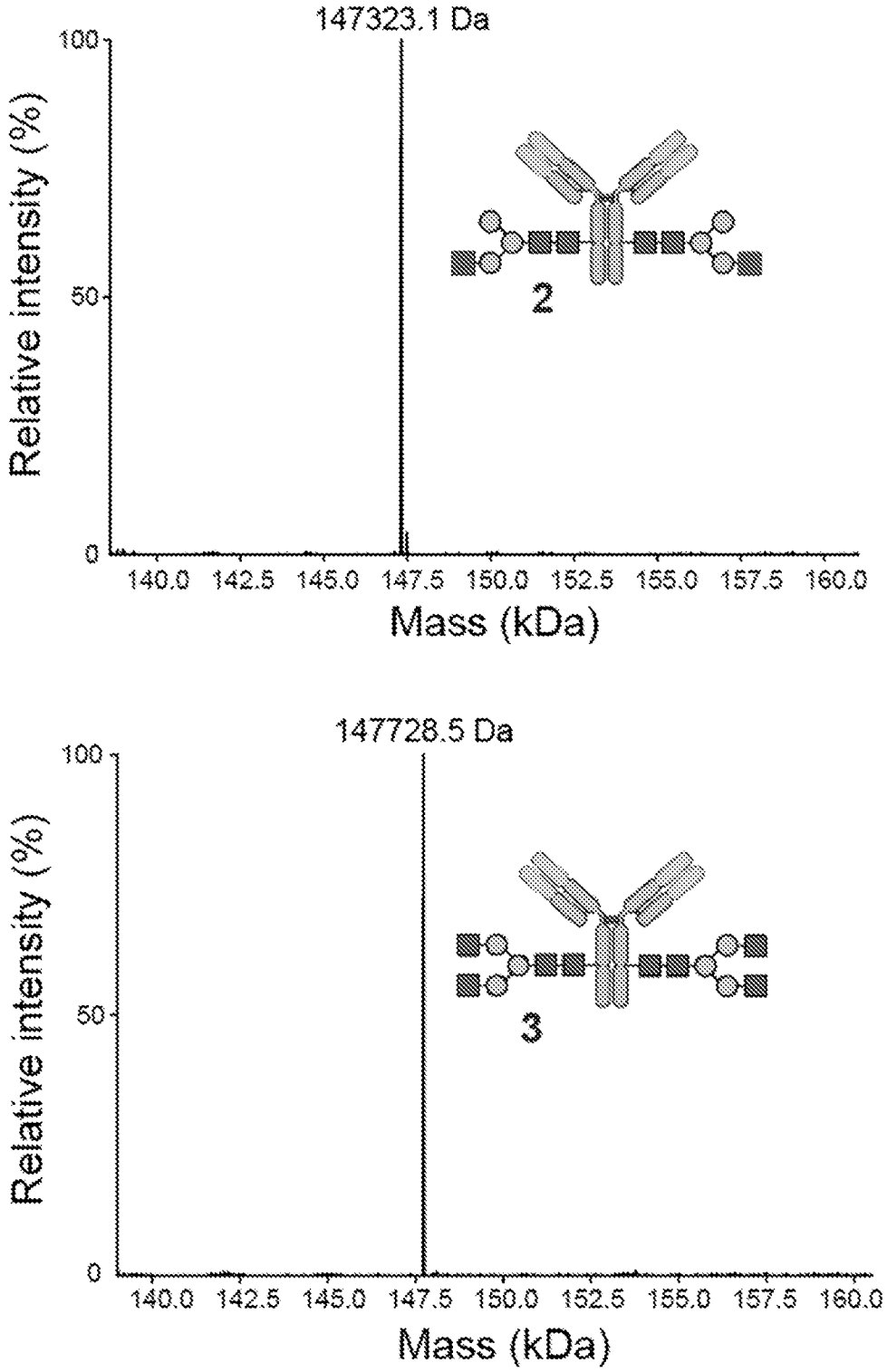


FIG. 7B (continued)

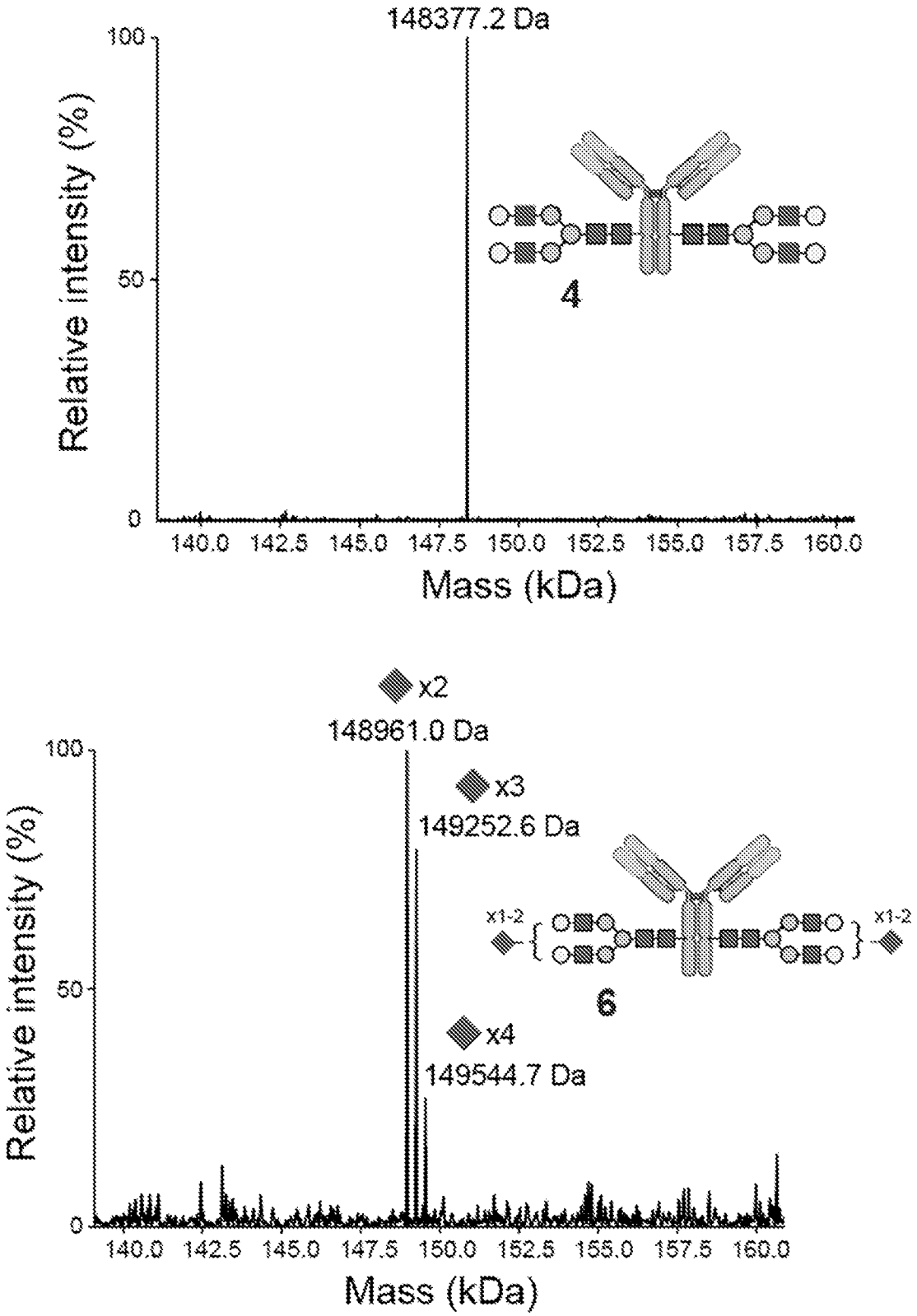


FIG. 7B (continued)

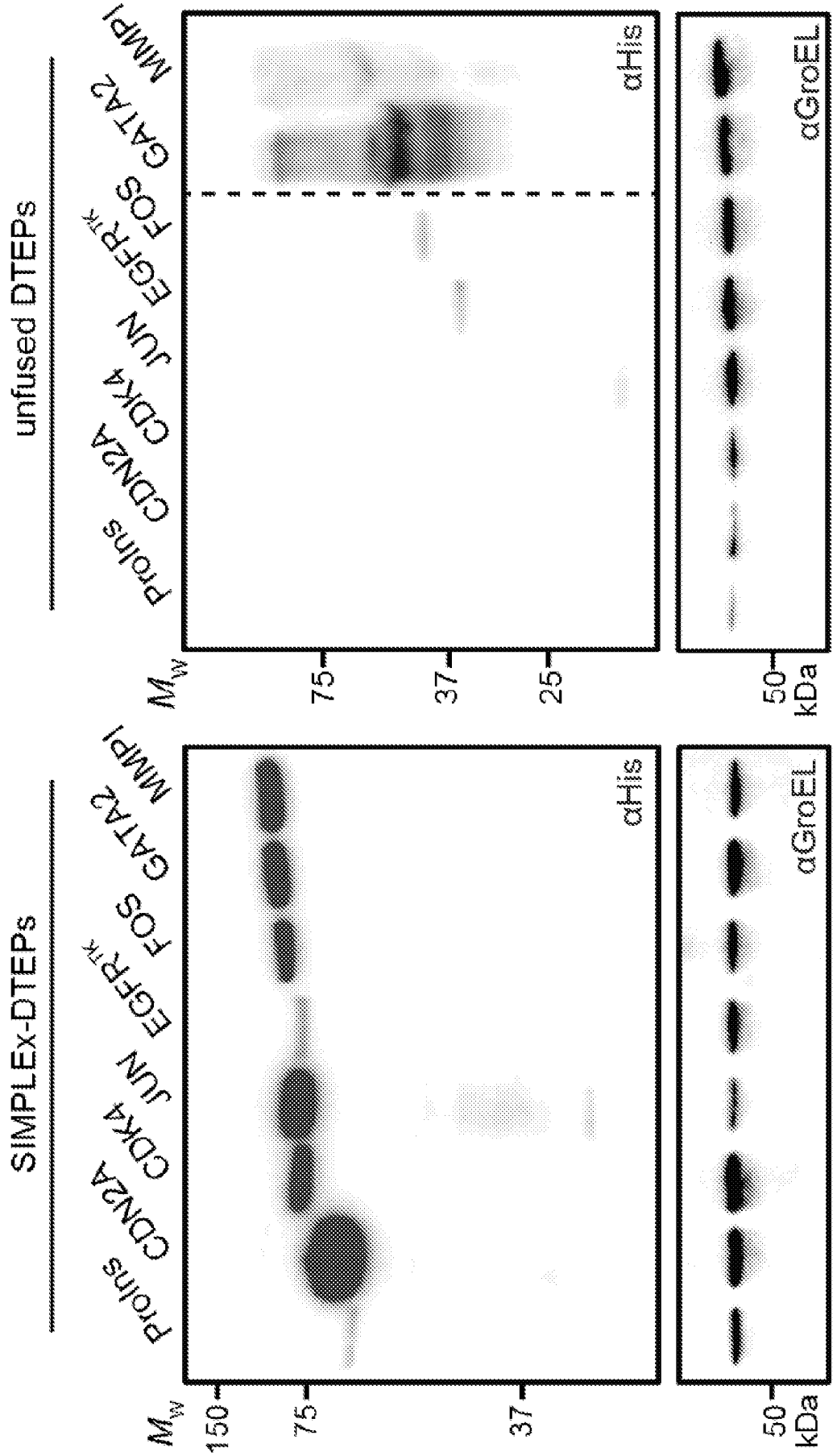

SIMPLEx-DTEPs
unfused DTEPs
Proins CDN2A CDK4 JUN EGFRTK FOS GATA2 MMP1
αHis
αGroEL
MW
150
75
37
25
50
kDa

FIG. 8A

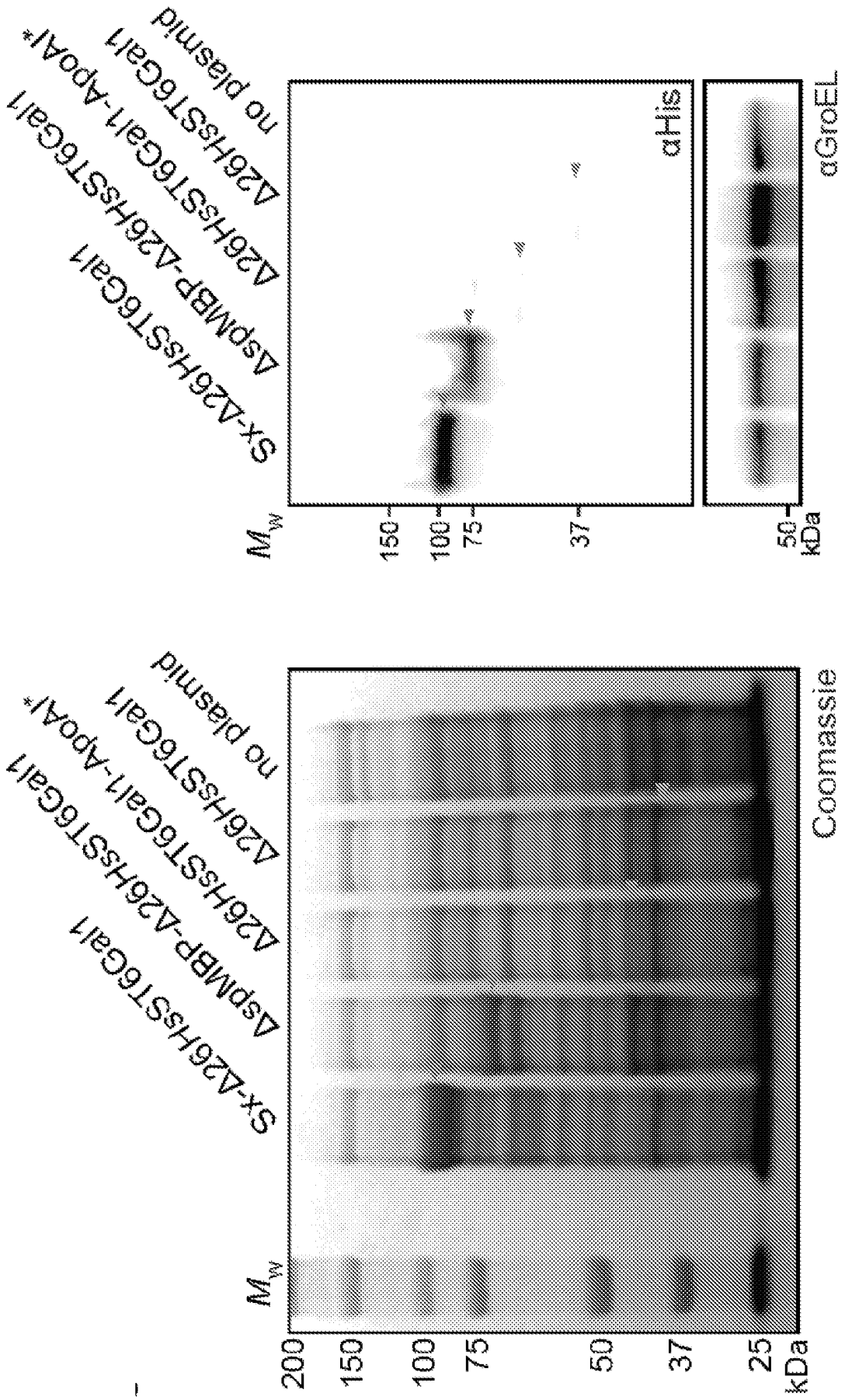
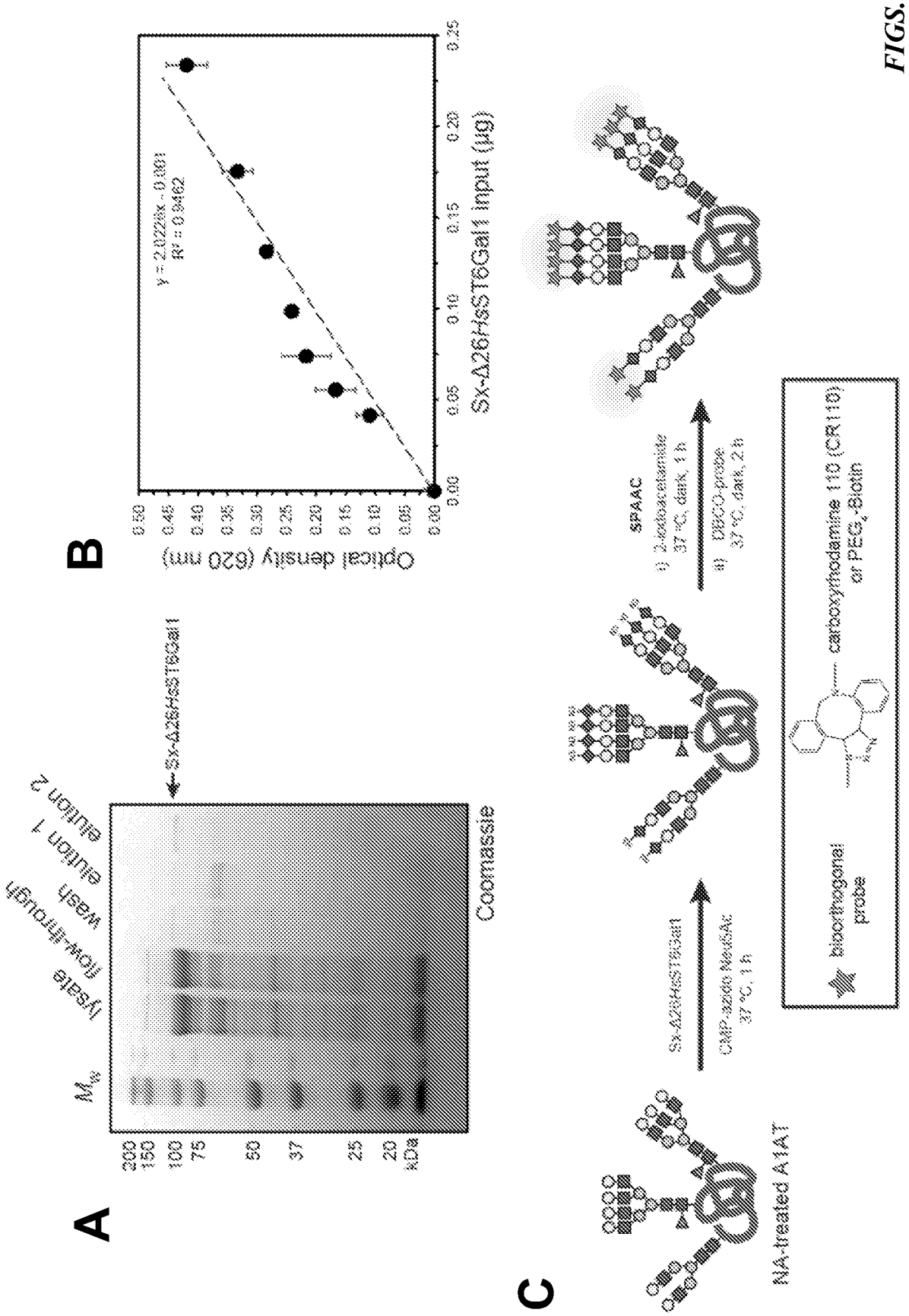
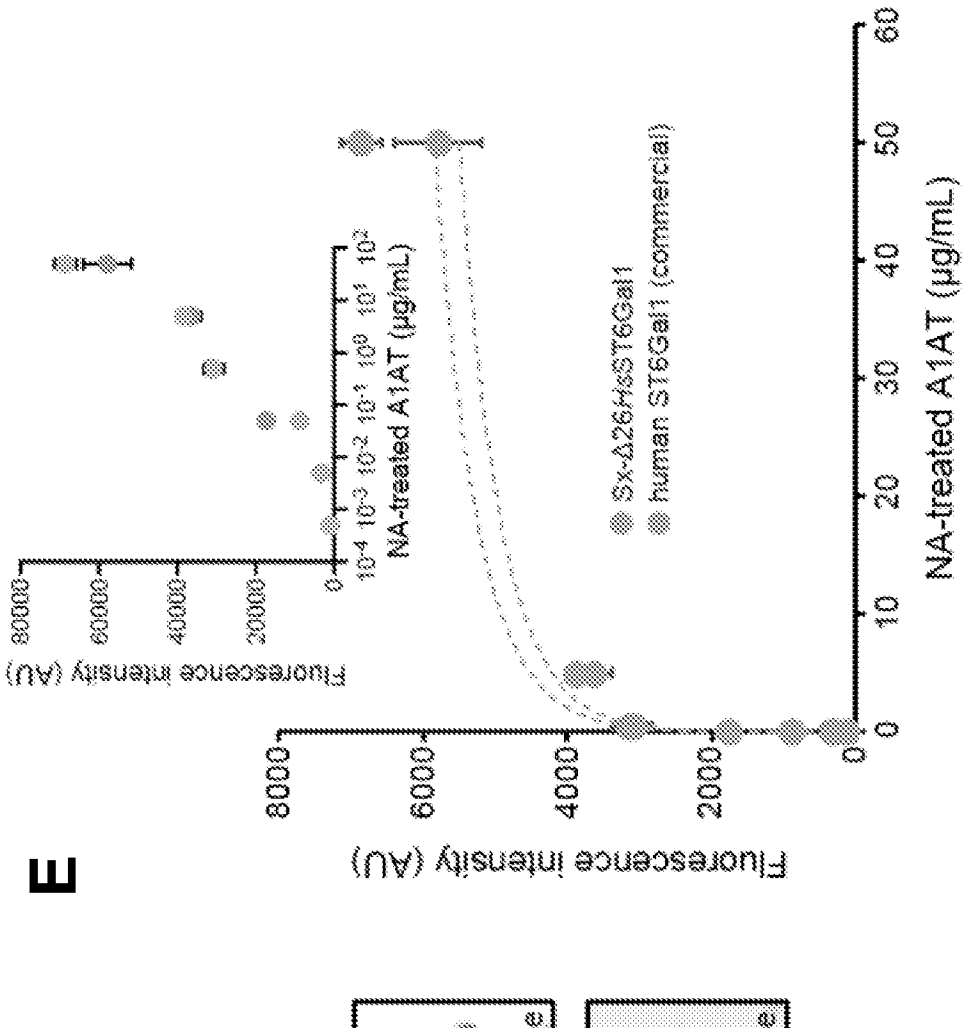

Sx-A26HisSTGGal1
AspMBP-A26HisSTGGal1
A26HisSTGGal1-ApoA1*
A26HisSTGGal1
no plasmid
M_W
200
150
100
75
50
37
25
kDa
Coomassie
αHis
αGroEL

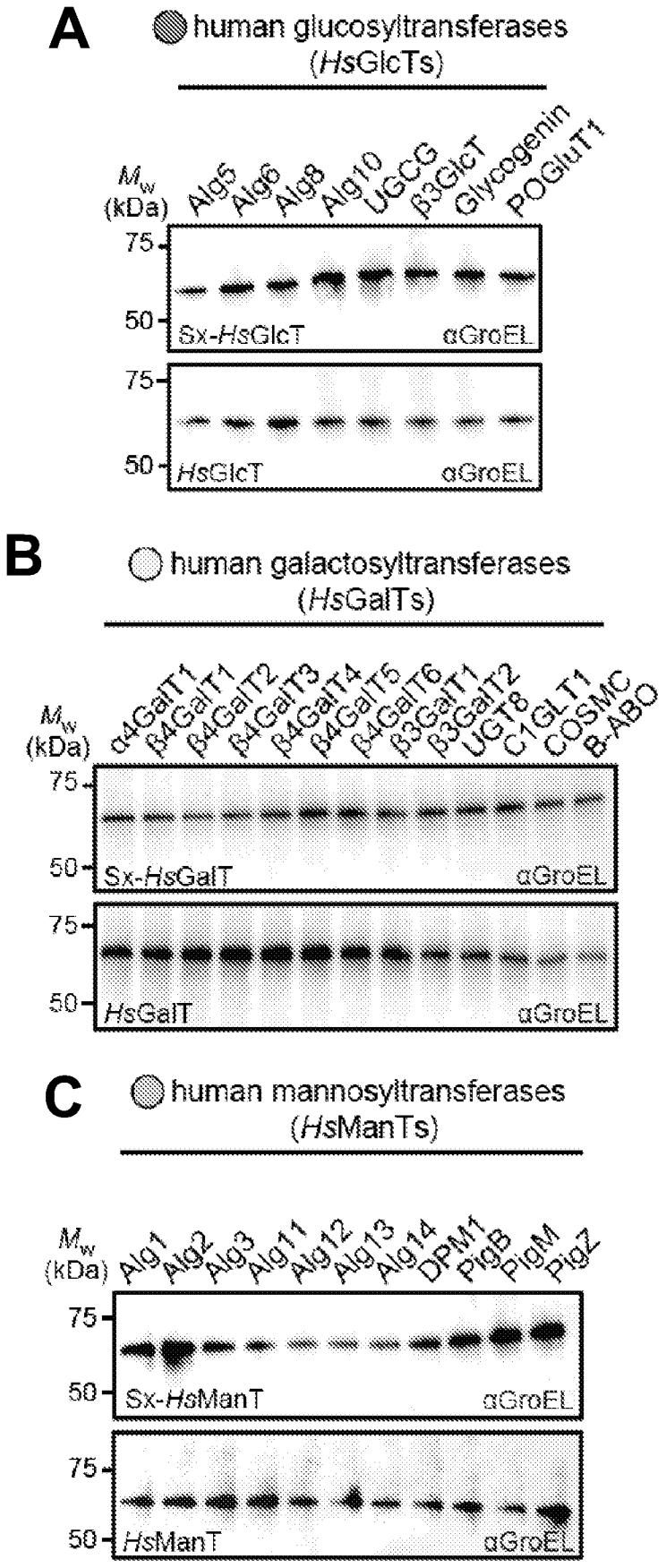
FIG. 8B


A
Mw
lysate
flow-through
wash
elution 1
elution 2
Sx-Δ26HsST6Gal1
Coomassie
200
150
100
75
50
37
25
20
kDa
B
Optical density (620 nm)
Sx-Δ26HsST6Gal1 input (µg)
y = 2.0226x - 0.0301
R² = 0.9462
C
NA-treated A1AT
Sx-Δ26HsST6Gal1
CMP-azido NeuSAc
37 °C, 1 h
SPAAC
i) 2-azidoacetamide
37 °C, dark, 1 h
ii) DBCO-probe
37 °C, dark, 2 h
biorthogonal probe
carboxyrhodamine 110 (CR110)
or PEG4-Biotin

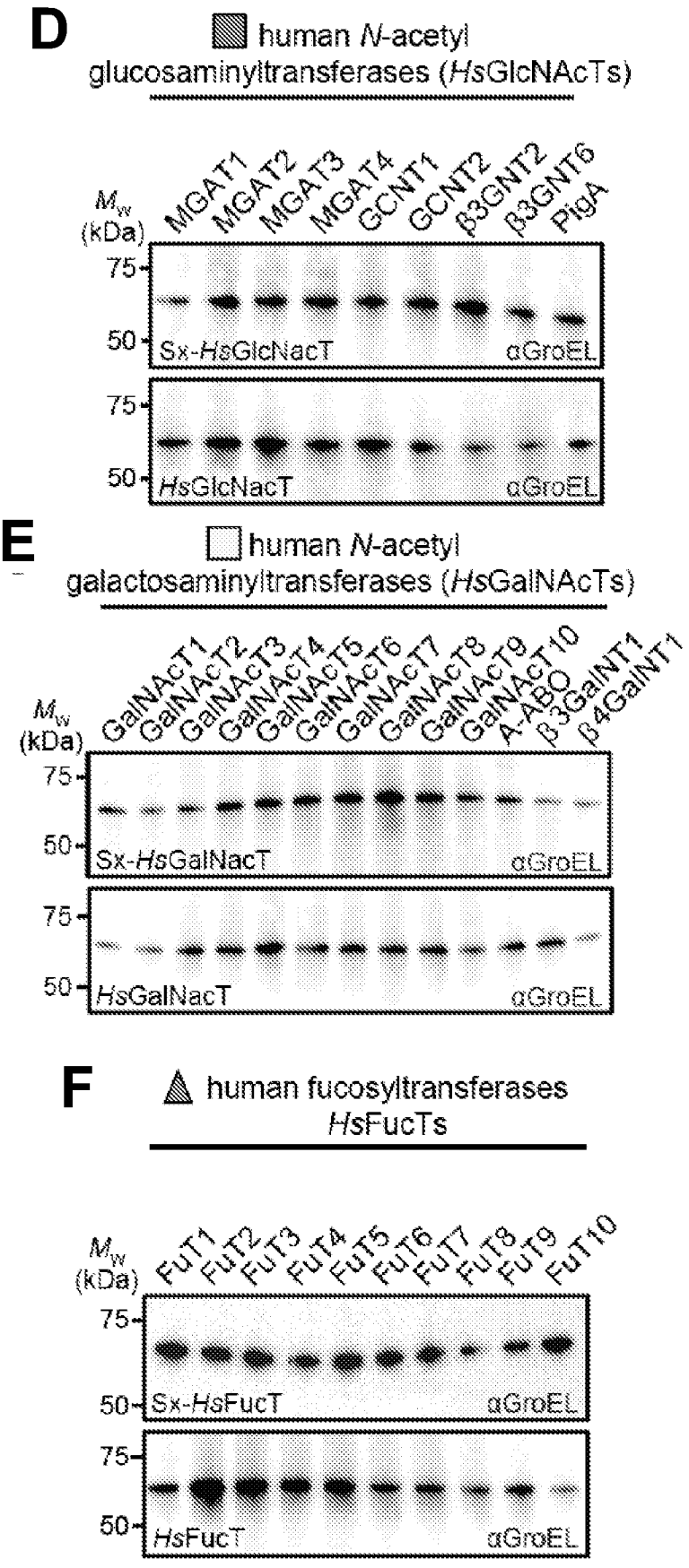
FIGS. 9A–9C


D
purified
Sx-Δ26HsST6Gal1
NA-treated A1AT
CMP-azido Neu5Ac
CMP-Neu5Ac
Mw
50
Coomassie
In-gel fluorescence
50
kDa
E
Fluorescence intensity (AU)
8000
6000
4000
2000
0
0
10
20
30
40
50
60
NA-treated A1AT (μg/mL)
Sx-Δ26HsST6Gal1
human ST6Gal1 (commercial)
Fluorescence intensity (AU)
80000
60000
40000
20000
0
10-4 10-3 10-2 10-1 100 101 102
NA-treated A1AT (μg/mL)

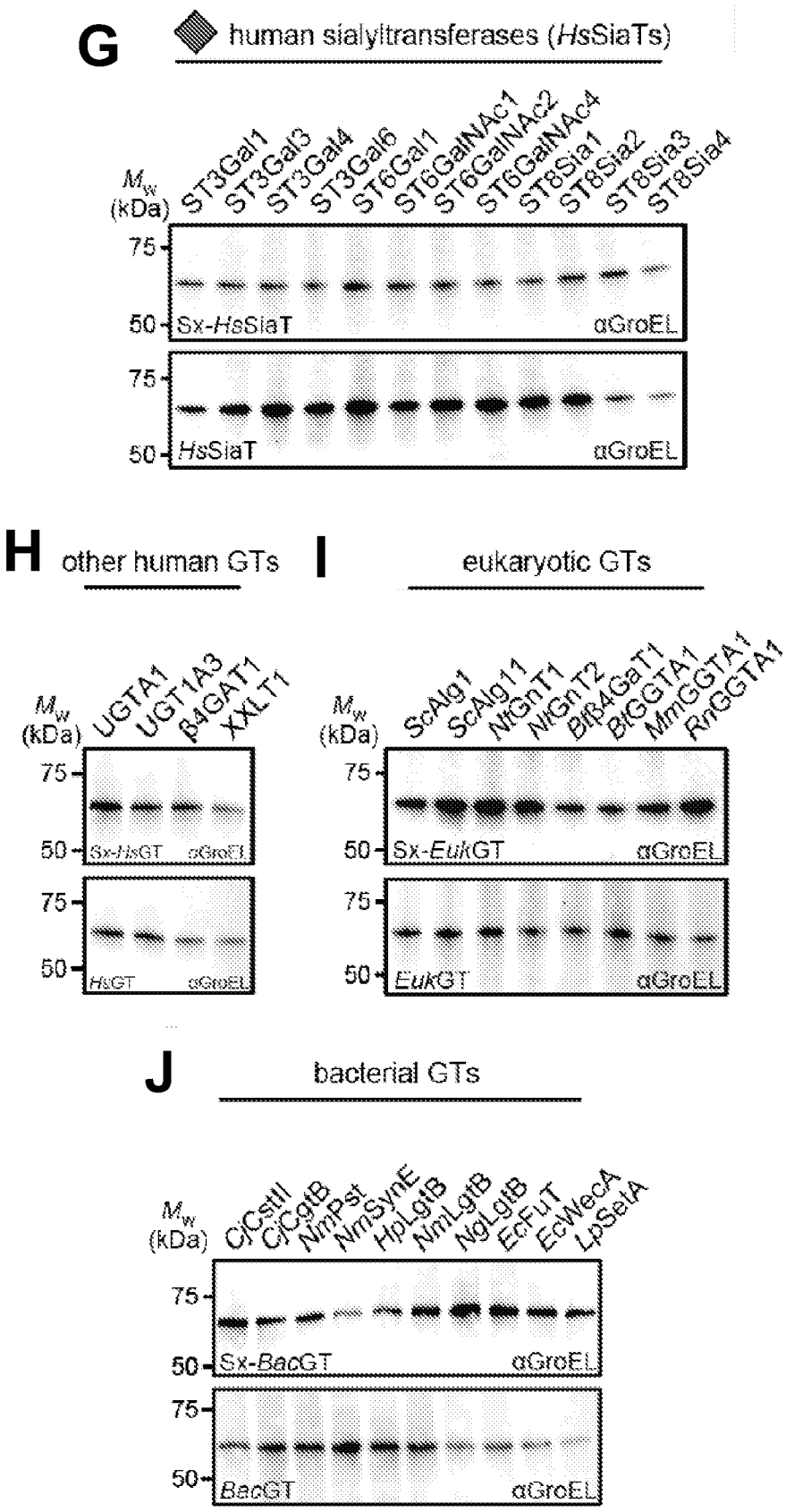
FIGS. 9D–9E


A
human glucosyltransferases
(HsGlcTs)
Mw (kDa)
Alg5
Alg6
Alg8
Alg10
UGCG
β3GlcT
Glycogenin
POGluT1
75
50
Sx-HsGlcT
αGroEL
75
50
HsGlcT
αGroEL
B
human galactosyltransferases
(HsGalTs)
Mw (kDa)
α4GalT1
β4GalT1
β4GalT2
β4GalT3
β4GalT4
β4GalT5
β4GalT6
β3GalT1
β3GalT2
UGT8
C1GLT1
COSMC
B-ABO
75
50
Sx-HsGalT
αGroEL
75
50
HsGalT
αGroEL
C
human mannosyltransferases
(HsManTs)
Mw (kDa)
Alg1
Alg2
Alg3
Alg11
Alg12
Alg13
Alg14
DPM1
PigB
PigM
PigZ
75
50
Sx-HsManT
αGroEL
75
50
HsManT
αGroEL

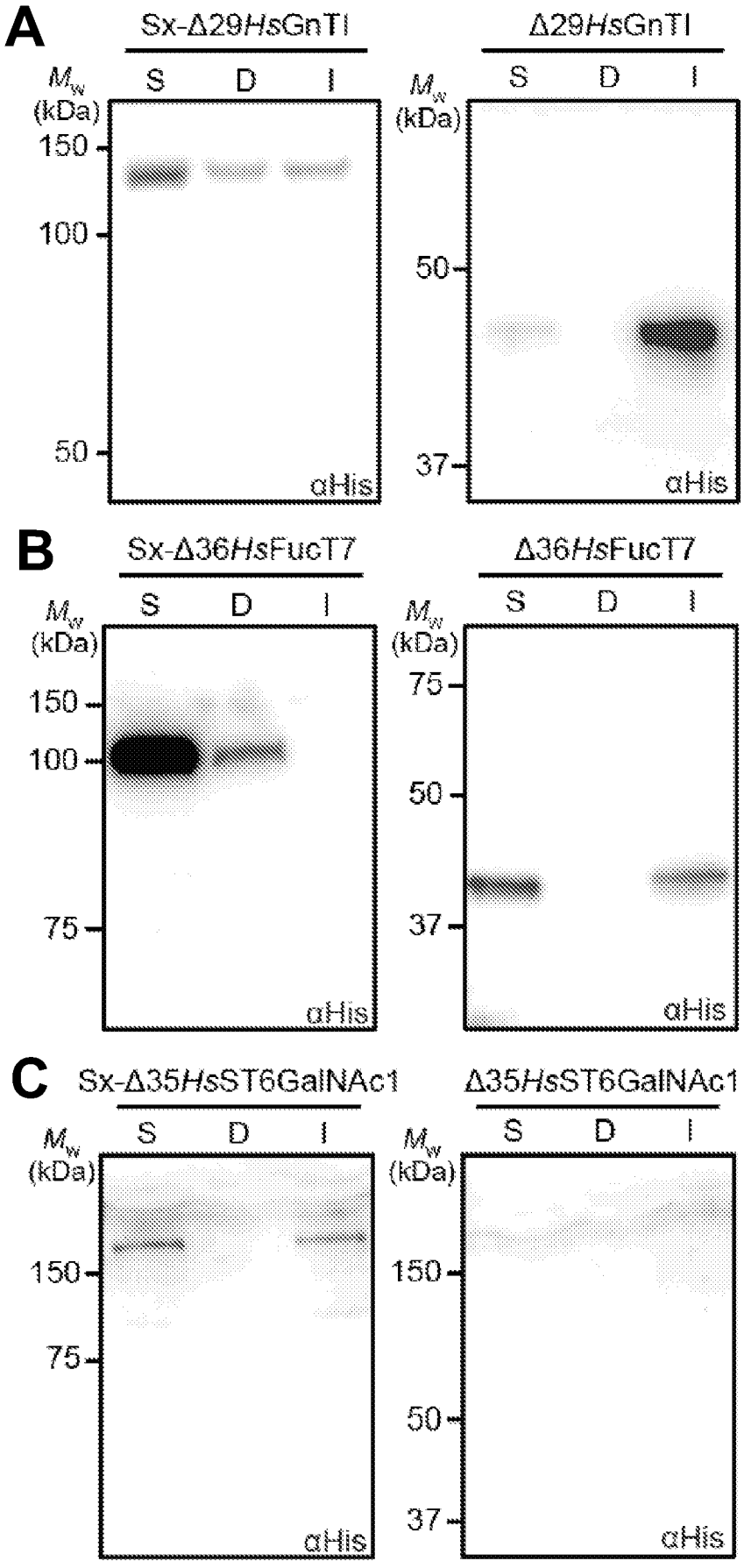
FIGS. 10A–10C


D
human N-acetyl glucosaminyltransferases (HsGlcNAcTs)
MGAT1
MGAT2
MGAT3
MGAT4
GCNT1
GCNT2
β3GNT2
β3GNT6
PigA
Mw (kDa)
75
50
Sx-HsGlcNacT
αGroEL
HsGlcNacT
E
human N-acetyl galactosaminyltransferases (HsGalNAcTs)
GalNAcT1
GalNAcT2
GalNAcT3
GalNAcT4
GalNAcT5
GalNAcT6
GalNAcT7
GalNAcT8
GalNAcT9
GalNAcT10
A-ABO
β3GalNT1
β4GalNT1
Sx-HsGalNacT
HsGalNacT
F
human fucosyltransferases HsFucTs
FuT1
FuT2
FuT3
FuT4
FuT5
FuT6
FuT7
FuT8
FuT9
FuT10
Sx-HsFucT
HsFucT

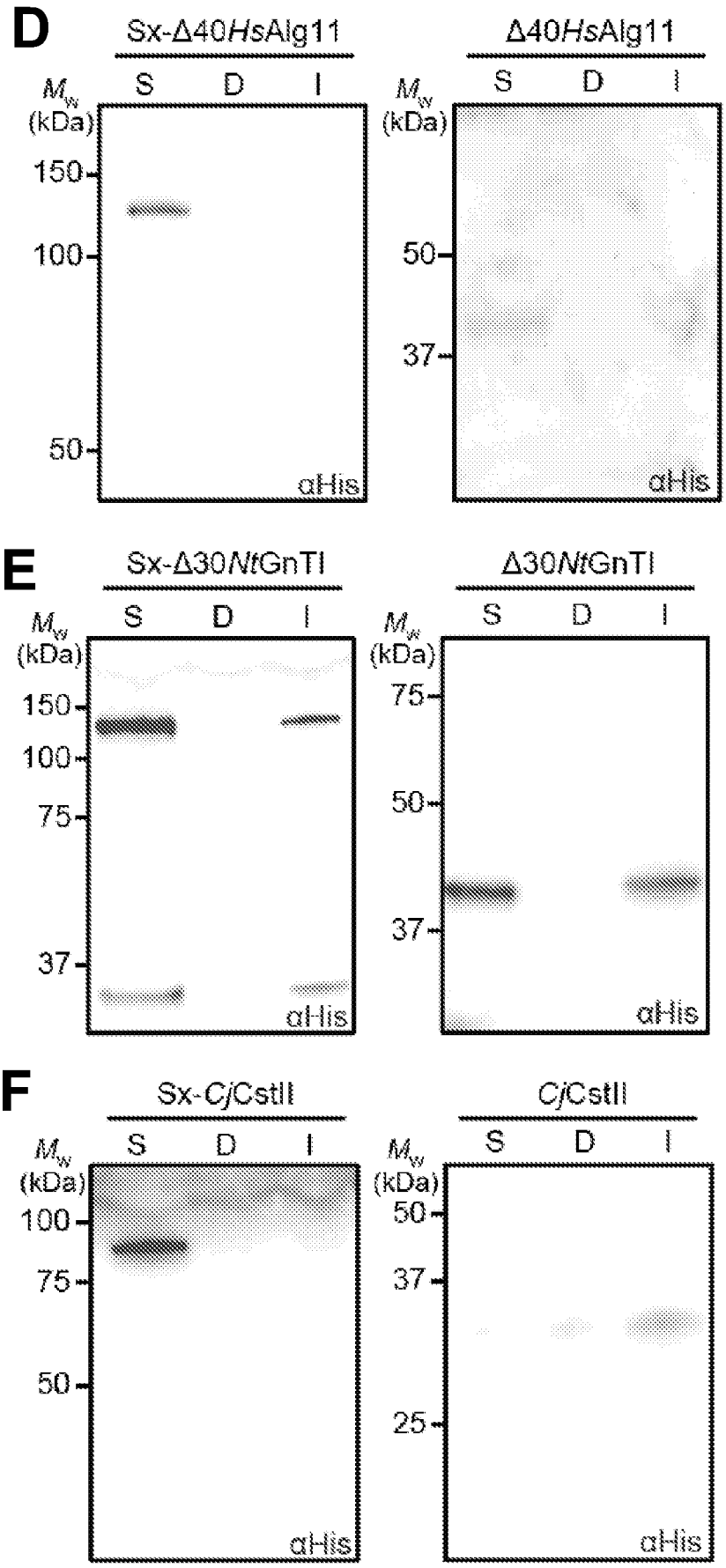
FIGS. 10D–10F


G
human sialyltransferases (HsSiaTs)
Mw (kDa)
ST3Gal1
ST3Gal3
ST3Gal4
ST3Gal6
ST6Gal1
ST6GalNAc1
ST6GalNAc2
ST6GalNAc4
ST8Sia1
ST8Sia2
ST8Sia3
ST8Sia4
75
50
Sx-HsSiaT
αGroEL
HsSiaT
αGroEL
H
other human GTs
UGTA1
UGT1A3
β4GAT1
XXLT1
Sx-HsGT
HsGT
I
eukaryotic GTs
ScAlg1
ScAlg11
NtGnT1
NtGnT2
Btβ4GaT1
BtGGTA1
MmGGTA1
RnGGTA1
Sx-EukGT
EukGT
J
bacterial GTs
CjCstII
CjCgtB
NmPst
NmSynE
HpLgtB
NmLgtB
NgLgtB
EcFuT
EcWecA
LpSetA
Sx-BacGT
BacGT

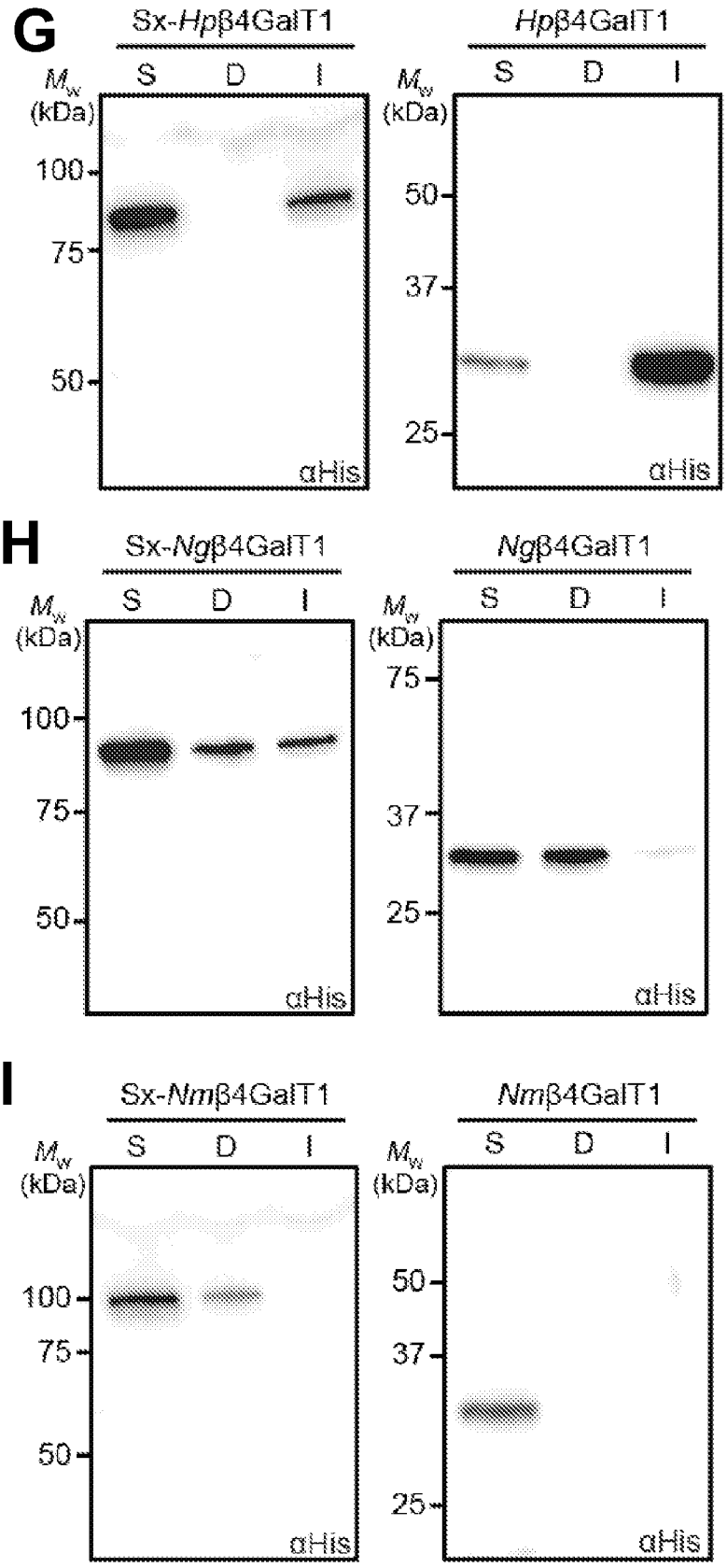
FIGS. 10G–10J


A
Sx-Δ29HsGnTI
Δ29HsGnTI
S D I
Mw (kDa)
150
100
50
37
αHis
B
Sx-Δ36HsFucT7
Δ36HsFucT7
75
C
Sx-Δ35HsST6GalNAc1
Δ35HsST6GalNAc1

FIGS. 11A–11C


D
Sx-Δ40HsAlg11
Δ40HsAlg11
S D I
Mw (kDa)
150
100
50
37
αHis
E
Sx-Δ30NtGnTI
Δ30NtGnTI
S D I
Mw (kDa)
150
100
75
50
37
αHis
F
Sx-CjCstII
CjCstII
S D I
Mw (kDa)
100
75
50
25
αHis

FIGS. 11D–11F


G
Sx-Hpβ4GalT1
Hpβ4GalT1
S D I
Mw (kDa)
100
75
50
37
25
αHis
H
Sx-Ngβ4GalT1
Ngβ4GalT1
I
Sx-Nmβ4GalT1
Nmβ4GalT1

FIGS. 11G–11I

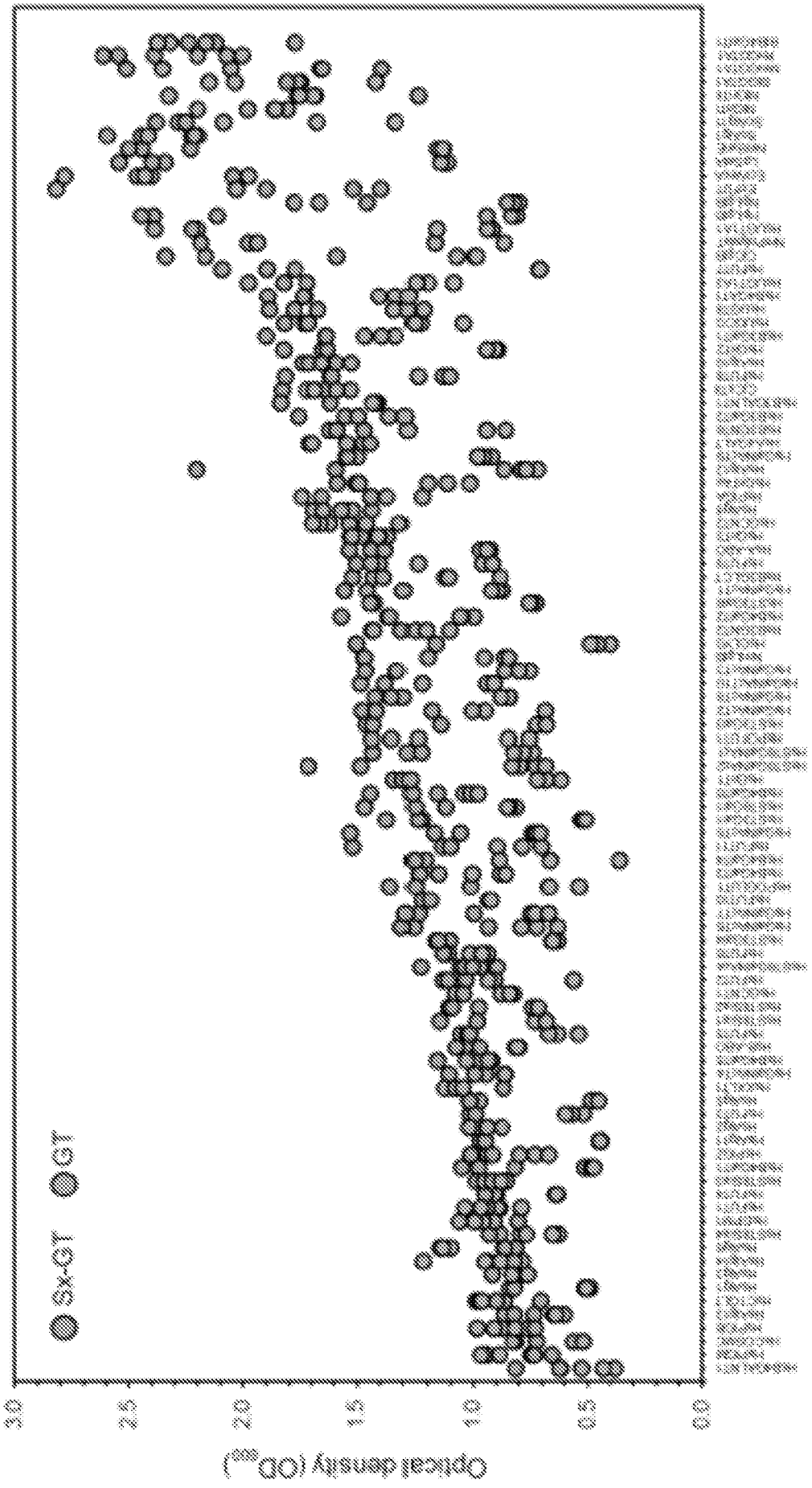
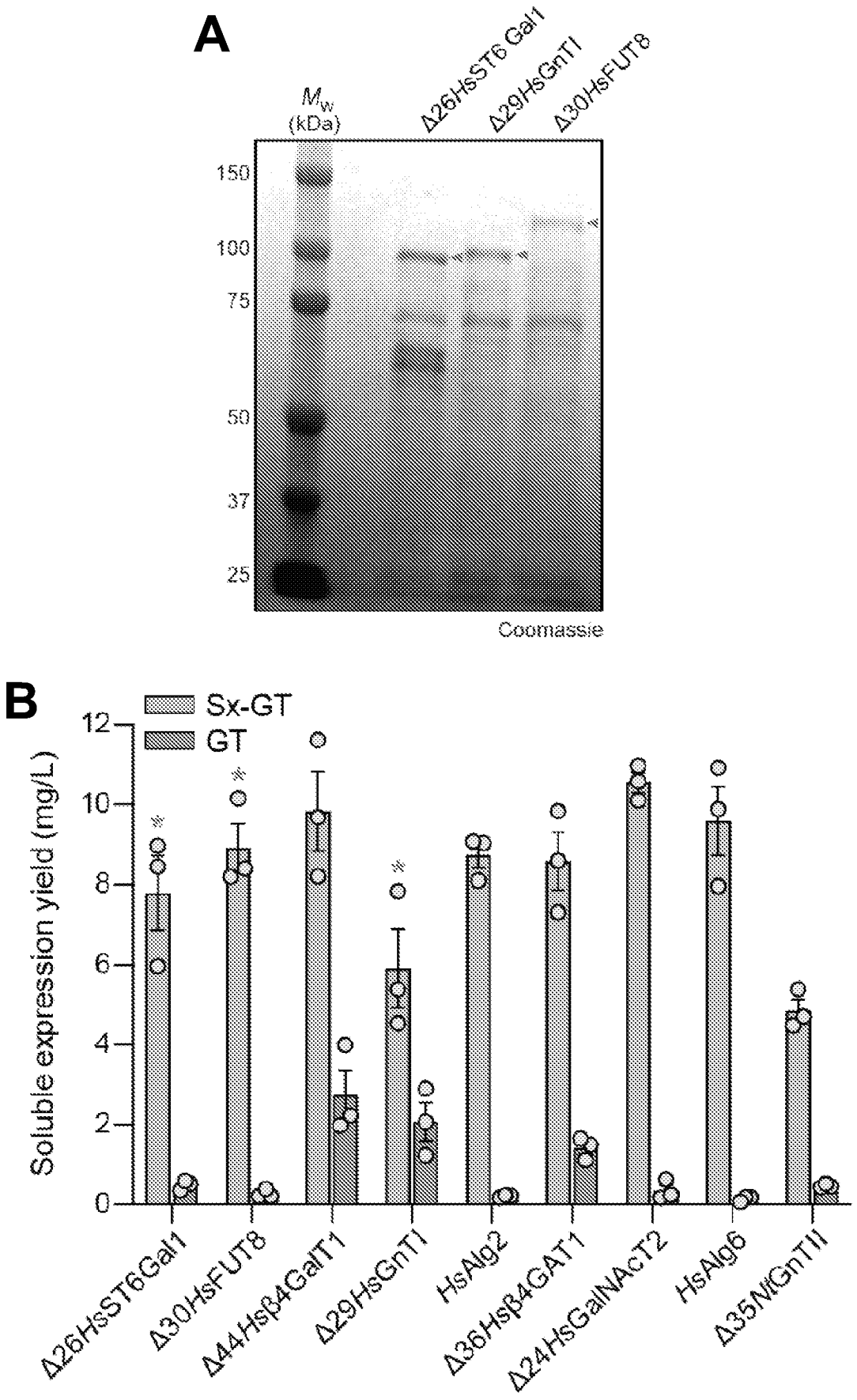


FIG. 12


A
M_w (kDa)
Δ26HsST6 Gal1
Δ29HsGnT1
Δ30HsFUT8
150
100
75
50
37
25
Coomassie
B
Sx-GT
GT
Soluble expression yield (mg/L)
12
10
8
6
4
2
0
Δ26HsST6Gal1
Δ30HsFUT8
Δ44Hsβ4GalT1
Δ29HsGnT1
HsAlg2
Δ36Hsβ4GAT1
Δ24HsGalNAcT2
HsAlg6
Δ35NtGnT1I

FIGS. 13A–13B

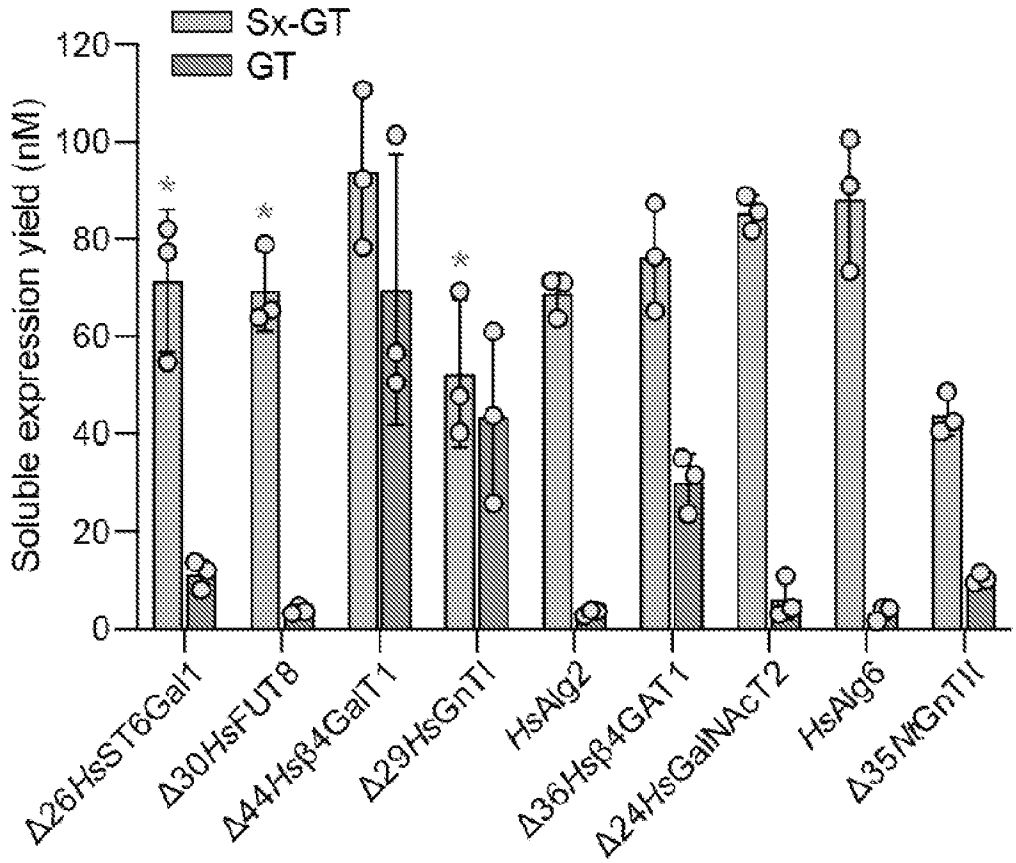

Sx-GT
GT
Soluble expression yield (nM)
120
100
80
60
40
20
0
*
*
*
Δ26HsST6Gal1
Δ30HsFUT8
Δ44Hsβ4GalT1
Δ29HsGnTI
HsAlg2
Δ36Hsβ4GAT1
Δ24HsGalNAcT2
HsAlg6
Δ35MtGnTII

FIG. 13C

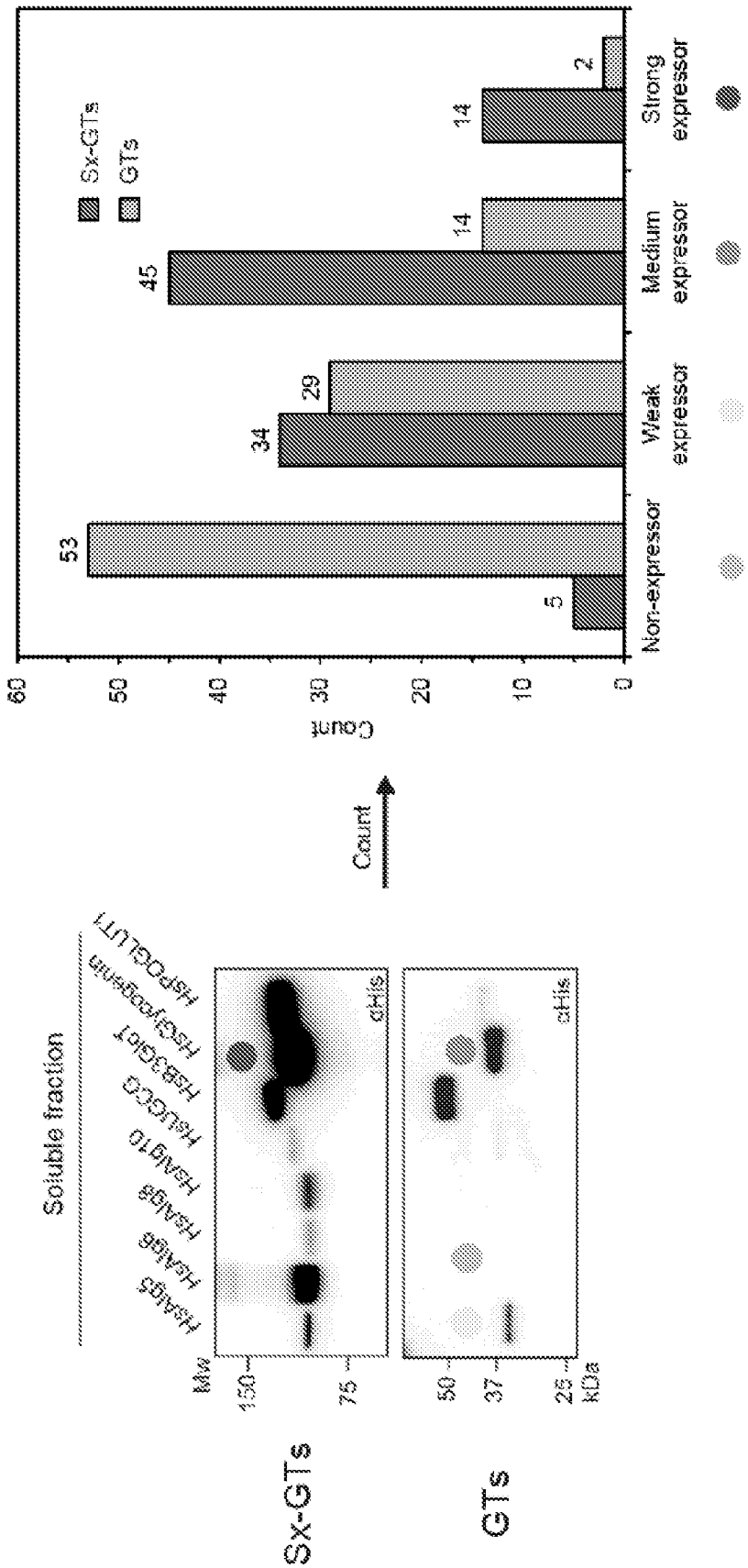

Sx-GTs
GTs
Count
60
50
40
30
20
10
0
53
5
34
29
45
14
14
2
Non-expressor
Weak expressor
Medium expressor
Strong expressor
Count
Soluble fraction
HsAlg5
HsAlg6
HsAlg8
HsAlg9
HsAlg10
HeLiGCG
HSB3Gal-T
HSGlycogenin
HSPOGLUT1
MW
150
75
50
37
25
kDa
αHis
αHis
Sx-GTs
GTs

FIG. 14A

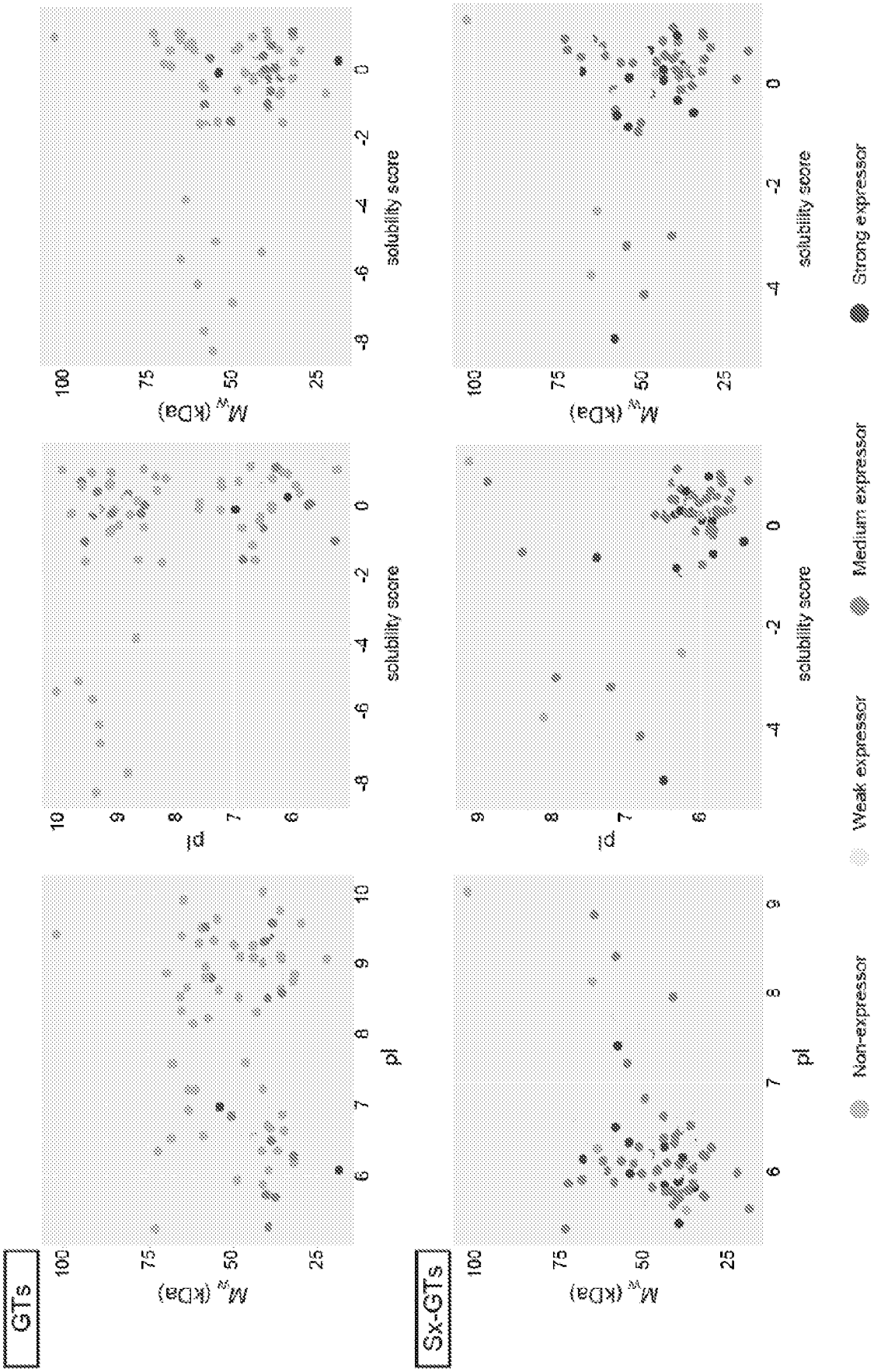

GTs
Sx-GTs
M_w (kDa)
pI
solubility score
Non-expressor
Weak expressor
Medium expressor
Strong expressor

FIG. 14B

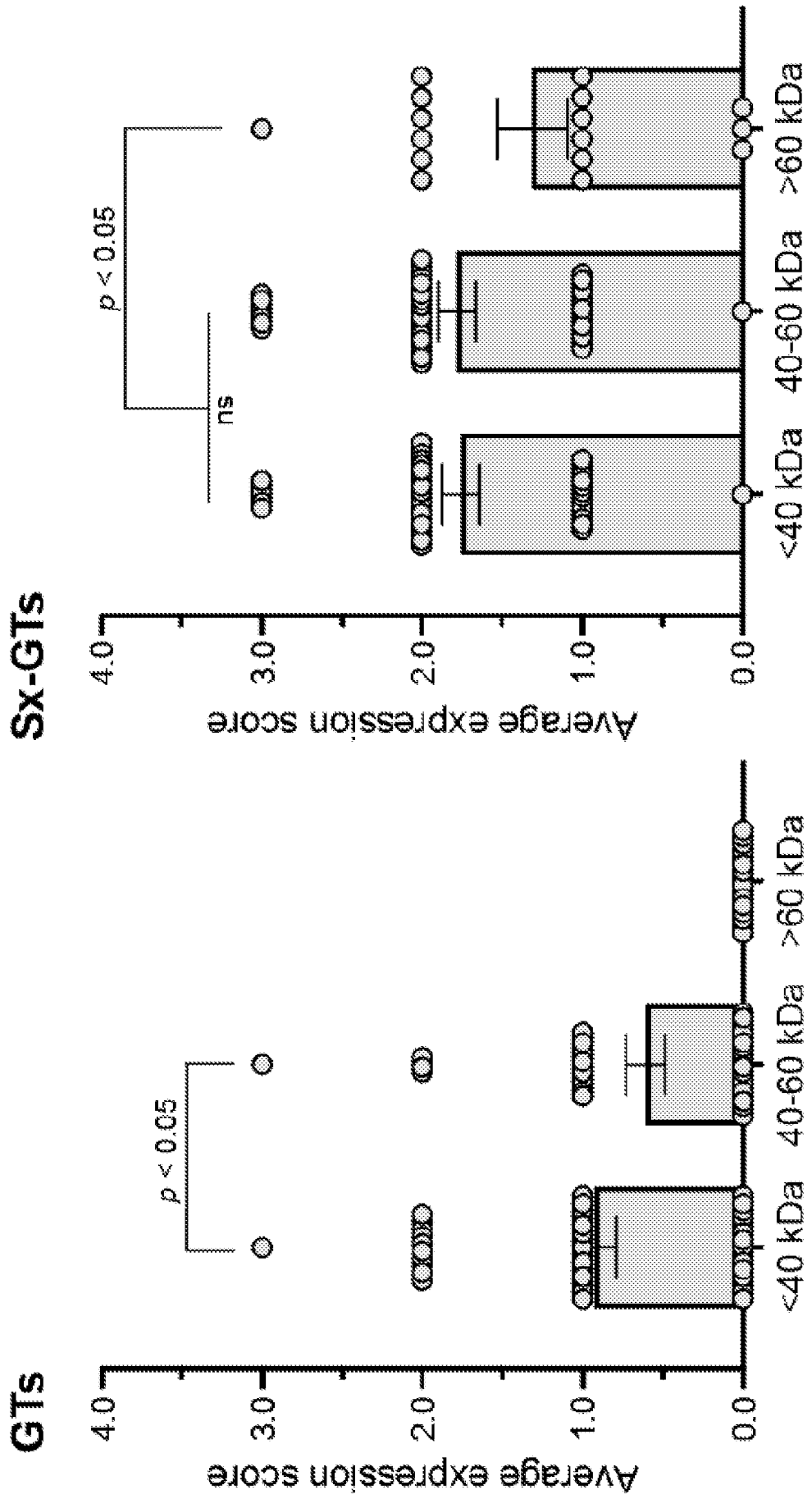

Sx-GTs
GTs
Average expression score
4.0
3.0
2.0
1.0
0.0
p < 0.05
ns
<40 kDa
40-60 kDa
>60 kDa

FIG. 15A

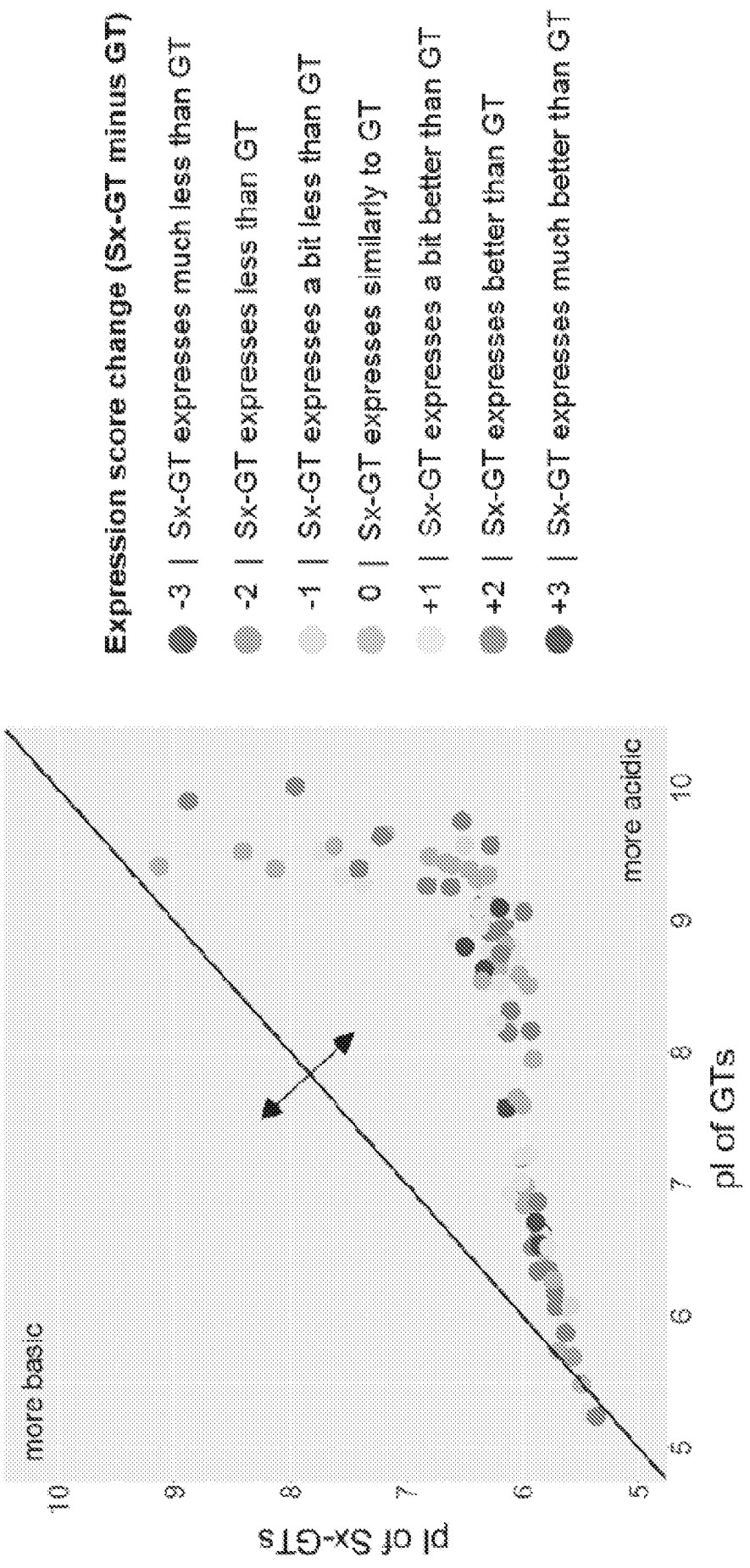
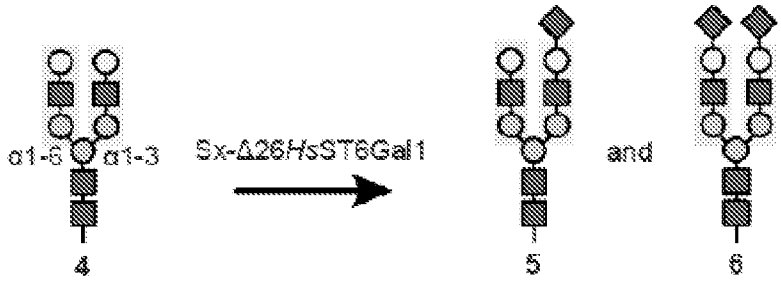

Expression score change (Sx-GT minus GT)
-3 | Sx-GT expresses much less than GT
-2 | Sx-GT expresses less than GT
-1 | Sx-GT expresses a bit less than GT
0 | Sx-GT expresses similarly to GT
+1 | Sx-GT expresses a bit better than GT
+2 | Sx-GT expresses better than GT
+3 | Sx-GT expresses much better than GT
more basic
more acidic
pI of Sx-GTs
pI of GTs
10
9
8
7
6
5

FIG. 15B



ID	Time (min)	Area	Height	Width	Width at 50%
4	2.625	9914.3	664.0	0.728	0.132
5	2.427	336409.0	58667.5	0.976	0.236
6	2.231	92802.9	15767.2	0.329	0.138

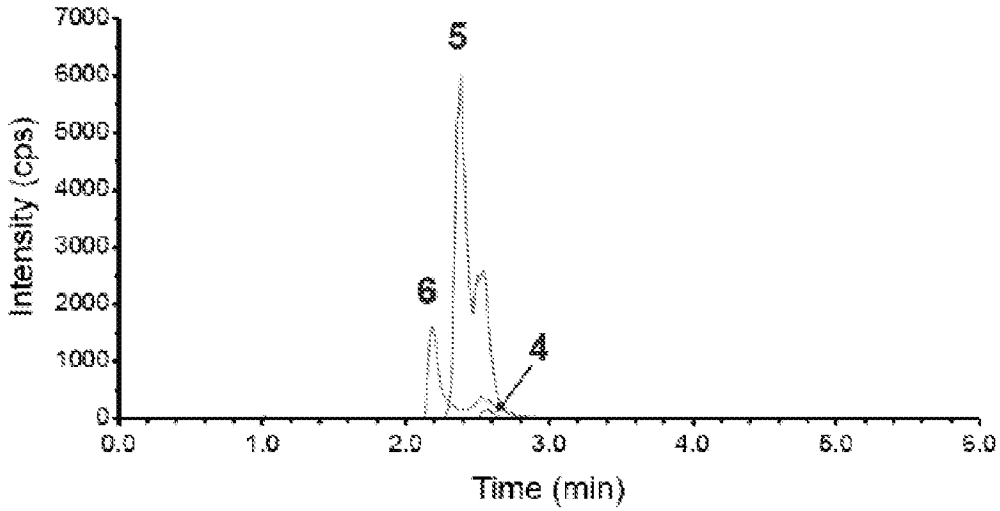


FIG. 16A

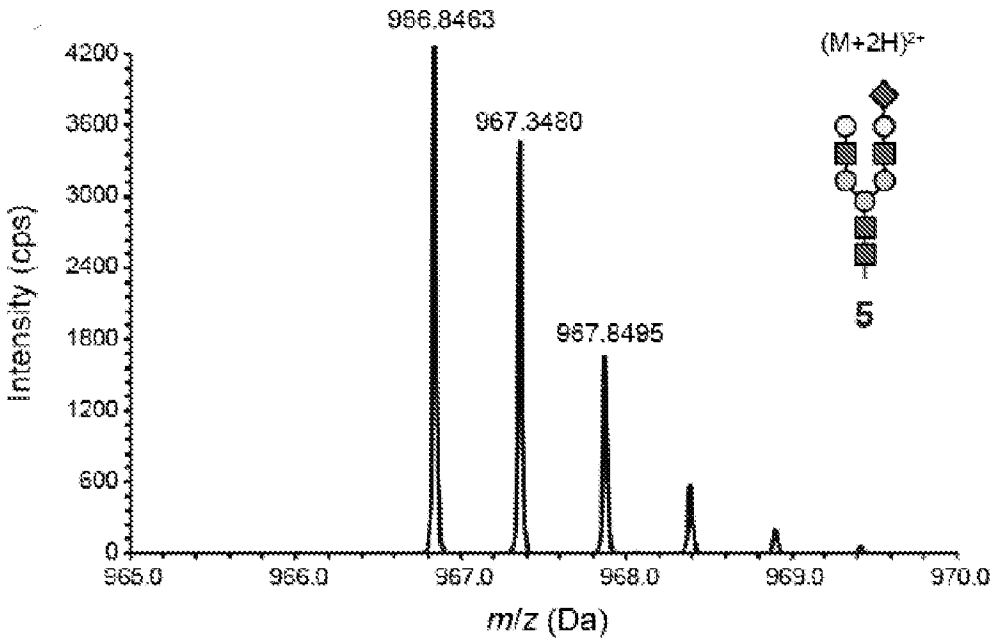


FIG. 16B

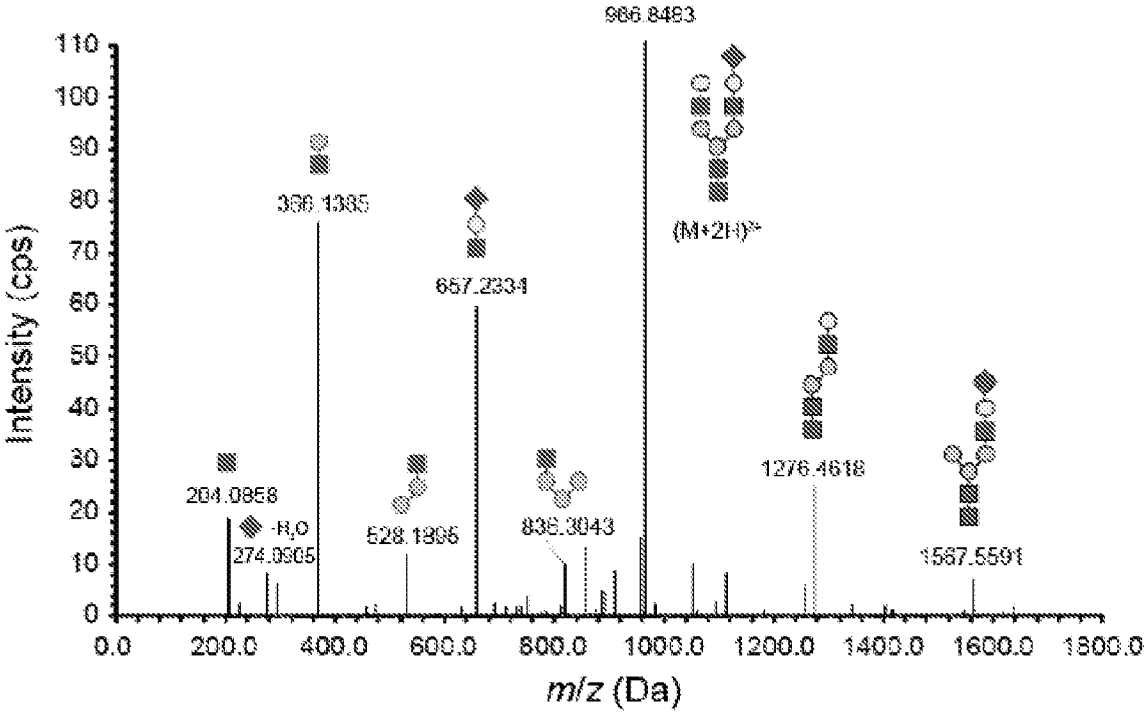


FIG. 16B (continued)

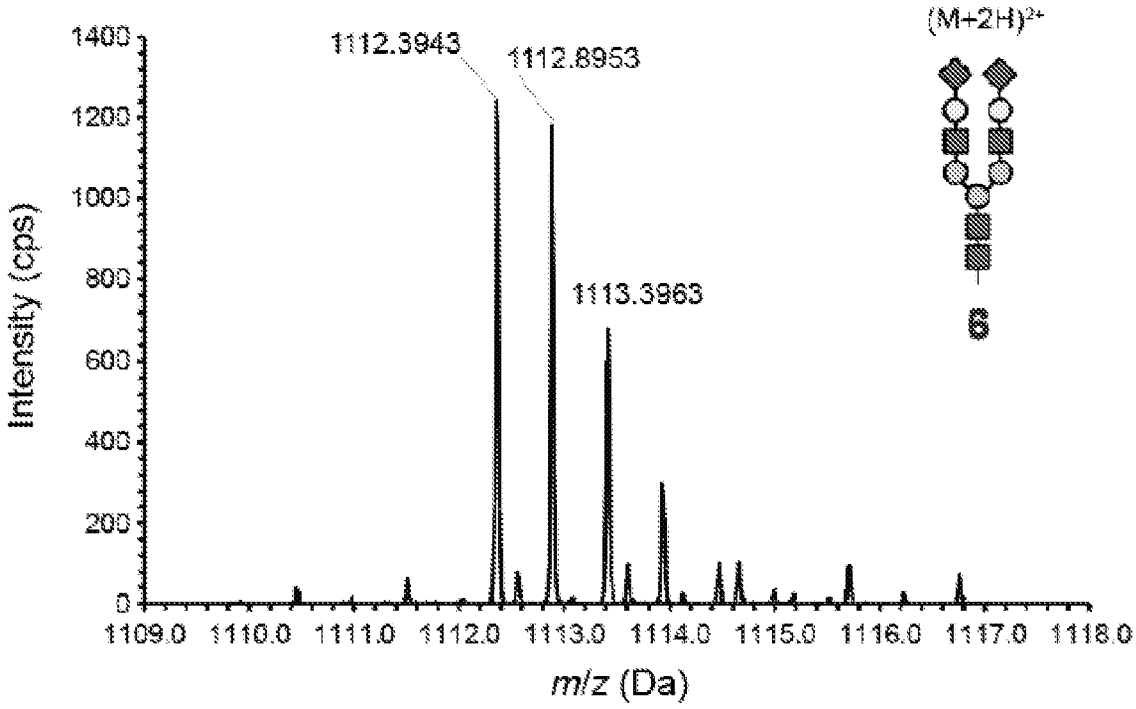


FIG. 16C

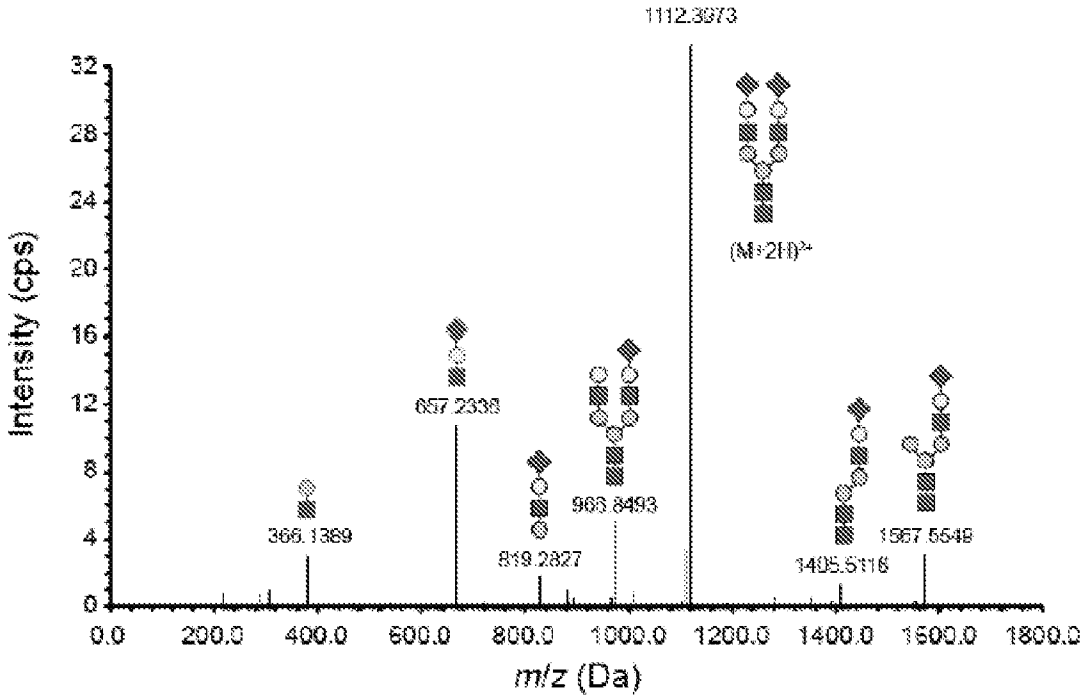
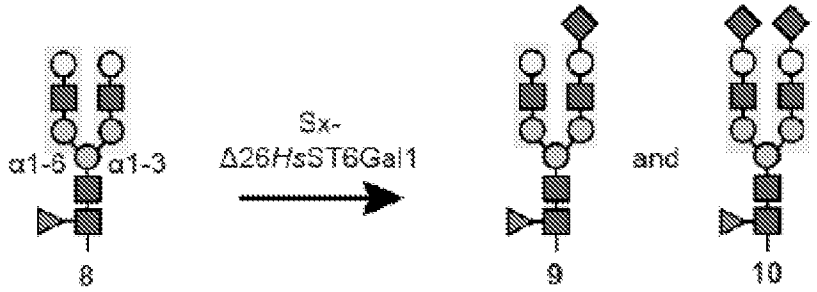


FIG. 16C (continued)



ID	Time (min)	Area	Height	Width	Width at 50%
9	1.278	132185.4	30847.44	0.238	0.063
10	1.205	16656.4	2380.64	0.228	0.113

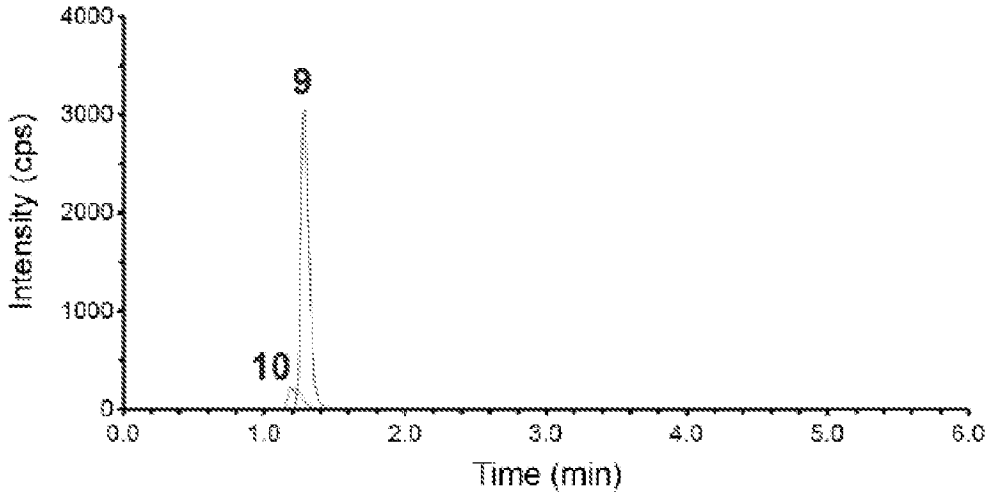


FIG. 17A

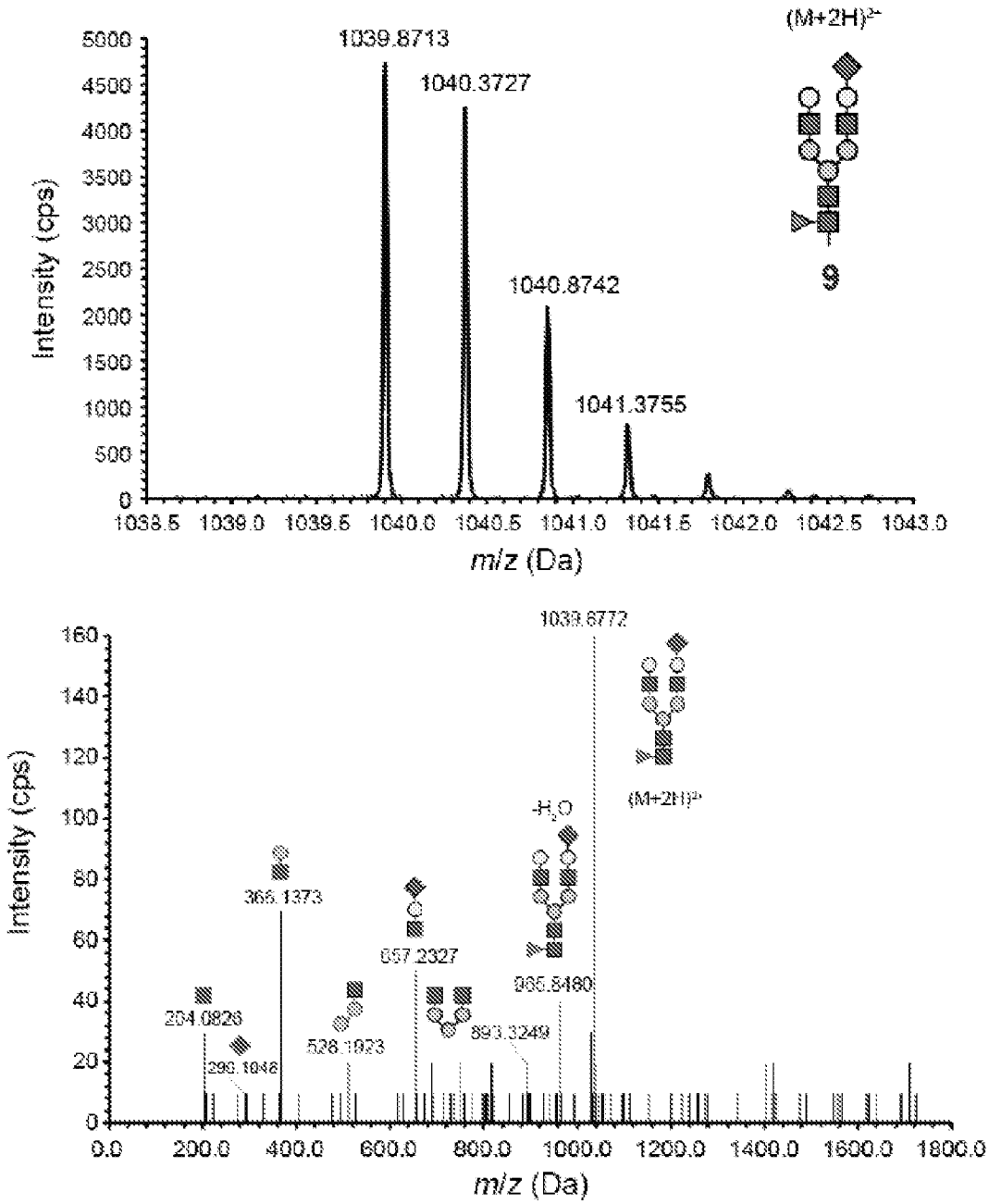

(M+2H)2+
1039.8713
1040.3727
1040.8742
1041.3755
9
Intensity (cps)
m/z (Da)
1039.8772
(M+2H)2+
-H2O
366.1373
657.2327
865.8480
204.0826
296.1048
528.1923
893.3249
Intensity (cps)
m/z (Da)

FIG. 17B

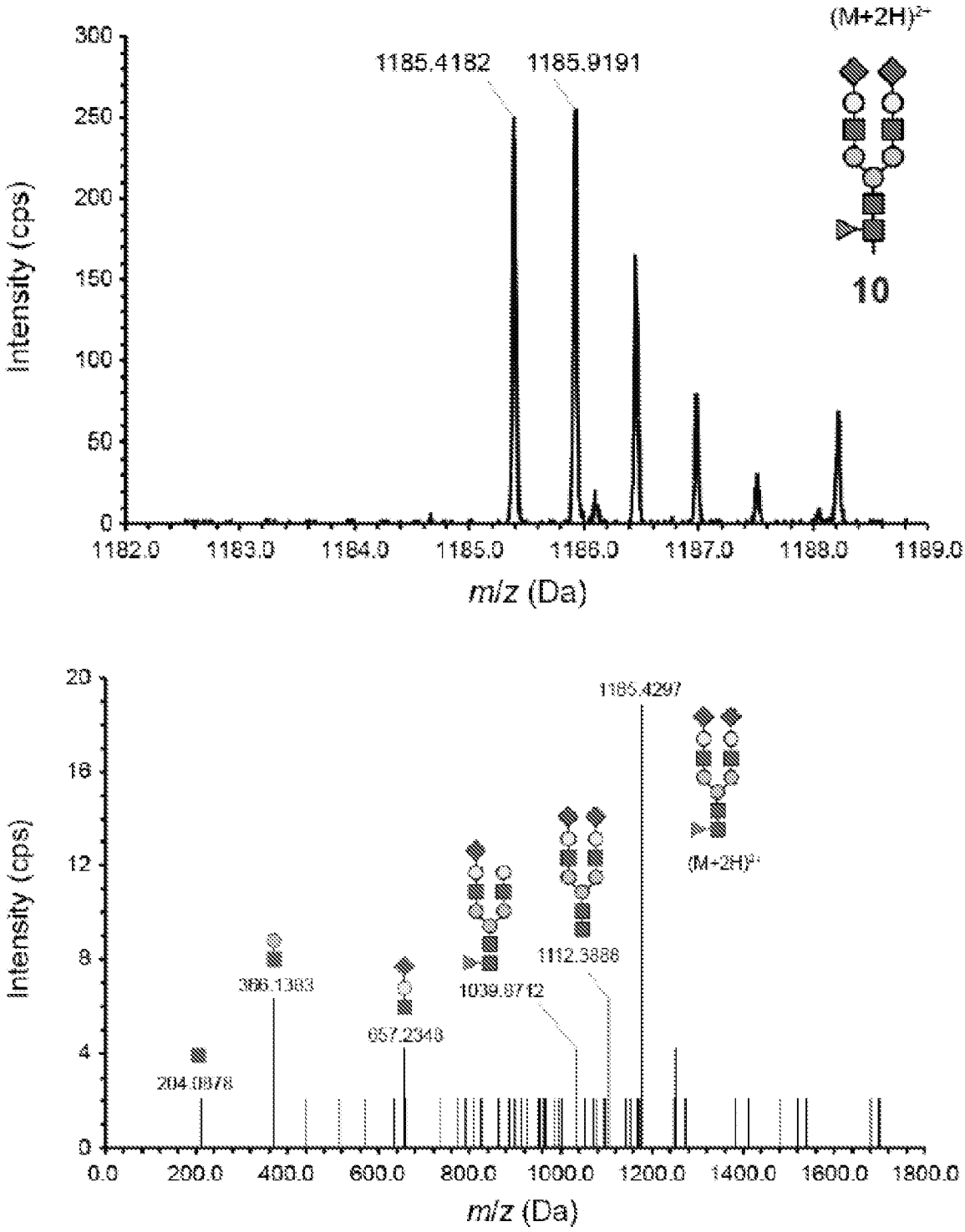
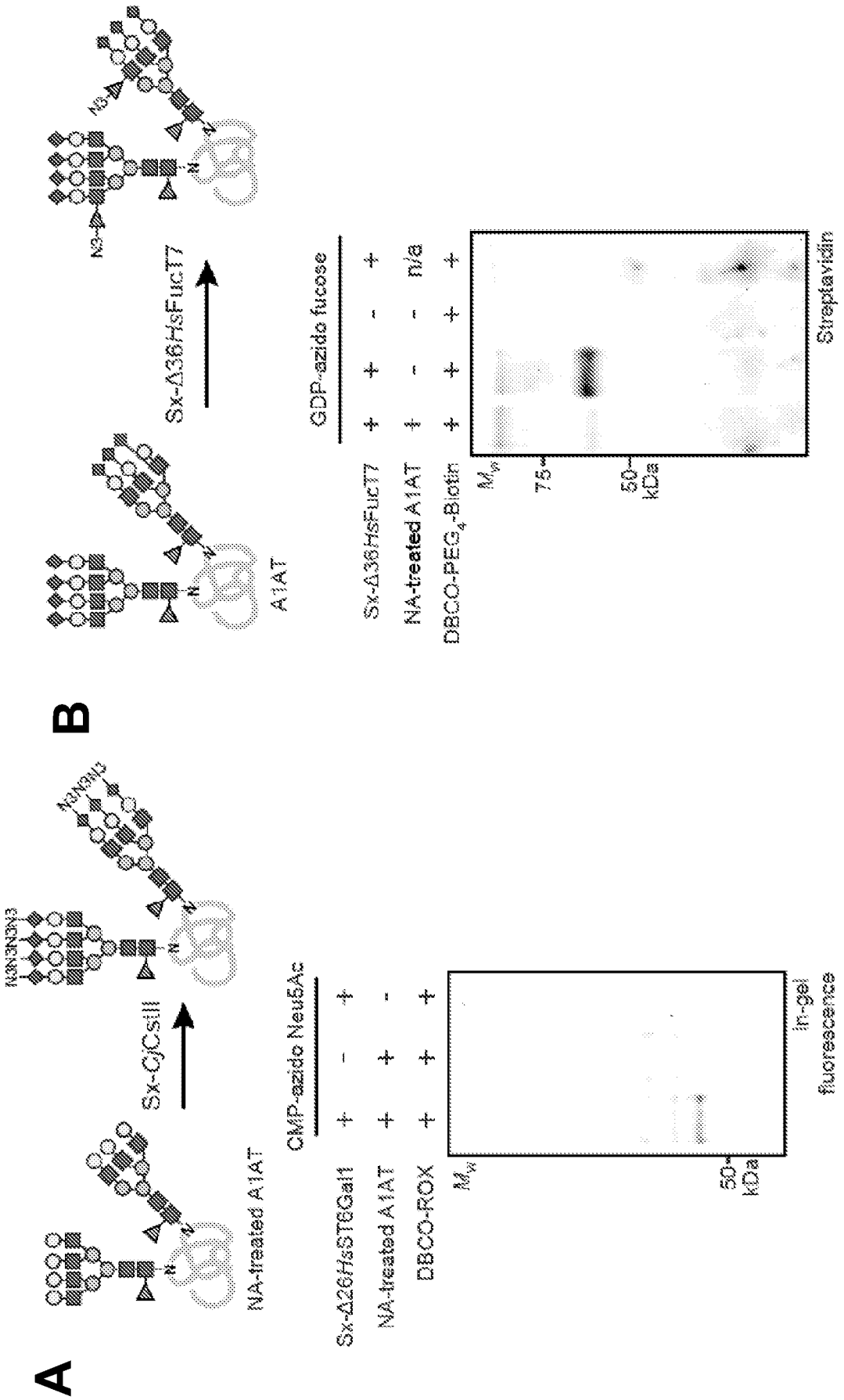

(M+2H)2+
1185.4182
1185.9191
10
Intensity (cps)
m/z (Da)
1185.4297
(M+2H)2+
1112.3886
1039.8712
396.1383
657.2348
204.0878
Intensity (cps)
m/z (Da)

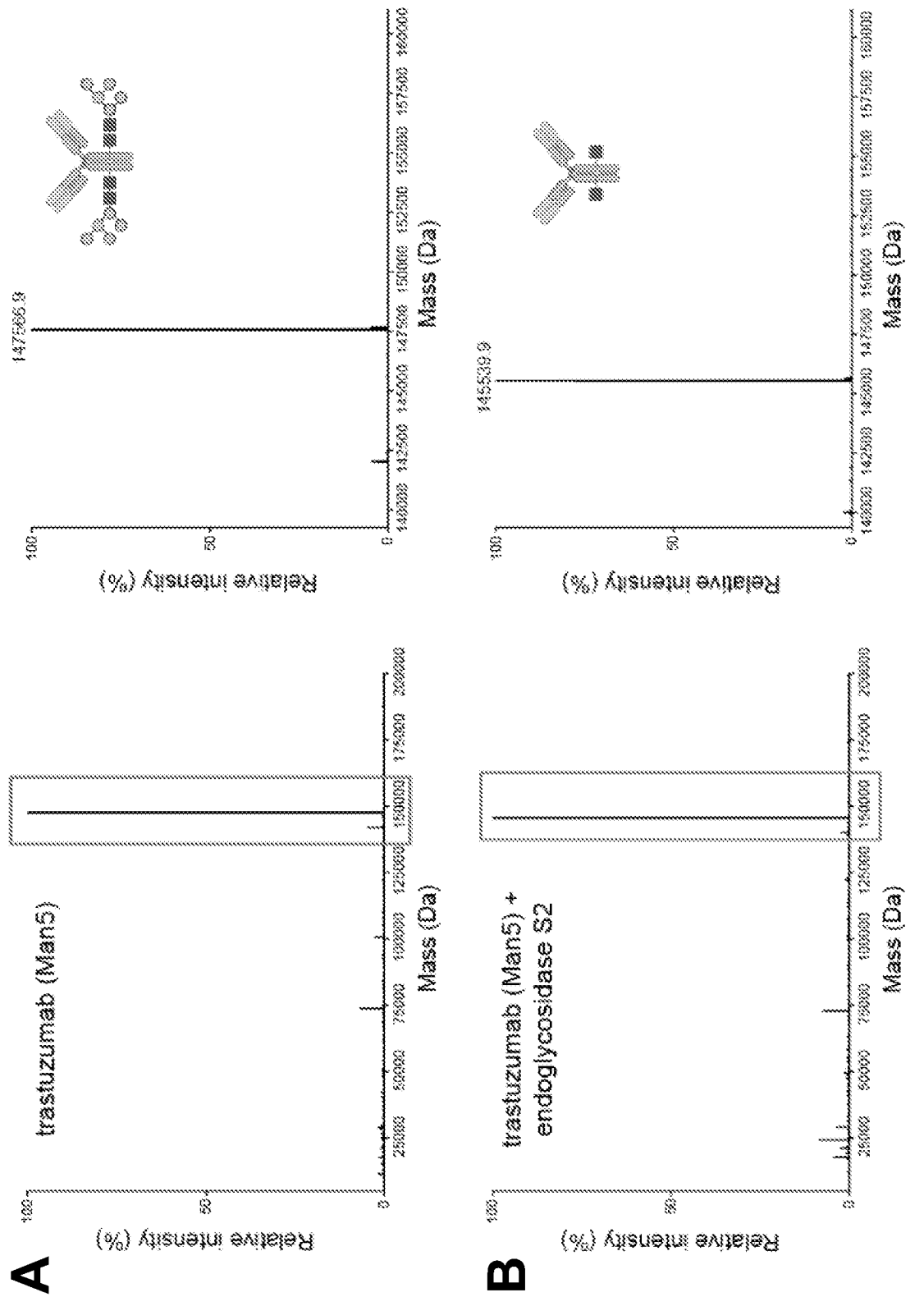
FIG. 17C


A
NA-treated A1AT
Sx-CjCstII
N3N3N3N3
N3N3N3
CMP-azido Neu5Ac
Sx-Δ26HsST6Gal1
NA-treated A1AT
DBCO-ROX
M_W
50–
kDa
in-gel
fluorescence
B
A1AT
Sx-Δ36HsFucT7
N3
GDP-azido fucose
Sx-Δ36HsFucT7
NA-treated A1AT
DBCO-PEG$_4$-Biotin
M_W
75–
50–
kDa
Streptavidin

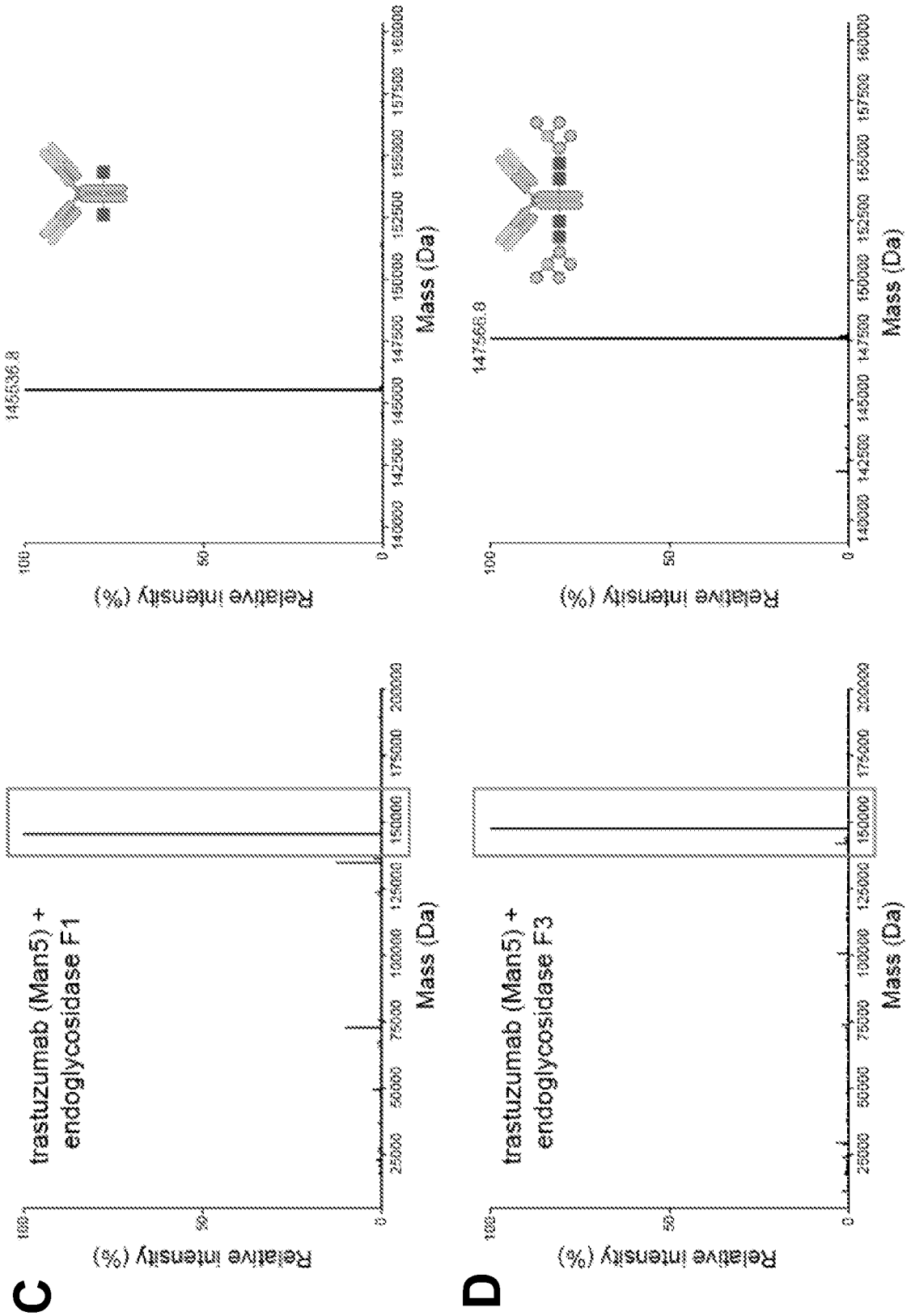
FIGS. 18A–18B

C
Sx-Δ34HsST3Gal1
bovine mucin
CMP-azido Neu5Ac
Sx-Δ34HsST3Gal1
NA-treated mucin
DBCO-ROX
Mw
150
100
75
kDa
In-gel fluorescence
D
Sx-Δ29HsGnT1
MBP-GCG
UDP-azido GlcNAc
Sx-Δ29HsGnT1
Man3GlcNAc2 MBP-GCG
DBCO-PEG4-Biotin
Mw
100
75
50
kDa
Streptavidin
αHis

FIGS. 18C–18D


A
trastuzumab (Man5)
Relative intensity (%)
Mass (Da)
1475866.9
B
trastuzumab (Man5) +
endoglycosidase S2
Relative intensity (%)
Mass (Da)
1455539.9

FIGS. 19A–19B


C
trastuzumab (Man5) + endoglycosidase F1
Relative Intensity (%)
Mass (Da)
D
trastuzumab (Man5) + endoglycosidase F3
Relative Intensity (%)
Mass (Da)

FIGS. 19C–19D

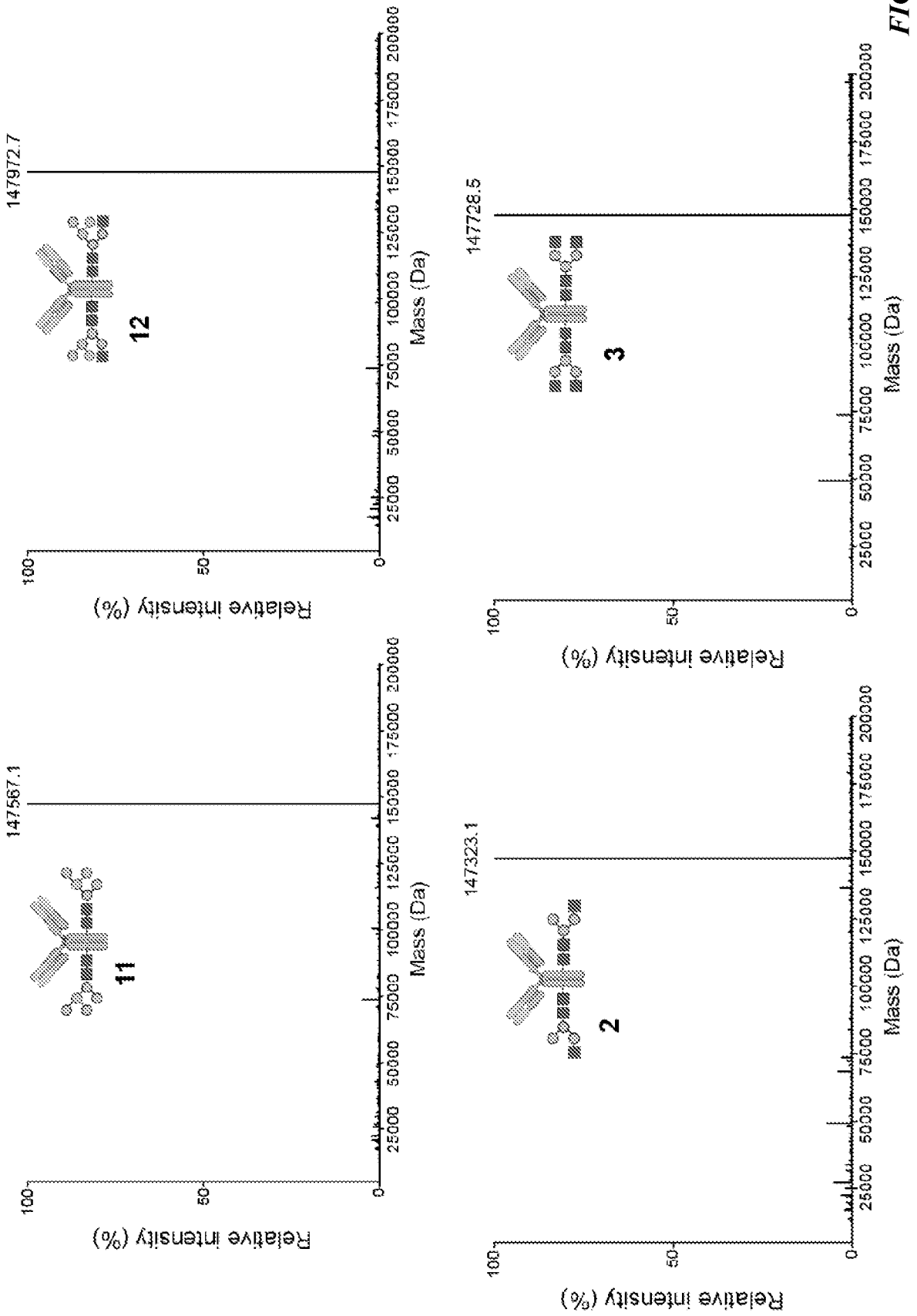

147567.1
11
Relative Intensity (%)
Mass (Da)
147972.7
12
Relative Intensity (%)
Mass (Da)
147323.1
2
Relative Intensity (%)
Mass (Da)
147728.5
3
Relative Intensity (%)
Mass (Da)

FIG. 20

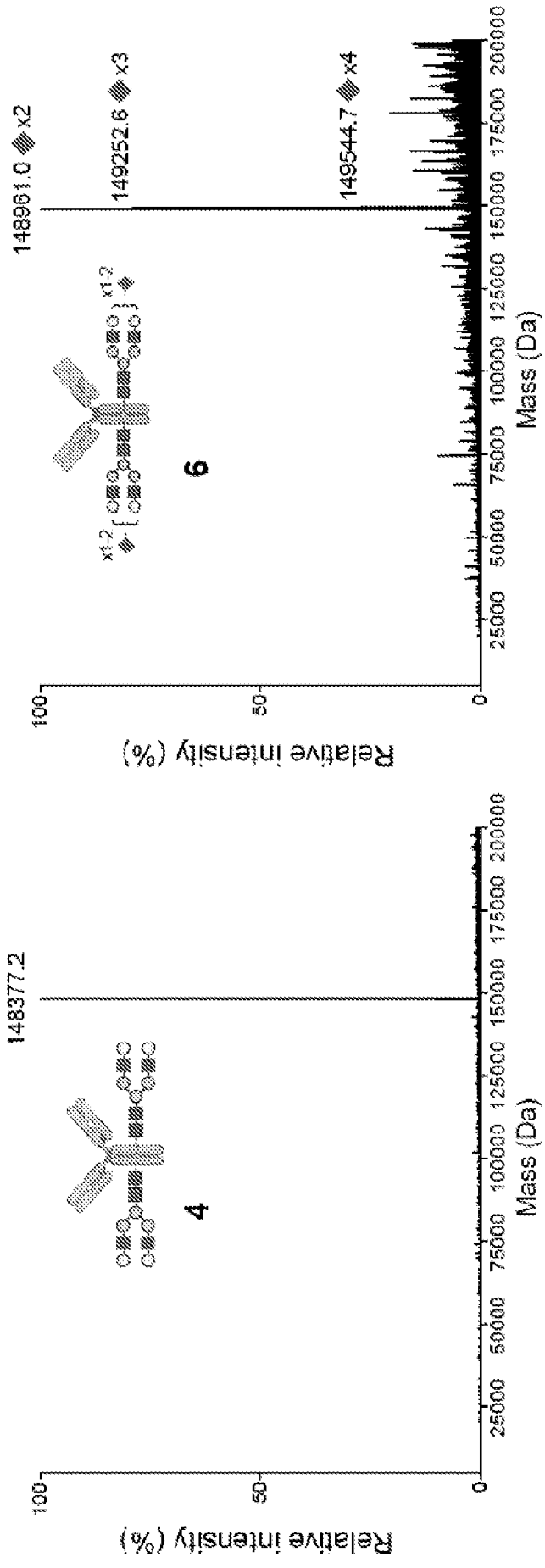


FIG. 20 (continued)

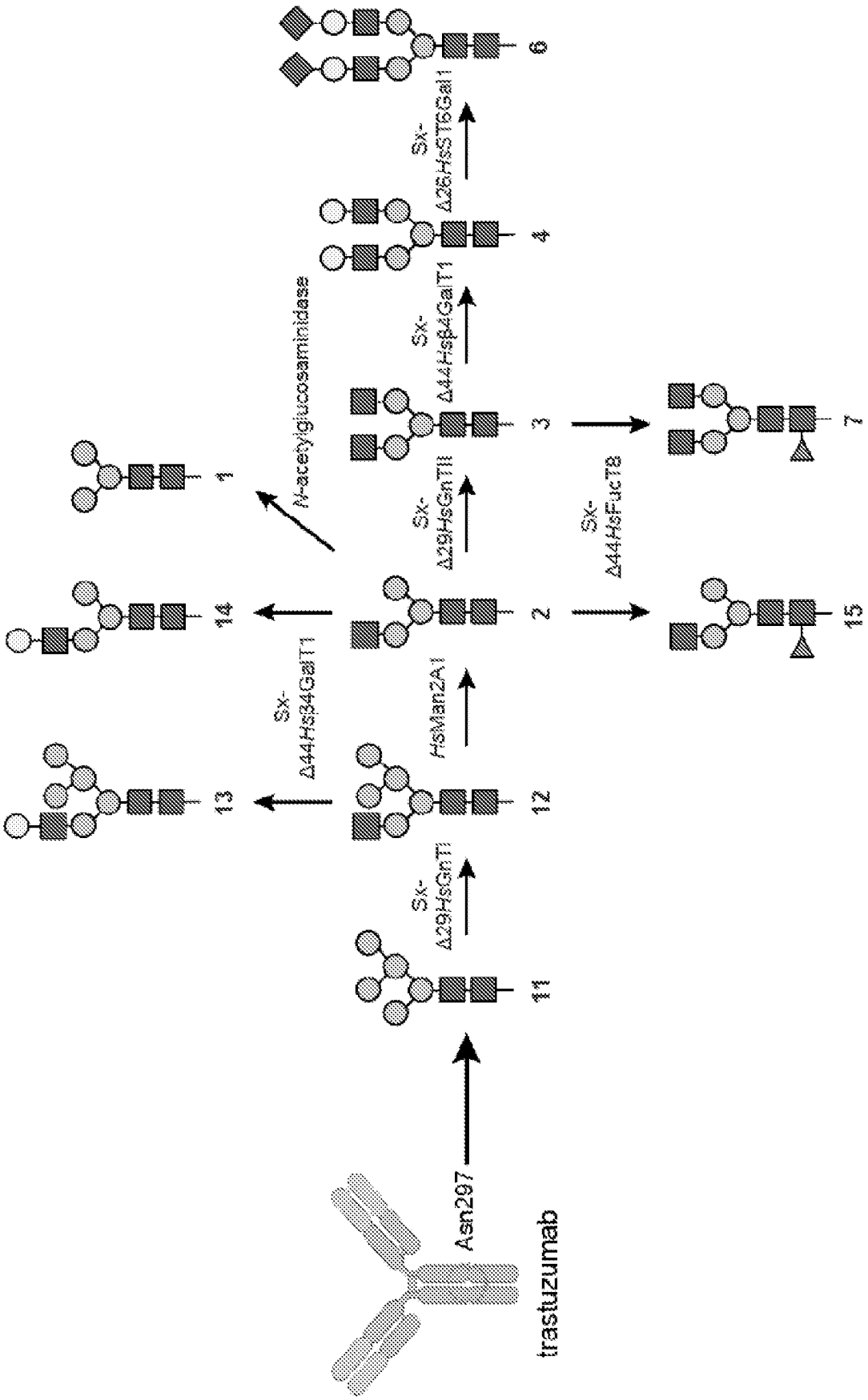

trastuzumab
Asn297
11
Sx-
Δ29HsGnTII
12
HsMan2A1
13
Sx-
Δ44HsB4GalT1
14
1
N-acetylglucosaminidase
2
Sx-
Δ29HsGnTII
3
Sx-
Δ44HsB4GalT1
4
Sx-
Δ26HsST6Gal1
6
Sx-
Δ44HsFuc78
15
7

FIG. 21A

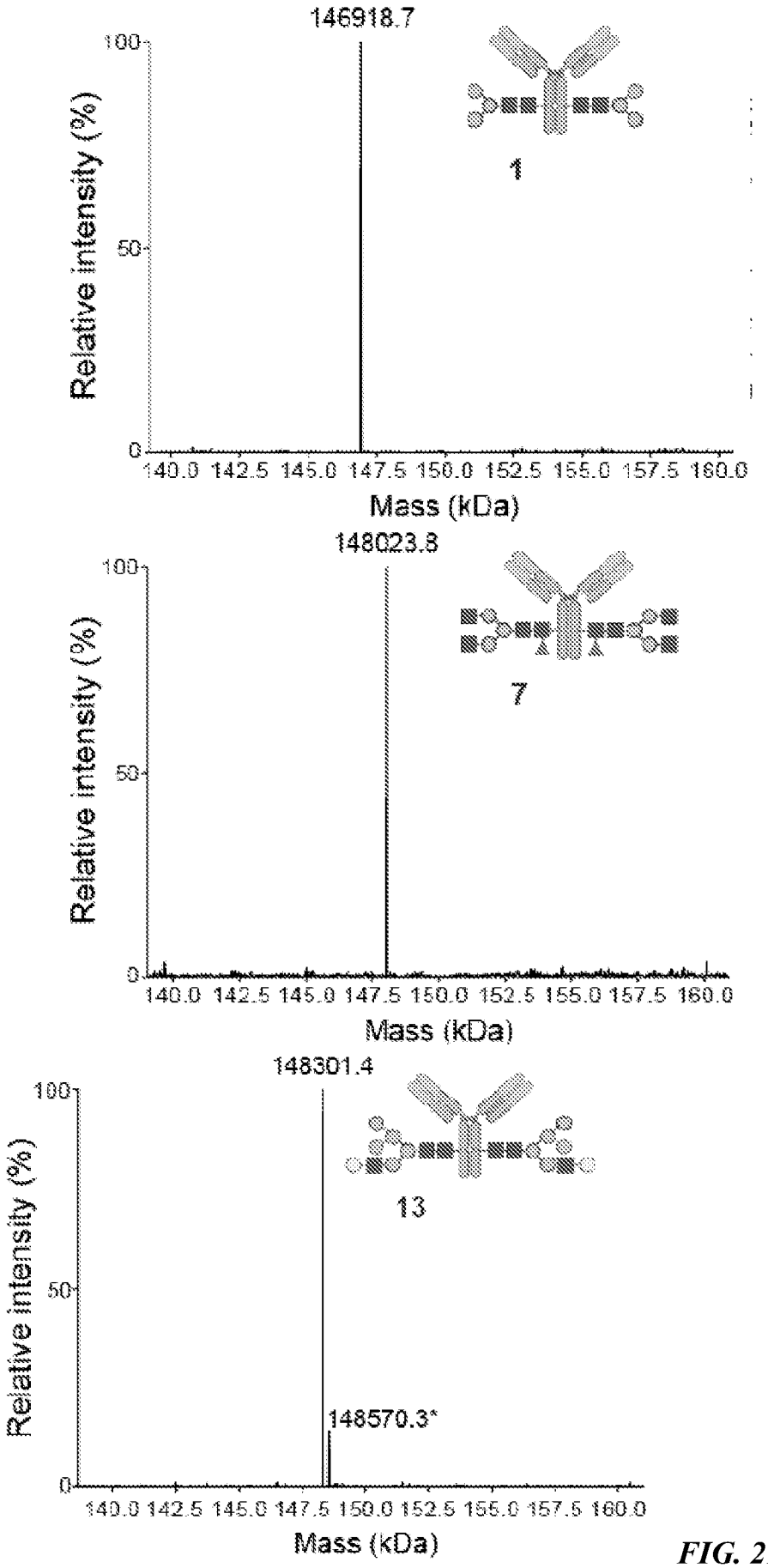

146918.7
100
50
0
Relative intensity (%)
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
1
148023.8
100
50
0
Relative intensity (%)
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
7
148301.4
100
50
0
Relative intensity (%)
148570.3*
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
13

FIG. 21B

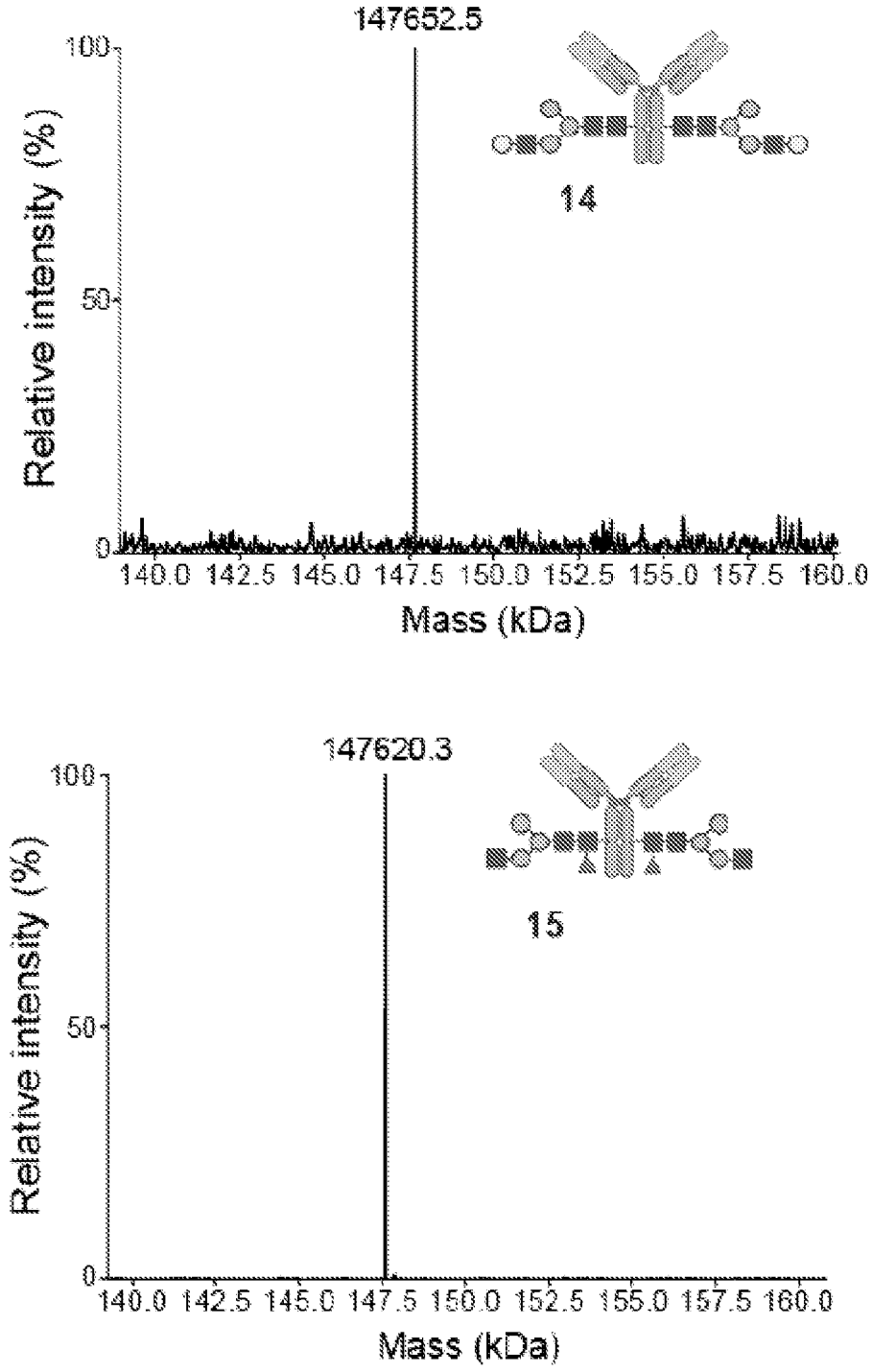


FIG. 21B (continued)

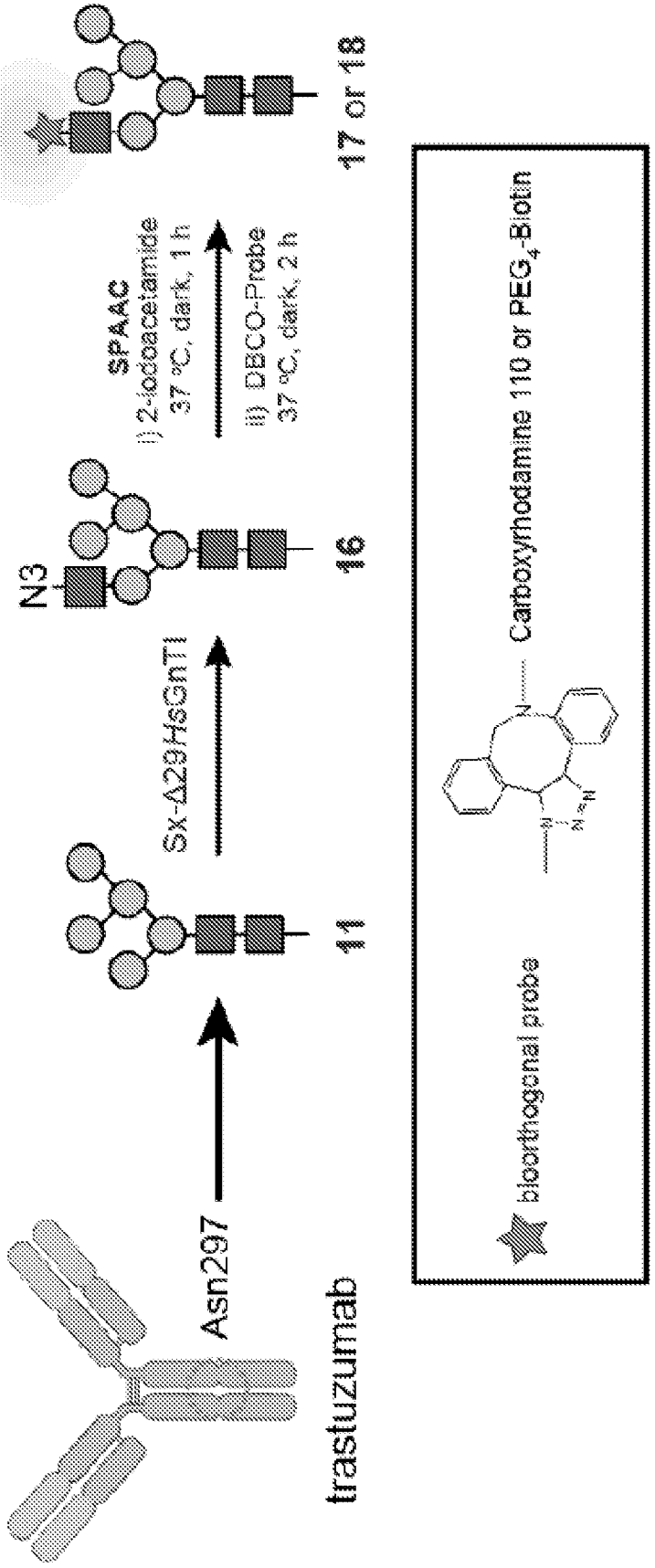

trastuzumab
Asn297
11
Sx-Δ29HsGnT1
N3
16
SPAAC
i) 2-iodoacetamide
37 °C, dark, 1 h
ii) DBCO-Probe
37 °C, dark, 2 h
17 or 18
bioorthogonal probe
Carboxyrhodamine 110 or PEG4-Biotin

FIG. 22A

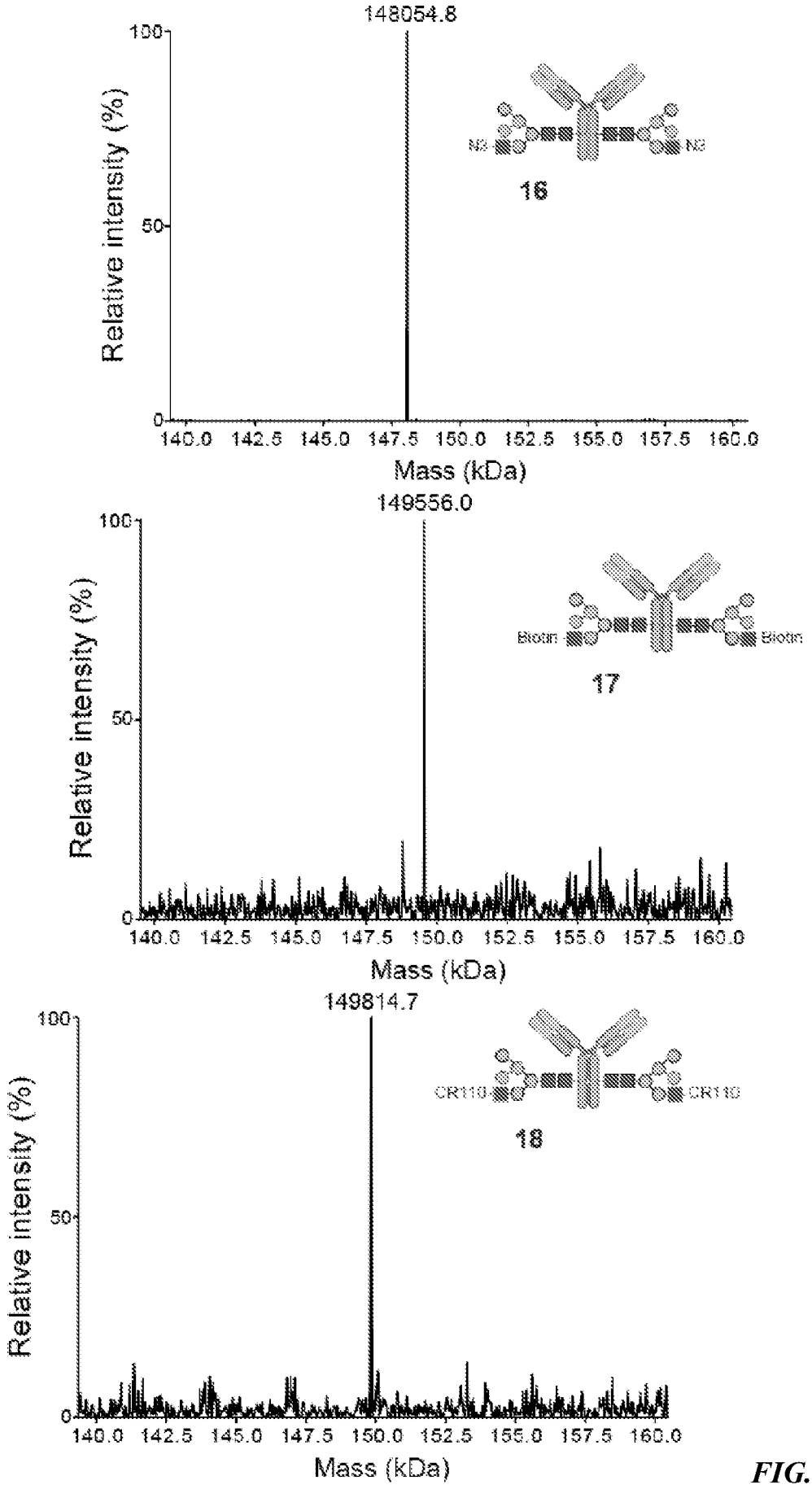

148054.8
100
50
0
Relative intensity (%)
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
N3
N3
16
149556.0
100
50
0
Relative intensity (%)
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
Biotin
Biotin
17
149814.7
100
50
0
Relative intensity (%)
140.0 142.5 145.0 147.5 150.0 152.5 155.0 157.5 160.0
Mass (kDa)
CR110
CR110
18

FIG. 22B

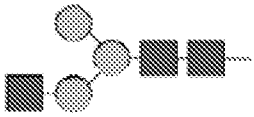
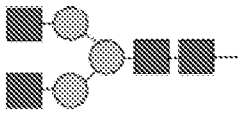
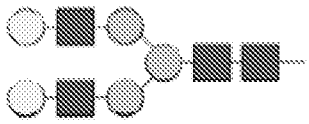
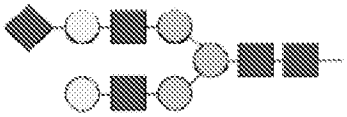
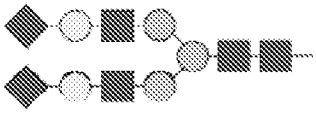
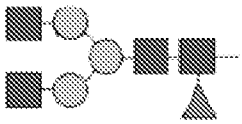
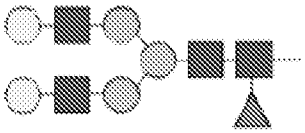
Glycan Number	Symbolic Representation	Oxford Notation	Structure
1		M3	Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
2		A1	Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAc β1-4GlcNAc
3		A2	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man β1-4GlcNAcβ1-4GlcNAc
4		A2G2	GalÎ²1-4GlcNAcÎ²1-2ManÎ±1-6(GalÎ²1-4GlcNAcÎ²1-2 ManÎ±1-3)ManÎ²1-4GlcNAcÎ²1-4GlcNAc
5		A2G2S1	Galβ1-4GlcNAcβ1-2Manα1-6(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4 GlcNAc
6		A2G2S2	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6 (Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3) Manβ1-4GlcNAcβ1-4GlcNAc
7		FA2	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3) Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc
8		FA2G2	Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-2 Manα1-3)Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc

FIG. 23

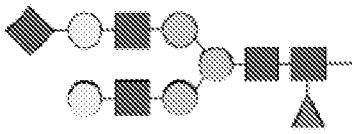
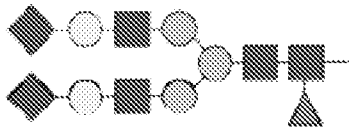
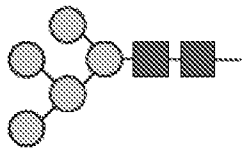
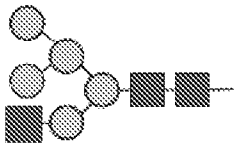
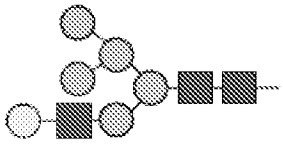
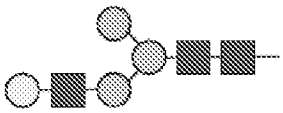
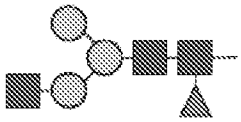
Glycan Number	Symbolic Representation	Oxford Notation	Structure
9		FA2G2S1	Galβ1-4GlcNAcβ1-2Manα1-6(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4 (Fucα1-6)GlcNAc
10		FA2G2S2	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6 (Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3) Manβ1-4(Fucα1-6)GlcNAcβ1-4GlcNAc
11		M5	Manα1-6(Manα1-3)Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
12		M5A1	Manα1-6(Manα1-3)Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
13		M5A1G1	Manα1-6(Manα1-3)Manα1-6(Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
14		A1G1	Manα1-6(Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
15		FA1	Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4 GlcNAc β1-4(Fucα1-6)GlcNAc

FIG. 23 (continued)

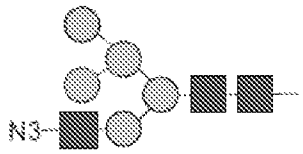
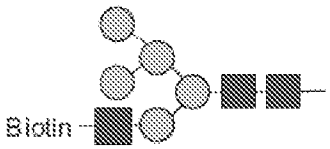
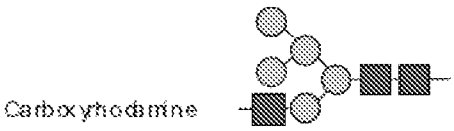
Glycan Number	Symbolic Representation	Oxford Notation	Structure
16		N/A	Manα1-6(Manα1-3)Manα1-6(GlcNAzβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
17		N/A	Manα1-6(Manα1-3)Manα1-6(Biotin-PEG4-DBCO-GlcNAzβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
18		N/A	Manα1-6(Manα1-3)Manα1-6(CR110-DBCO-GlcNAzβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc

FIG. 23 (continued)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US23/10330

A. CLASSIFICATION OF SUBJECT MATTER

IPC - INV. C12N 15/62; C12N 9/10; C12N 15/52 (2023.01)
ADD. C12N 15/67 (2023.01)

CPC - INV. C12N 15/62; C12N 9/1048; C12N 15/52
ADD. C07K 2319/03; C07K 2319/70; C12N 15/67

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic database consulted during the international search (name of database and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005/0112141 A1 (TERMAN, D.) 26 May 2005; Paragraphs [0015], [0016], [0022], [0042], [0049], [0225]	1-9, 38-43, 53
A	WO 2020/056239 A1 (MODERNATX, INC.) 19 March 2020; Paragraphs [0007], [0080], [0207], [0971]	1-9
A	US 7,405,198 B2 (DEFREES SHAWN) 29 July 2008; Columns 68 and 83	1-9, 38-43, 53
A	US 2008/0255040 A1 (DEFREES SHAWN) 16 October 2008; Paragraph [0014]	1-9
A	MIZRACHI, ET AL.. "Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy" Nature Communications. April 2015; Abstract; Page 2; DOI: 10.1038/ncomms7826	38-43, 53

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 March 2023 (27.03.2023)	APR 14 2023
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US23/10330

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed.

   b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),

      ☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US23/10330

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 10-37, 44-52, 54-64
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.