Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2827732 A1 2012/08/30

(21) 2 827 732

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2012/02/20
(87) Date publication PCT/PCT Publication Date: 2012/08/30
(85) Entrée phase nationale/National Entry: 2013/08/19
(86) N° demande PCT/PCT Application No.: US 2012/025809
(87) N° publication PCT/PCT Publication No.: 2012/115903
(30) Priorité/Priority: 2011/02/25 (US61/446,846)

(51) Cl.Int./Int.Cl. *C12N 1/19* (2006.01), *C12N 15/00* (2006.01), *C12N 15/54* (2006.01), *C12N 15/61* (2006.01), *C12N 15/79* (2006.01), *C12N 15/81* (2006.01), *C12P 21/02* (2006.01)

(71) Demandeur/Applicant:
MERCK SHARP & DOHME CORP., US

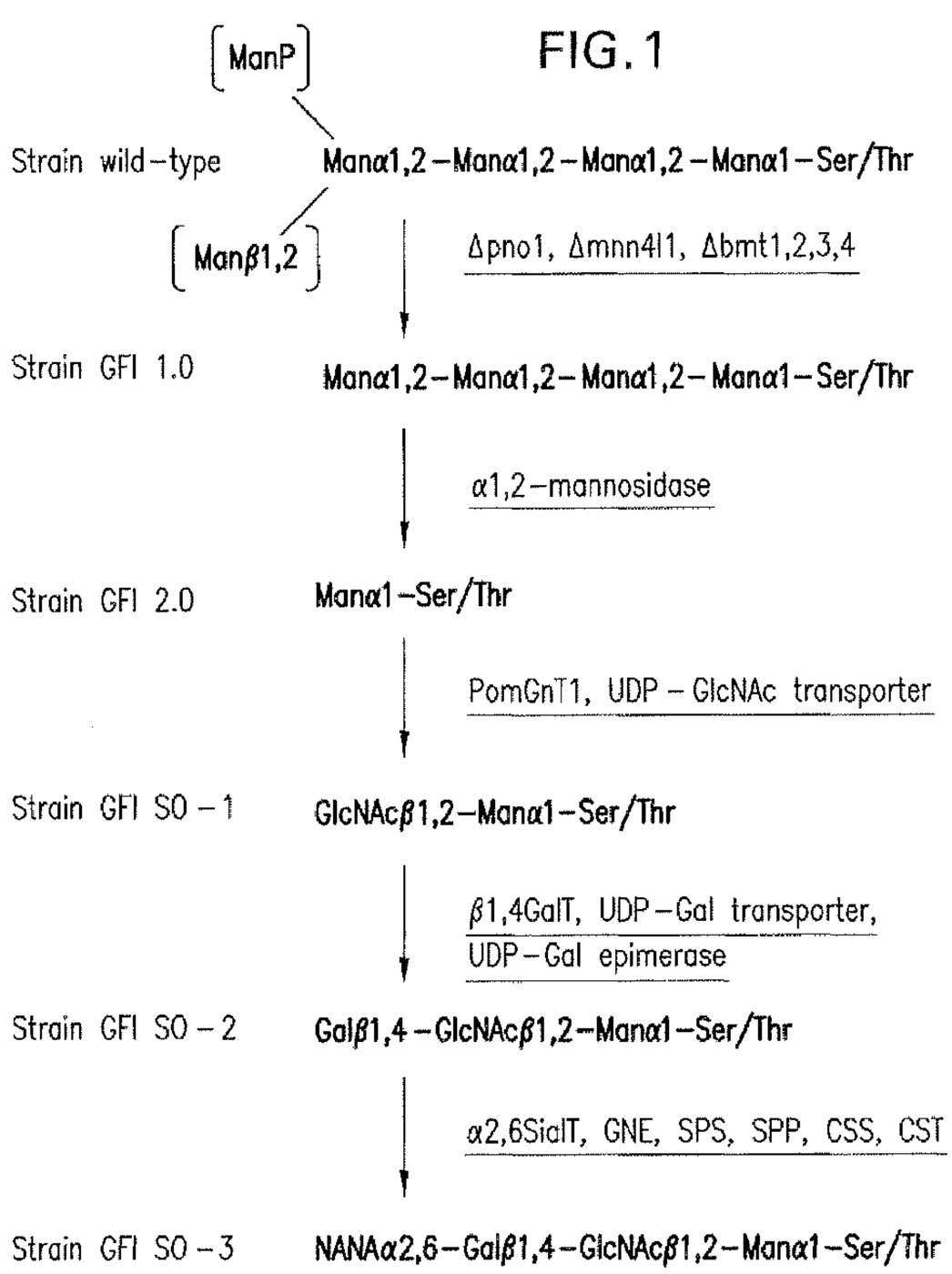
(72) Inventeurs/Inventors:
BOBROWICZ, PIOTR, US;
COOK, WILLIAM J., US;
HAMILTON, STEPHEN, US;
NETT, JUERGEN, US;
STADHEIM, TERRANCE A., US;
WILDT, STEFAN, US

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : SOUCHE DE LEVURE POUR LA PRODUCTION DE PROTEINES AVEC UNE O-GLYCOSYLATION MODIFIEE

(54) Title: YEAST STRAIN FOR THE PRODUCTION OF PROTEINS WITH MODIFIED O-GLYCOSYLATION

FIG. 1



(57) Abrégé/Abstract:
Lower eukaryotic host cells have been recombinantly engineered to produce glycoprotein having human-like O-glycosylation. The glycoproteins are useful for the production of glycoprotein compositions with advantages for the production of human therapeutics.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 August 2012 (30.08.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/115903 A2

(51) International Patent Classification:
C12P 21/02 (2006.01)

(21) International Application Number:
PCT/US2012/025809

(22) International Filing Date:
20 February 2012 (20.02.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/446,846 25 February 2011 (25.02.2011) US

(71) Applicant *(for all designated States except US)*: **MERCK SHARP & DOHME CORP.** [US/US]; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BOBROWICZ, Piotri** [PL/US]; 111 Blueberry Hill Drive, Hanover, New Hampshire 03755 (US). **COOK, William, J.** [US/US]; 29 Carriage Lane, Hanover, New Hampshire 03755 (US). **HAMILTON, Stephen** [GB/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **NETT, Juergen** [DE/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **STADHEIM, Terrance, A.** [US/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **WILDT, Stefan** [DE/US]; 450 North End Avenue, Apt. 5A, New York, New York 10282 (US).

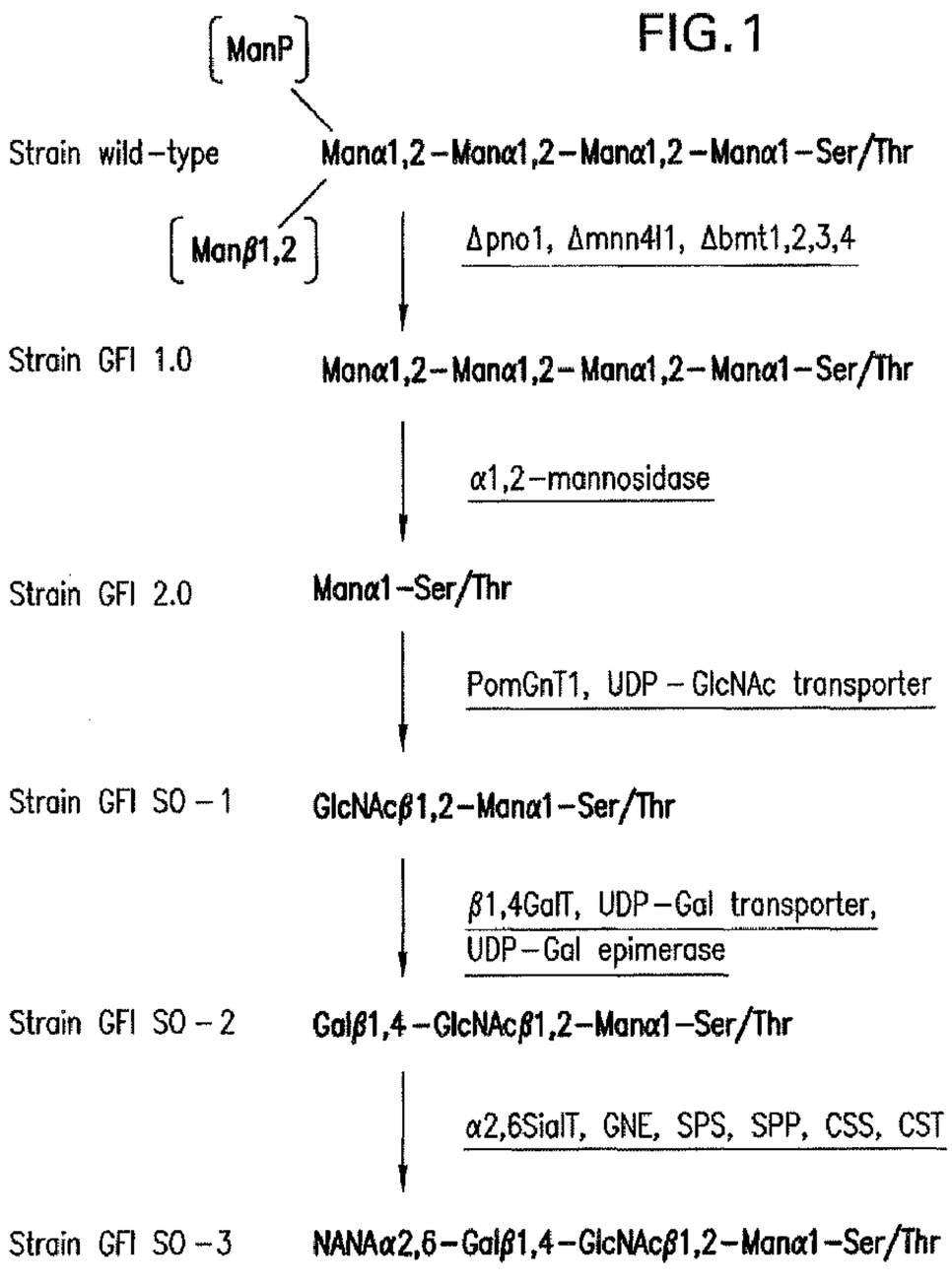
(74) Common Representative: **MERCK SHARP & DOHME CORP.**; 126 East Lincoln Avenue, Rahway, New Jersey 07065-0907 (US).

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

[Continued on next page]

(54) Title: YEAST STRAIN FOR THE PRODUCTION OF PROTEINS WITH MODIFIED O-GLYCOSYLATION

FIG. 1



(57) Abstract: Lower eukaryotic host cells have been recombinantly engineered to produce glycoprotein having human-like O-glycosylation. The glycoproteins are useful for the production of glycoprotein compositions with advantages for the production of human therapeutics.

WO 2012/115903 A2

WO 2012/115903 A2

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

YEAST STRAIN FOR THE PRODUCTION OF PROTEINS WITH MODIFIED O-GLYCOSYLATION

FIELD OF THE INVENTION

[0001] The present invention relates to the field of molecular biology, in particular the invention is concerned with lower eukaryotic cells, such as yeast strains, genetically engineered to produce glycoproteins having humanized O-glycosylation and their production from recombinant expression systems.

BACKGROUND OF THE INVENTION

[0002] Significant structural differences exist between the O-glycosylation pathways and patterns of lower eukaryotes (*i.e.,* yeast and filamentous fungi) and mammals (particularly humans). Because the human immune system may recognize the alternative glycosylation of lower eukaryotes as foreign, any protein-based therapeutic products produced in fungal systems have the potential to provoke an immunogenic response when injected into humans. This response may limit the effectiveness of a therapeutic over multiple administrations and, in the most serious cases, may cause adverse effects in the patient.

[0003] In fact, the presence of fungal glycosylation is a common signal for clearance by the innate human immune system; *see, e.g.,* Ballou, C.E., 1990 *Methods Enzymol.* 185:440. This raises concern about the potential for rapid clearance or immunogenicity of therapeutic proteins produced in yeast, such as *Pichia pastoris* and injected into humans, including but not limited to proteins having N and O glycans typical of yeast.

[0004] Previous attempts at reducing the immunogenicity of therapeutic proteins produced in yeast have largely focused on reducing the immunogenicity of N glycans (see, e.g., Gerngross US Patent 7,029,872). More recent attempts have focused on reducing or eliminating altogether fungal O-glycosylation in order to reduce or eliminate this response; *see, e.g.,* Tanner, US Patent 5,714,377; and Bobrowicz et al., WO2007/061631. By contrast, the present inventors have surprisingly found unexpected advantages in producing a protein with certain O-glycosylation patterns similar to the O-glycosylation observed on native human glycoproteins, or closer to the O-glycosylation observed on recombinant glycoproteins produced in mammalian cells.

SUMMARY OF THE INVENTION

[0005] The present invention provides methods and materials for the production of recombinant glycoproteins with improved properties useful in the development of human therapeutics or veterinary therapeutic products. In certain embodiments, the invention comprises lower eukaryotic host cells (including yeast and filamentous fungi) which have been engineered to produce glycoproteins having a predominant human-like O-glycan. In preferred embodiments, the host cells produce glycoproteins having predominantly a human-like O-glycan selected from O-Man-GlcNAc, O-Man-GlcNAc-Gal, or O-Man-GlcNAc-Gal-Sia.

[0006] As a first aspect, the present invention provides lower eukaryotic host cells for use in the methods of the present invention which are modified by the knock-out and/or inactivation of one or more genes involved in O-glycosylation, including but not limited to those encoding beta-mannosyltransferases (bmts; *see, e.g.,* US 2006/0211085), phosphomannose transferases, or one or more additional genes involved in O-glycosylation.

[0007] As a second aspect, the present invention provides lower eukaryotic host cells such as those described above for use in the methods of the present invention which are modified to express exogenous (non-native) genes involved in the human or mammalian O-linked glycosylation pathways by transfection or transformation with exogenous genes involved in O-glycosylation, and in particular genes encoding protein O-mannose β-1,2-*N*-acetylglucosaminyltransferase 1 ("PomGnT1") or the catalytic domain thereof as well as one or more additional genes encoding for one or more steps in the human or mammalian O-glycosylation pathways. In particular embodiments, these one or more additional genes encode, without limitation, the following enzymes or catalytic domains thereof: UDP-GlcNAc transporter; α-1,2-mannosidase; β-1,4-galactose transferase ("β1,4GalT"); UDP-galactose transporter (UGT); UDP-Gal epimerase; α-2,6-sialic acid transferase ("α2,6SialT"); α-2,3-sialic acid transferase ("α2,3SialT"); UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase("GNE"); *N*-acetylneuraminate-9-phosphate synthase("SPS"); sialylate-9-P phosphatase ("SPP"); CMP-sialic acid synthase("CSS"); and CMP-sialic acid transporter ("CST").

[0008] As a third aspect, the present invention provides lower eukaryotic host cells such as those described above for use in the methods of the present invention which are modified to express a recombinant glycoprotein of interest by the transfection or transformation of the host cell with an exogenous gene encoding the glycoprotein. Thereby, when cultured under appropriate conditions for expression, the host cells as

describe herein will express and secrete improved recombinant glycoproteins comprising particular human-like O-glycosylation.

[0009] The present invention further comprises methods for producing a recombinant glycoprotein having predominantly a human-like O-glycan, said method comprising a) selecting a lower eukaryotic host cell; b) attenuating one or more endogenous glycosylation enzymes in the host cell (where said host cell of step (a) is not already so attenuated); c) transforming the host cell with nucleic acid sequence encoding an O-linked mannose β1,2-N-acetylglucosaminyltransferase 1 (POMGnT1), a UDP-GlcNAc transporter, an alpha-1,2 mannosidase, and the glycoprotein; and d) culturing the cell under conditions suitable for expression of the nucleic acid sequence to produce the recombinant glycoprotein having predominantly a human-like O-glycan. In particular embodiments, the host cell is further transformed with nucleic acid encoding, without limitation, the following enzymes or catalytic domains thereof: β-1,4-galactose transferase ("β1,4GalT"); UDP-galactose transporter (UGT); UDP-Gal epimerase; α-2,6-sialic acid transferase ("α2,6SialT"); α-2,3-sialic acid transferase ("α2,3SialT"); UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase("GNE"); *N*-acetylneuraminate-9-phosphate synthase("SPS"); sialylate-9-P phosphatase ("SPP"); CMP-sialic acid synthase("CSS"); and CMP-sialic acid transporter ("CST").

[0010] In other embodiments, the present invention comprises recombinant glycoprotein compositions produced from the lower eukaryotic host cells of the invention.

[0011] The lower eukaryotic host cells of the present invention may optionally be engineered to produce glycoproteins having a predominant human-like N-glycan. In preferred embodiments, the host cells produce glycoproteins having a predominant N-glycan selected from: Man5GlcNAc2, GlcNAcMan5GlcNAc2, GalGlcNAcMan5GlcNAc2, SiaGalGlcNAcMan5GlcNAc2, Man3GlcNAc2, GlcNAcMan3GlcNAc2, GalGlcNAcMan3GlcNAc2, SiaGalGlcNAcMan3GlcNAc2, GlcNAc2Man3GlcNAc2, GalGlcNAc2Man3GlcNAc2, Gal2GlcNAc2Man3GlcNAc2, SiaGal2GlcNAc2Man3GlcNAc2, or Sia2Gal2GlcNAc2Man3GlcNAc2.

[0012] In particular embodiments, the glycoprotein compositions provided herein comprise glycoproteins having fucosylated and non-fucosylated hybrid and complex *N*-glycans, including bisected and multiantennary species, including but not limited to *N*-glycans such as $GlcNAc_{(1-4)}Man_3GlcNAc_2$; $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$; $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$.

[0013] In particular embodiments, the glycoprotein compositions provided herein comprise glycoproteins having at least one hybrid *N*-glycan selected from the group consisting of $GlcNAcMan_3GlcNAc_2$; $GalGlcNAcMan_3GlcNAc_2$; $NANAGalGlcNAcMan_3GlcNAc_2$; $GlcNAcMan_5GlcNAc_2$; $GalGlcNAcMan_5GlcNAc_2$; and $NANAGalGlcNAcMan_5GlcNAc_2$. In particular aspects, the hybrid *N*-glycan is the predominant *N*-glycan species in the composition. In further aspects, the hybrid *N*-glycan is a particular *N*-glycan species that comprises about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, or 100% of the hybrid *N*-glycans in the composition.

[0014] In particular embodiments, the glycoprotein compositions provided herein comprise glycoproteins having at least one complex *N*-glycan selected from the group consisting of $GlcNAc_2Man_3GlcNAc_2$; $GalGlcNAc_2Man_3GlcNAc_2$; $Gal_2GlcNAc_2Man_3GlcNAc_2$; $NANAGal_2GlcNAc_2Man_3GlcNAc_2$; and $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$. In particular aspects, the complex *N*-glycan is the predominant *N*-glycan species in the composition. In further aspects, the complex *N*-glycan is a particular *N*-glycan species that comprises about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, or 100% of the complex *N*-glycans in the composition.

[0015] In particular embodiments, the *N*-glycan is fusosylated. In general, the fucose is in an α1,3-linkage with the GlcNAc at the reducing end of the *N*-glycan, an α1,6-linkage with the GlcNAc at the reducing end of the *N*-glycan, an α1,2-linkage with the Gal at the non-reducing end of the *N*-glycan, an α1,3-linkage with the GlcNac at the non-reducing end of the *N*-glycan, or an α1,4-linkage with a GlcNAc at the non-reducing end of the *N*-glycan.

[0016] Therefore, in particular aspects of the above the glycoprotein compositions, the glycoform is in an α1,3-linkage or α1,6-linkage fucose to produce a glycoform selected from the group consisting of $GlcNAcMan_5GlcNAc_2$(Fuc), $GlcNAcMan_3GlcNAc_2$(Fuc), $GlcNAc_2Man_3GlcNAc_2$(Fuc), $GalGlcNAc_2Man_3GlcNAc_2$(Fuc), $Gal_2GlcNAc_2Man_3GlcNAc_2$(Fuc), $NANAGal_2GlcNAc_2Man_3GlcNAc_2$(Fuc), and $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$(Fuc); in an α1,3-linkage or α1,4-linkage fucose to produce a glycoform selected from the group consisting of $GlcNAc(Fuc)Man_5GlcNAc_2$, $GlcNAc(Fuc)Man_3GlcNAc_2$, $GlcNAc_2(Fuc_{1-2})Man_3GlcNAc_2$, $GalGlcNAc_2(Fuc_{1-2})Man_3GlcNAc_2$, $Gal_2GlcNAc_2$(Fuc1-2)Man3GlcNAc2, NANAGal2GlcNAc2(Fuc_{1-2})$Man_3GlcNAc_2$, and $NANA_2Gal_2GlcNAc_2(Fuc_{1-2})Man_3GlcNAc_2$; or in an α1,2-linkage fucose to produce a glycoform selected from the group consisting of $Gal(Fuc)GlcNAc_2Man_3GlcNAc_2$,

$Gal_2(Fuc_{1-2})GlcNAc_2Man_3GlcNAc_2$, $NANAGal_2(Fuc_{1-2})GlcNAc_2Man_3GlcNAc_2$, and $NANA_2Gal_2(Fuc_{1-2})GlcNAc_2Man_3GlcNAc_2$.

[0017] In further aspects of the above, the complex *N*-glycans further include fucosylated and non-fucosylated bisected and multiantennary species.

[0018] In further aspects, the glycoproteins comprise high mannose *N*-glycans, including but not limited to, $Man_5GlcNAc_2$, or *N*-glycans that consist of the $Man_3GlcNAc_2$ *N*-glycan structure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] FIGURE 1 illustrates an overview of a process for generating sialylated O-mannosyl glycans in *Pichia pastoris.*

[0020] FIGURES 2A-B illustrate TNFRII-Fc O-glycans from Strain GFI SO-1 (see Figure 1) compared to that from Strain GFI 2.0. Shown are traces from HPAEC-PAD (see below for method) for O-glycans on a reporter protein (TNFRII-Fc) expressed in strain YGLY6428 (Panel A) which has no POMGnT1, and strain YGLY7879 (Panel B) which harbors the mouse POMGnT1-ScMNN6-s fusion. Results indicate a POMGnT1-dependent O-glycan in YGLY7879 that co-migrates with the arabitol standard. Arabitol is added to the samples at a concentration that typically gives a maximum reading of ~100 nano coulombs (nC) in strains lacking POMGnT1. However, with co-elution of both arabitol and Man-GlcNAc, which have an additive effect, the magnitude is increased to approximately 450nC in strains harboring POMGnT1. This increase in magnitude is highlighted by the circles positioned on the Y-axis.

[0021] FIGURES 3A-B illustrate TNFRII-Fc O-glycans from Strain GFI SO-1 (YGLY7879 [mouse POMGnT1-ScMNN6-s]) following hexosaminidase treatment to remove GlcNAc. Shown are HPAEC-PAD traces of glycans from strain YGLY7879 without (PANEL A) and with (PANEL B) hexosaminidase treatment to remove terminal GlcNAc. The treatment results in the reduction of the putative arabitol + Man-GlcNAc peak, appearance of a new peak at T=23.909 which is free GlcNAc, and an increase in the manitol (man1) peak. This verifies that the sugar co-migrating with arabitol in POMGnT1-containing GFI SO-1 strains is Man-GlcNAc.

[0022] FIGURES 4A-B illustrate TNFRII-Fc O-glycans from Strain GFI SO-2 compared to that from Strain GFI 2.0. Shown are HPAEC-PAD traces of TNFRII-Fc O-glycans from GFI 2.0 strain YGLY6428 (PANEL A) which has no POMGnT1, and GFI SO-2 strain YGLY7880 (PANEL B) which harbors galactose transfer genes and human POMGnT1-ScMNN2-s. Results indicate a novel POMGnT1-dependent O-glycan in YGLY7880 migrating at ~T=21.38 that is Man-GlcNAc-Gal. Note, the peak eluting at T=21.38 has been described as a sorbitol peak by the instrument, as this peak elutes within the range of the sorbitol standard peak, which has been defined during calibration. It is obvious from Figure 5 that this peak is Man-GlcNAc-Gal, due to its hydrolysis with galactosidase and hexosaminidase treatments.

[0023] FIGURES 5A-B illustrate TNFRII-Fc O-glycans from GFI SO-2 strain YGLY7880 treated with galactosidase and hexosaminidase. Shown are HPAEC-PAD traces of TNFRII-Fc O-glycans from GFI SO-2 strainYGLY7880 untreated (PANEL A), or treated with galactosidase to remove terminal galactose plus hexosaminidase to remove terminal GlcNAc (PANEL B). The treatment results in the elimination of the putative Man-GlcNAc-Gal peak, appearance of a peak for free GlcNAc at T=23.909, an increase in the manitol (man1) peak, and the appearance of a galactose peak at T=38.35. These results indicate that the POMGnT1-dependent O-glycan in GFI SO-2 strains that migrates at ~T=21.38 is Man-GlcNAc-Gal. Note, the peak eluting at T=21.38 has been labeled as a sorbitol peak by the instrument, see Figure 4.

[0024] FIGURES 6A-B illustrate TNFRII-Fc O-glycans from GFI SO-3 strain YGLY8750 (musPOMGnT1-ScMNN6-s), treated with neuraminidase, galactosidase and hexosaminidase. Shown are HPAEC-PAD traces of TNFRII-Fc O-glycans from YGLY8750 untreated (PANEL A), or treated with neuraminidase to remove terminal sialic acid, galactosidase to remove terminal galactose, plus hexosaminidase to remove terminal GlcNAc (PANEL B). A sub-optimal dose of enzymes was added in order to generate all possible species. The treatment results in the reduction of the putative Man-GlcNAc-Gal-Sialic acid peak at T=45.8, the reduction of the Man-GlcNAc-Gal peak at T=21.7, appearance of a peak for free GlcNAc at T=23.5, an increase in the manitol (man1) peak, and the appearance of a galactose peak at T=37.9. These results indicate that the POMGnT1-dependent O-glycan in GFI SO-3 strains that migrates at ~T=45.8 is Man-GlcNAc-Gal-sialic acid.

[0025] FIGURE 7 illustrates Western blotting with lectin SNA-1-AP to detect terminal sialic acid. Shown is a coomassie stained SDS-PAGE gel (bottom) showing TNFRII-Fc protein levels in samples 1-12, and a Western blot (top) using detection by alkaline phosphatase-conjugated lectin SNA-1. Lanes 1-9 are supernatants (2, 6, and 10 uL per strain) from three separate colonies of GFI SO-2 strain YGLY7880 transformed with sialylation vector pGLY1758; lanes 10-12 are from non-transformed YGLY7880. Lectin SNA-1 binds to terminal sialic acid, and thus will detect terminal sialic acid on O-glycans in Strain GFI SO-3 that harbors genes for sialic acid synthesis and transfer. Results indicate strong lectin binding to TNFRII-Fc from samples 1-9 which are from GFI SO-3 strains, but no binding to TNFRII-Fc from samples 10-12 which are from GFI SO-2 strains that lack the ability to transfer sialic acid.

[0026] FIGURES 8A-B illustrate TNFRII-Fc O-glycans from GFI SO-3 strain YGLY11603 (musPOMGnT1-ScMNN6-s) which has been optimized for O-sialylation as described in the text. PANEL A shows HPAEC-PAD traces of TNFRII-Fc O-glycans; PANEL B lists percentages of each major O-glycan species.

[0027] FIGURE 9 illustrates how increased O-sialylation of a glycoprotein results in increased bioavailability in B6 mice. Serum concentrations of mice 48 hours after subcutaneous administration of different doses of TNFRII-Fc are shown. Total sialic acid content (mol/mol) was as follows: Form 5-A: 21; Form 5-B: 16; and Form 5-C: 11.

[0028] FIGURES 10A-B illustrate how increased O-sialylation of a glycoprotein results in increased serum half-life in rats. Serum time-concentration curves following intravenous ("IV") [Figure 10A] and subcutaneous ("SC") [Figure 10B] administrations at 1 mg/kg are shown. Total sialic acid content (mol/mol) was as follows: Form 5-A: 21; Form 5-B: 16; and Form 5-C: 11.

[0029] FIGURE 11 illustrates the TD-POMGnT1 expression/integration vector pGLY3136.

BRIEF DESCRIPTION OF THE SEQUENCES

[0030] SEQUENCE ID NOs: 1-108 illustrate various ER and/or Golgi localization leader sequences.

[0031] SEQUENCE ID NOs: 109-118 illustrate POMGnT1 sequences.

[0032] SEQUENCE ID NOs: 119-120, 122-123, 125-126, 133-134, 135-136 and 141-148 illustrate PCR primers utilized in Examples.

[0033] SEQUENCE ID NOs: 121 and 124 illustrate *PpHI3* ORF and 3' untranslated fragments, respectively.

[0034] SEQUENCE ID NO: 127 illustrates a DNA fragment encompassing the *PpALG3* transcriptional termination sequence.

[0035] SEQUENCE ID NO: 128 illustrates a *PpGAP* promoter sequence.

[0036] SEQUENCE ID NO: 129 illustrates a *ScCYC1* transcriptional terminator sequence.

[0037] SEQUENCE ID NO: 130 illustrates a *PpAOX1* promoter sequence.

[0038] SEQUENCE ID NOs: 131-132 and 138-140 illustrate sequences comprising TNFR-IgG1 sequence.

[0039] SEQUENCE ID NO: 137 illustrates sequence encoding the HSA pre-signal peptide.

DETAILED DESCRIPTION OF THE INVENTION

[0040] The present invention relates to methods and materials for generating glycoproteins having human O-glycosylation in lower eukaryotes, *e.g.,* yeast (including but not limited to *Pichia pastoris*), or filamentous fungi which do not inherently have human O-glycosylation synthesis and transfer machinery. Recombinant glycoproteins produced using the disclosed methods and materials bear glycosylation patterns common with human or mammalian-produced proteins and, as such, are able to be injected into humans with greatly reduced or eliminated potential for an immunogenic response.

[0041] Applicants have found, moreover, that humanizing the O-glycans can enhance the bioactivity (such as bioavailability and serum half-life) of therapeutic proteins particularly sialylated O-glycans, for example, by improving pharmacokinetic properties, hence facilitating better control over *in vivo* drug activity.

[0042] Accordingly, the present invention relates to the development of protein expression systems for yeasts and filamentous fungi, such as *Pichia pastoris*, based on improved vectors, novel nucleic acids, host cell lines, and methods for using the foregoing in the production of recombinant glycoproteins with decreased immunogenicity and/or better pharmacokinetic properties.

[0043] The present inventors have found that improved glycoprotein therapeutics can be obtained in recombinant lower eukaryotic host cells by modifying the glycosylation machinery present in the lower eukaryotic cells. The cells as will be described herein are or have been genetically modified (*e.g.*, through particular deletions and insertions) so as to specifically produce recombinant glycoproteins having certain human like O-glycosylation. More specifically, Applicants found that, through the steps of (i) attenuating the lower eukaryotic host cell's endogenous O-glycosylation pathway (or, alternatively, exploiting host cells so attenuated) and (ii) expressing heterologous genes encoding particular O-glycosylation enzymes, the host cell was able to produce proteins having particular human-like O-glycans instead of the native, lower eukaryotic immunogenic O-glycans. Following the changes described and transfection with an exogenous gene of interest, Applicants found that the modified lower eukaryotic host cells were able to produce glycoproteins (including therapeutic glycoproteins) with particular human-like O-glycosylation (and, in particular embodiments human-like N-glycosylation as well). Figures 9 and 10 illustrate how increased O-sialylation of a glycoprotein as disclosed in the present invention results in increased bioavailability in B6 mice and both increased bioavailability and serum half-life in rats. In addition, the lower eukaryotic host cells, such as *Pichia pastoris*, are able to produce therapeutic glycoproteins in high titers,

with the predominant species of glycoprotein having human O-glycans (and optionally N-glycans) which exhibit improved efficacy compared to therapeutic glycoprotein produced in mammalian cells or in lower eukaryotic host cells retaining the endogenous glycosylation machinery.

[0044] Although a previous invention from Bobrowicz et al., WO2007/061631, has described methods for reducing fungal O-glycans to a single O-linked mannose, it may be advantageous to produce a protein that has similar or identical O-glycosylation to certain O-glycoforms observed from mammalian cell culture or of human origins. For purposes of the present invention, such proteins with similar or identical O-glycosylation to certain O-glycoforms observed from mammalian cell culture or of human origin ("human-like" O-glycans for purposes herein) are proteins having as the predominant O-glycan species, glycans comprising (1) a terminal GlcNAc linked to a single mannose residue which is linked to a serine or threonine residue on the protein (*i.e.,* O-Man-GlcNAc); (2) a terminal GlcNAc-Gal (*i.e.,* galactose linked to the terminal GlcNAc referred to above) linked to a single mannose residue which is linked to a serine or threonine residue on the protein (*i.e.,* O-Man-GlcNAc-Gal); or (3) terminal GlcNAc-Gal-Sia (*i.e.,* sialic acid linked to the terminal GlcNAc-Gal referred to in (2)) linked to a single mannose residue which is linked to a serine or threonine residue on the protein (*i.e.,* O-Man-GlcNAc-Gal-Sia). These structures are closely related to O-glycan structures typically observed on nerve cells and other brain-related or neural-related cells (See, Zamze et al., *Glycobiology;* 9:823-831 (1999). The most well studied example of the mammalian sialylated O-mannose glycan is on alpha-dystroglycan ("Alpha-DG"); Chiba *et al.*, 1997 *J Biol. Chem* 272:2156; Sasaki *et al.*, 1998 *Biochem. Biophys. Acta* 1425:599. Alpha-DG contains the O-glycan (sialic acid-alpha-2,3-Gal-beta-1,4-GlcNAc-beta-1,2-mannose) in "high abundance" on Ser/Thr within a mucin-like domain located at the N-terminal end of mature alpha-DG. Manya *et al.*, 2007 *J Biol Chem* 282:20200, identified at least some of the O-mannosylated residues.

[0045] The present invention provides methods and materials useful for producing recombinant glycoproteins bearing predominantly human-like O-glycan structures (as defined herein) in lower eukaryotic expression systems. Lower eukaryotes such as yeast are preferred for expression of glycoproteins because they can be economically cultured, give high yields, and when appropriately modified are capable of suitable glycosylation. Various yeasts, such as *Kluyveromyces lactis, Pichia pastoris, Pichia methanolica,* and *Hansenula polymorpha* are preferred for cell culture because they are able to grow to high cell densities and secrete large quantities of recombinant protein. In specific embodiments, the lower eukaryote is *Pichia finlandica, Pichia*

trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia minuta (Ogataea minuta, Pichia lindneri), Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia guercuum, Pichia pijperi or *Pichia stiptis.* Likewise, filamentous fungi, such as *Aspergillus niger, Fusarium sp, Neurospora crassa* and others can be used to produce glycoproteins of the invention at an industrial scale.

[0046] As a first aspect, the present invention provides lower eukaryotic host cells for use in the methods of the present invention which are modified by the knock-out and/or inactivation of one or more genes involved in O-glycosylation, including but not limited to those encoding beta-mannosyltransferases (bmts; *see, e.g.,* US 2006/0211085 and Trimble RB *et al.,* 2003 *Glycobiology* 14:265-274), phosphomannose transferase (*see, e.g.,* Trimble RB *et al.,* 2003 *Glycobiology* 14:265-274), or one or more additional genes involved in O-glycosylation. In particular embodiments, the one or more genes involved in O-glycosylation which are knocked out and/or inactivated encode enzymes selected from: beta-mannosyl transferases (*e.g., BMT 1, 2, 3* and *4* genes; *see, e.g.,* Mille *et al.*, 2008 *J. Biol. Chem.* 283:9724-9736; US 2006/0211085) or phospho-mannose transferases (*e.g., MNN4A* [also known as *MNN4*], *MNN4B* [also known as *MNN4L1 or MNNS*; *see, e.g.,* US 7,259,001] and *PNO1* [*see, e.g.,* US 7,198,921] genes; Li *et al.*, 2006 *Nat. Biotechnol.* 24:210-215). In specific embodiments, genes encoding Ktr and Kre2/Mnt1 mannosyltransferases, including Ktr1 and Ktr3 (*KTR1* and *KTR3* genes) and alpha1,2-mannosyltransferase Kre2 (*KRE2* gene, *see, e.g.,* US Patent 7,217,548) are left intact or in a state that enables them to produce functional protein. In alternative embodiments, the gene encoding Ktr1 is deleted or inactivated while the other MNT genes are left in a state that enables them to produce functional protein. It has been found that inactivation of *KTR1* increases level of O-Man1 2-fold. In specific embodiments, at least one of the genes encoding Mnn4a, Mnn4b and Pno1 are deleted. In alternative embodiments, genes encoding Ktr1, Ktr3 and Kre2 are deleted or inactivated.

[0047] Disruption of genes can be done by any method suitable in the art to disrupt the gene and/or translation into the encoded protein; and includes but is not limited to disrupting the open reading frame of the gene, disrupting expression of the open reading frame or abrogating translation of RNAs encoding the intended protein and using interfering RNA, antisense RNA, or the like. Suitable methods include, for example, the procedure of Rothstein, 1991 *Methods Enzymol.* 194:281-301, or the PCR-mediated approach of Baudin *et al.*, 1993 *Nucleic Acids Res.* 21:3329-3330. In specific embodiments, alternative genes in the above families not specifically mentioned above are left intact (*i.e.,* not disrupted).

[0048] As a second aspect, the present invention provides host cells (in particular, those described above in the first aspect) for use in the methods of the present invention which are modified to express recombinant, exogenous (non-native) genes involved in the human or mammalian O-linked glycosylation pathways. In specific embodiments, the host cells express an extra copy of each of the exogenous genes involved in the human or mammalian O-linked glycosylation pathways. The cells may be obtained through transfection or transformation of cells with exogenous genes involved in O-glycosylation, and in particular genes encoding protein O-mannose β-1,2-*N*-acetylglucosaminyltransferase 1 ("PomGnT1"), or the catalytic domain thereof, as well as one or more additional genes encoding for enzymes responsible for one or more steps in the human or mammalian O-glycosylation pathways, including but not limited to the following enzymes or the catalytic domains thereof: UDP-GlcNAc transporter; α-1,2-mannosidase (*see, e.g.,* WO 2007/061631); β-1,4-galactose transferase ("β1,4GalT"); UDP-galactose transporter (UGT); UDP-Gal epimerase; α-2,6-sialic acid transferase ("α2,6SialT"); α-2,3-sialic acid transferase ("α2,3SialT"); UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase ("GNE"); *N*-acetylneuraminate-9-phosphate synthase ("SPS"); sialylate-9-P phosphatase ("SPP"); CMP-sialic acid synthase ("CSS"); and CMP-sialic acid transporter ("CST"). In additional embodiments, nucleic acid encoding genes for *O*-fucosylation may be added. *O*-fucose glycans may, in specific embodiments, range from the monosaccharide Fuc-*O*-Ser/Thr to the tetrasaccharide NeuAcα2, 3/6 Galβ1, 3 Fuc-*O*-Ser/Thr and the di- and tri-intermediates. In specific embodiments, the host cell is transformed with nucleic acid encoding genes for adding fucose, NeuAcα2, 3/6 Galβ1, or Fuc-*O*-Ser/Thr (GDP-fucose protein *O*-fucosyltransferase; POFUT1); and/or one or more genes for adding GlcNAc-linked β1,3 to Fuc-*O*-Ser/Thr: Manic, Radical and/or Lunatic Fringe. Nucleic acid introduced may be operatively linked to a fungal-derived promoter and transcription terminator as well as a fungal-derived leader sequence having a signal sequence and an endoplasmic reticulum ("ER")- or Golgi-localizing transmembrane domain. The genes encoding PomGnT1, β1,4GalT, α2,6SialT and α2,3SialT, in particular embodiments, are operatively linked to a fungal-derived (in particular embodiments, that derived from *Saccharomyces cerevisiae*, *Pichia pastoris*, or *Kluyveromyces lactis*) promoter and transcription terminator as well as a fungal-derived leader sequence having a signal sequence and an endoplasmic reticulum ("ER")- or Golgi-localizing transmembrane domain. The signal sequence serves to direct the nascent protein to the ER. The transmembrane domain is utilized to localize and anchor the protein to an ER or Golgi membrane. The gene encoding α-1,2-mannosidase is operatively linked to a fungal-based (in

particular embodiments, *Pichia*-derived) promoter and transcription terminator as well as a fungus-derived signal sequence (in specific embodiments wherein the signal sequence is the *Saccharomyces* alphaMAT pre-signal sequence). The genes encoding UDP-GlcNAc transporter, UDP-galactose transporter, UDP-Gal epimerase, GNE, SPS, SPP, CSS and CST are operatively linked to a fungal-based (in particular embodiments, *Pichia*-derived) promoter and transcription terminator. In particular embodiments, the promoter utilized is the *Pichia* AOX1 promoter or the GAP promoter. In other embodiments, the promoter utilized is the *Pichia* TEF promoter or the PMA promoter; *see, e.g.,* Hamilton *et al.*, 2006 *Science* 303:1441-1443. In specific embodiments, the α2,6SialT gene is operatively linked to the *Pichia* TEF promoter. In specific embodiments, the CST gene is operatively linked to the PMA promoter.

[0049] In specific embodiments, the POMGnT1 enzyme or catalytic domain encoded is that of, or derived from that of, human, mouse or frog.

[0050] In specific embodiments, the POMGnT1 enzyme comprises sequence SEQ ID NO: 110, SEQ ID NO: 112, or SEQ ID NO: 118. In specific embodiments, the POMGnT sequence is selected from: SEQ ID NO: 109, SEQ ID NO: 111, or SEQ ID NO: 117. SEQ ID NOs: 109, 111 and 117, codon-optimized sequences encoding the POMGnT1 enzyme form specific embodiments of the present invention.

[0051] In specific embodiments, the UDP-GlcNAc transporter is that of, or derived from that of, *Kluyveromyces lactis* or mouse UDP-GlcNAc transporter.

[0052] The alpha 1,2-mannosidase of the cell lines and methods of the present invention must be active on O-linked alpha 1,2-mannose. In specific embodiments, the alpha 1,2-mannosidase is a fungal alpha 1,2-mannosidase and in specific embodiments is that of, or derived from that of, *Trichoderma reesei* alpha-1,2-mannosidase, *Saccharomyces sp.*, *Coccidiodes* species (*e.g., C. immitis* such as that described in accession no. EAS32290, or *C. posadasii* mannosidase such as that described in accession no. ABA54911; *see, e.g.,* U.S. application serial no. 61/369157, filed July 30, 2010 [Attorney Docket No. GF2010.7158L01US]) or *Aspergillus sp.* (*see, e.g.*, Bobrowicz *et. al.*, WO2007/061631) alpha-1,2-mannosidase. In specific embodiments, the alpha-1,2-mannosidase is under the control of the *Pichia* AOX1 promoter. Thus, upon switching the methanol media, the mannosidase is expressed generating single mannose structures.

[0053] Cells in accordance with the present invention may be generated utilizing procedures available in the art, within publications referenced herein, and as described herein; *see, e.g.,* Hamilton *et al.*, 2006 *Science* 313: 1441-1443; and WO 07/136752.

[0054] Expression of exogenous genes in a host requires the use of regulatory regions functional in such hosts, whether native or non-native, or introduced or present in cell line utilized. A wide variety of transcriptional, translational, localization and other regulatory sequences can be employed. In specific embodiments, the regulatory sequences are those of other yeast or filamentous fungi. In specific embodiments, the regulatory sequences are vertebrate, including but not limited to frog, murine or human sequences.

[0055] In particular instances, as described above, the exogenous genes are operatively linked to an endoplasmic reticulum ("ER") or Golgi localization sequence, *see, e.g.,* Choi et al., 2003 PNAS 100:5022, and US 7,449,308. The ER or Golgi leader/localization sequence utilized targets the sequence operatively linked therewith to the endoplasmic reticulum or the early Golgi. In specific embodiments, the ER/Golgi leader sequence is a membrane targeting region from an ER and/or Golgi residing membrane protein of *S. cerevisiae*, *P. pastoris*, or *K. lactis*. In specific embodiments, the ER/Golgi leader sequence comprises a signal sequence to direct the nascent protein to the ER and a localization signal consisting essentially of: (i) the cytoplasmic tail and a segment of the transmembrane domain capable of retaining the operatively linked protein in the ER and/or Golgi for a period of time to permit the desired function; (ii) part or all of the entire stem region; or (iii) the entire stem region and optionally parts of the catalytic domain. In specific embodiments, the ER and/or Golgi leader sequence is an ER and/or Golgi leader sequence, or derived from an ER and/or Golgi leader sequence of a protein selected from: ScMNS1-s, ScSEC12-m, ScMNN9-s, ScKRE2-s, ScMNN2-s or ScMNN6-s. In specific embodiments, the ER and/or Golgi leader sequence is derived from an ER and/or Golgi leader sequence of a protein selected from: ScMNS1-s, ScMNN9-s, PpKRE2-s, K1GNT1-s, ScMNN2-s, ScMNN2-m, ScMNN5-s, ScMNN6-s, ScPMT5-m, PpPMT1-m or PpBET1-s. Certain leaders falling into this category are disclosed in the prior art, *see, e.g.,* Choi et al., 2003 PNAS 100:5022 and US 7,449,308. The disclosed leaders and combinations thereof comprising the leader and exogenous gene as described herein, optionally containing a transcription terminator sequence, form specific embodiments hereof. Specific embodiments relate to the *PMT, BET1, BOS1* and *SEC22* leaders and leader combinations as described herein. In particular, the present invention relates to leaders and their corresponding SEQ ID NOs disclosed in Table 1, in specific embodiments, Leader #s 33-34, 27-32 and 35-54, as well as combinations comprising the foregoing leaders and exogenous genes.

[0056] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, the leader sequence of a protein selected from: S. *cerevisiae*

alpha-glucosidase I encoded by *GLS1*, *S. cerevisiae* alpha-1,2-mannosidase encoded by *MNS1*, *S. cerevisiae* and *P. pastoris* nucleotide exchange factor encoded by *SEC12* which cycles between the ER and Golgi, *S. cerevisiae* and *P. pastoris* protein O-mannosyltransferases encoded by *PMT 1,2,3,4,5&6* that encode integral ER membrane spanning domains, or SNARE proteins encoded by *BOS1*, *BET1*, and *SEC22* from *S. cerevisiae* and *P. pastoris*. The region typically is an amino-terminal signal sequence followed by a transmembrane or other anchoring domain. In specific embodiments, the sequences listed immediately above are used as ER localization sequences.

[0057] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, the leader sequence of a protein selected from: *P. pastoris* Och1p, *S. cerevisiae* Mnn9p, Van1p, Anp1p, Hoc1p, Mnn10p or Mnn11p (and in particular embodiments is an amino-terminal fragment thereof). In specific embodiments, the sequences listed immediately above are used as early or cis-Golgi localization sequences.

[0058] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, the leader sequence of a protein selected from: *S. cerevisiae* Mnn1p or Mnn6p (and in particular embodiments is an amino-terminal fragment thereof). In specific embodiments, the sequences listed immediately above are used as late Golgi localization sequences.

[0059] Table 1 provides specific embodiments of ER and/or Golgi leader sequences contemplated herein. Additional sequences (with SEQ ID NOs) are also provided herein. Use of any of SEQ ID NOs: 1-108 is contemplated herein. Table 2 provides results achieved upon fusing particular disclosed leader sequences to human "Hs", murine "Mm" or frog "Xen" POMGnT1 catalytic domain sequences and expressing them from the methanol-inducible AOX1 promoter. The present invention contemplates the use of human, murine, or frog POMGnT1 sequences (inclusive but not limited to the catalytic domain sequences) and expressing them from a suitable promoter, including but not limited to the methanol-inducible AOX1 promoter or the GAP promoter.

[0060] In specific embodiments, the ER and/or Golgi leader sequence is one selected from: Tables 1 or 2, SEQ ID NOs: 1-108, or is derived from the leader sequence of ScMNN6-s, ScPMT1-s, ScPMT2-s, ScPMT3-s, ScPMT4-s, ScPMT5-s or ScPMT6-s.

[0061] In particular embodiments, human POMGnT1 sequence or a catalytic domain sequence thereof is fused to a leader sequence which is that of, or derived from that of, a protein selected from: PpSEC12-s, ScMNN9-s, K1GNT1-s, ScMNN2-s, ScMNN2-m, ScPMT5-m, PpPMT1-m or PpBET1-S. In specific embodiments,

human POMGnT1 sequence or a catalytic domain sequence thereof is fused to a leader sequence selected from: SEQ ID NO 3 (or sequence encoding SEQ ID NO 4); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 73 (or sequence encoding SEQ ID NO 74); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88) or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

[0062] In particular embodiments, murine POMGnT1 sequence or catalytic domain sequence thereof is fused to a leader sequence which is that of, or derived from that of, a protein selected from: ScMNN9-s, PpKRE2-s, ScKTR2-s, K1GNT1-s, ScMNN2-s, ScMNN2-m, ScMNN5-s, ScMNN6-s, ScPMT5-m, ScPMT6-m, PpPMT1-s, PpPMT2-s, PpPMT4-s, or PpBET1-s. In specific embodiments, murine POMGnT1 sequence or a catalytic domain sequence thereof is fused to a leader sequence selected from: SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 37 (or sequence encoding SEQ ID NO 38); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 45 (or sequence encoding SEQ ID NO 46); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 73 (or sequence encoding SEQ ID NO: 74); SEQ ID NO 75 (or sequence encoding SEQ ID NO 76); SEQ ID NO 77 (or sequence encoding SEQ ID NO 78); SEQ ID NO 79 (or sequence encoding SEQ ID NO 80); SEQ ID NO 81 (or sequence encoding SEQ ID NO 82); or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

[0063] In particular embodiments frog POMGnT1 sequence or catalytic domain sequence thereof is fused to a leader sequence which is that of, or derived from that of, a protein selected from: ScMNS1-s, ScSEC12-m, PpSEC12-s, ScMNN9-s, ScANP1-s, ScHOC1-s, ScMNN10-s, ScMNN11-s, PpKRE2-s, ScKTR2-s, K1GNT1-s, ScMNN2-s, ScMNN2-m, ScMNN1-s, ScMNN6-s, ScPMT5-m, PpPMT1-m, or PpBET1-s. In specific embodiments, frog POMGnT1 sequence or a catalytic domain sequence thereof is fused to a leader sequence selected from: SEQ ID NO 3 (or sequence encoding SEQ ID NO 4); SEQ ID NO 7 (or sequence encoding SEQ ID NO 8); SEQ ID NO 9 (or sequence encoding SEQ ID NO 10); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 17 (or sequence encoding SEQ ID NO 18); SEQ ID NO 19 (or sequence encoding SEQ ID NO 20); SEQ ID NO 21 (or sequence encoding SEQ ID NO 22); SEQ ID NO 23 (or sequence encoding SEQ ID NO 24); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 37 (or sequence encoding SEQ ID NO 38); SEQ ID NO 39 (or sequence encoding SEQ ID

NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 49 (or sequence encoding SEQ ID NO 50); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 73 (or sequence encoding SEQ ID NO 74); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88); or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

[0064] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, a protein selected from: ScMNN1-s, ScMNN9-s, PpKre2-s, K1Gnt1-s, ScMNN2-s, ScMNN2-m, ScMNN5-s, ScMNN6-s, ScPMT5-m, PpPMT1-m or PpBET1-s. In specific embodiments, the ER and/or Golgi leader sequence is selected from: SEQ ID NO 49 (or sequence encoding SEQ ID NO 50); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 45 (or sequence encoding SEQ ID NO 46); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 93 (or sequence encoding SEQ ID NO 94); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88); or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

[0065] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, a protein selected from: PpSEC12, ScMNN9, PpKRE2, ScKTR2, K1GNT1, ScMNN2, ScMNN6, ScPMT5, PpPMT1 or PpBET1.

[0066] In specific embodiments, the ER and/or Golgi leader sequence is that of, or derived from that of, a protein selected from: ScMNN9; K1GNT1, ScMNN2, ScPMT5 or PpBET1.

[0067] The exogenous gene fused to the ER and/or Golgi leader sequence is operatively linked to a promoter. The promoter may be any promoter element which drives expression of the exogenous gene to which it is fused including, but not limited to, the *Pichia* AOX1 promoter or the GAP promoter. In particular preferred embodiments, where the leader is ScMNN2-s, ScMNN2–m, ScPMT5-m, or PpPMT1-m, the promoter is AOX1.

[0068] Particular embodiments comprise the following promoter-leader-POMGnT1 sequence combinations: AOX1-PpKRE2s-mouse POMGnT1; GAP-K1GNT1s-mouse POMGnT1; AOX1-K1GNT1s-mouse POMGnT1; GAP-K1GNT1s-frog POMGnT1; AOX1-ScMNN2s-mouse POMGnT1; GAP-ScMNN5s-mouse POMGnT1; AOX1-ScMNN5s-mouse POMGnT1; GAP-ScMNN6s-mouse POMGnT1; or AOX1-ScPMT5m-mouse POMGnT1.

[0069] Particular embodiments comprise the GAP-ScMNN6s-human POMGnT1 promoter-leader-POMGnT1 sequence combination.

[0070] The exogenous genes may also be operatively linked to a transcription termination sequence including but not limited to that of ScCYC1 and PpAOX1.

[0071] The exogenous gene may optionally be operatively linked to additional regulatory sequences, including but not limited to signal sequences.

[0072] These leaders performed well in assays designed to measure GlcNAc transfer onto O-Man1, for example by lectin staining (lectin GS-II) and Dionex-HPLC (HPAEC-PAD) analysis of released O-glycans.

[0073] As a third aspect, the present invention provides host cells (in particular, those described in the first and second aspects) for use in the methods of the present invention which are modified to express the recombinant glycoprotein of interest by the transfection or transformation of the host cell with an exogenous gene encoding the glycoprotein. Thereby, when cultured under appropriate conditions for expression, the host cell will express and secrete improved recombinant glycoproteins comprising particular human-like O-glycosylation.

[0074] The lower eukaryotic host cells of the present invention may be transformed with a recombinant vector comprising a nucleic acid encoding a desired human glycoprotein, such as a human therapeutic drug or a veterinary drug. The nucleic acids can be DNA or RNA, typically DNA. The nucleic acid encoding the glycoprotein is operably linked to regulatory sequences that allow expression and secretion of the glycoprotein. Such regulatory sequences include a promoter and optionally an enhancer upstream, or 5', to the nucleic acid encoding the fusion protein and a transcription termination site 3' or downstream from the nucleic acid encoding the glycoprotein. For secreted glycoproteins, the nucleic acid typically includes a signal peptide. The signal peptide is responsible for targeting the protein to the secretory pathway for glycosylation and secretion, typically the endoplasmic reticulum. The nucleic acid also typically encodes a 5' untranslated region having a ribosome binding site and a 3' untranslated region. The nucleic acid is often a component of a vector replicable in cells in which the glycoprotein is expressed. The vector can also contain a marker to allow recognition of transformed cells. However, some cell types, particularly yeast, can be successfully transformed with a nucleic acid lacking extraneous vector sequences.

[0075] Nucleic acids encoding desired glycoproteins can be obtained from several sources. cDNA sequences can be amplified from cell lines known to express the glycoprotein using primers to conserved regions (see, e.g., Marks et al., *J. Mol. Biol. 581-596* (1991)). Nucleic acids can also be synthesized *de novo* based on sequences in the scientific literature. Nucleic acids can also be synthesized by extension of overlapping oligonucleotides spanning a desired sequence (see, e.g., Caldas et al.,

Protein Engineering, 13, 353-360 (2000)). Using known techniques, the nucleic acid sequences used in the present invention can be codon-optimized for better expression in the host cells of the invention. Preferably, using the degeneracy of the genetic code, the primary amino acid sequence encoded by any DNA sequences encoding the therapeutic glycoprotein will be unchanged by such codon-optimization. If desired, one can add and/or remove one or more N-glycosylation sites by altering the chimeric DNA sequence used to express the primary amino acid sequence of the recombinant glycoprotein. In preferred embodiments, the host cells are transformed with a vector comprising nucleic acid encoding a soluble TNF-receptor fusion molecule (TNFRII-Fc); human granulocyte-colony stimulating factor (hG-CSF;) human granulocyte-macrophage colony-stimulating factor (hGM-CSF); or human erythropoietin (hEPO).

[0076] The recombinant glycoproteins produced in the present invention preferably comprise human therapeutic glycoproteins, including, but not limited to, granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), erythropoietin (EPO) and TNF-receptors; or soluble TNF-receptor fusion molecules, such as Enbrel (Amgen). For human therapeutics, the recombinant glycoprotein preferably originates from humans or closely related species. Recombinant glycoproteins may also be produced for veterinary indications, in which case the recombinant glycoprotein sequence preferably originates from the same or a closely relates species to that of the intended veterinary subject.

[0077] A particular embodiment of the present invention is generally illustrated in Figure 1.

[0078] In other embodiments, the present invention further comprises a method for producing a recombinant glycoprotein having predominantly a human-like *O*-glycan, said method comprising a) selecting a lower eukaryotic host cell; b) attenuating the activity of one or more endogenous O-glycosylation enzymes in the host cell (where said host cell of step (a) is not already so attenuated); c) transforming the host cell with nucleic acid sequence encoding an O-linked mannose β1,2-N-acetylglucosaminyltransferase 1 (POMGnT1), a UDP-GlcNAc transporter, an alpha-1,2 mannosidase; and the glycoprotein; and d) culturing the cell under conditions suitable for expression of the nucleic acid sequence to produce the recombinant glycoprotein having predominantly a human-like O-glycan. In particular embodiments, the host cell is further transformed with nucleic acid encoding, without limitation, the following enzymes or catalytic domains thereof: β-1,4-galactose transferase ("β1,4GalT"); UDP-galactose transporter (UGT); UDP-Gal epimerase; α-2,6-sialic acid transferase ("α2,6SialT"); α-2,3-sialic acid transferase ("α2,3SialT"); UDP-*N*-acetylglucosamine-2-epimerase/*N*-acetylmannosamine kinase("GNE"); *N*-

acetylneuraminate-9-phosphate synthase("SPS"); sialylate-9-P phosphatase ("SPP"); CMP-sialic acid synthase("CSS"); and CMP-sialic acid transporter ("CST"). In specific embodiments, the host cell is further transformed with nucleic acid encoding the following enzymes or catalytic domains thereof: β1,4GalT, UDP-galactose transporter and UDP-Gal epimerase. In other embodiments, the host cell is further transformed with nucleic acid encoding the following enzymes or catalytic domains thereof: β1,4GalT; UDP-galactose transporter; UDP-Gal epimerase; α2,6SialT; GNE; SPS; SPP; CSS; and CST. In other embodiments, the host cell is further transformed with nucleic acid encoding the following enzymes or catalytic domains thereof: β1,4GalT; UDP-galactose transporter; UDP-Gal epimerase; α2,3SialT; GNE; SPS; SPP; CSS; and CST. In particular embodiments, the host cell is transformed with an extra copy of nucleic acid encoding at least one of the five genes required for sialylation: α2,3SialT (or α2,6SialT); GNE; SPS; SPP; CSS; and/or CST. In incorporating these five genes into a different loci in the genome, Applicants found that providing the extra copy results in a higher percentage of desired sialylated O-glycans. Further embodiments are wherein any of the cells above cells further do not express *KTR1*. The recombinant glycoprotein can be isolated, purified and formulated with pharmaceutically acceptable excipients to produce a therapeutic glycoprotein composition. In particular embodiments, the predominant O-glycan is O-Man-GlcNAc; O-Man-GlcNAc-Gal; or O-Man-GlcNAc-Gal-Sia.

[0079] Specific embodiments of the present invention relate to a method for producing a recombinant glycoprotein having human or mammalian O-like glycosylation in lower eukaryotic host cells which comprise: (1) providing lower eukaryotic host cells which (a) do not express functional: (i) beta-mannosyltransferase enzymes BMT 1, 2, 3 and 4, and (ii) phospho-mannose transferase enzymes Mnn4a, Mnn4b and Pno1; and (b) do express functional (i) POMGnT1; (ii) UDP-GlcNAc transporter; and (iii) α-1,2-mannosidase enzymes; (2) transfecting or transforming the lower eukaryotic host cells with exogenous nucleic acid encoding the glycoprotein of interest, and (3) culturing the host cells under conditions permitting the exogenous nucleic acid to be expressed and the recombinant glycoprotein to be produced. In specific embodiments, the cells of the above method comprise exogenous nucleic acid encoding the enzymes of step (1)(b). In specific embodiments, the cells further express nucleic acid encoding β1,4GalT, UDP-galactose transporter and UDP-Gal epimerase. These additional genes enable the production of terminal GlcNAc-Gal. In other specific embodiments, the cells further express nucleic acid encoding (i) β1,4GalT, (ii) UDP-galactose transporter, (iii) UDP-Gal epimerase, (iv) α2,6SialT, (v) GNE, (vi) SPS, (vii) SPP, (viii) CSS and (ix) CST. These additional genes enable

the production of terminal GlcNAc-Gal-Sia. In other specific embodiments, the cells further express nucleic acid encoding (i) β1,4GalT, (ii) UDP-galactose transporter, (iii) UDP-Gal epimerase, (iv) α2,3SialT, (v) GNE, (vi) SPS, (vii) SPP, (viii) CSS and (ix) CST. Further embodiments are wherein any of the cells above cells further do not express Ktr1.

[0080] The determination of whether a cell expresses 'functional' enzyme is based upon whether the enzyme is carrying out its expected function at levels capable of affecting the desired O-glycosylation step. Where the activity is not detectable, the cell is understood herein to not express functional enzyme. In particular embodiments of step (1)(a) above, the cell does not express the enzymes or enzymatic activity at all. In particular embodiments of step (1)(b) where the cell expresses functional enzyme, the cells express native or expected activity of the enzyme. In specific embodiments of (1)(b), the cells are transfected with nucleic acid encoding the entire enzyme, or nucleic acid encoding the catalytic domain of the enzyme sufficient to enable the activity.

[0081] The methods, in specific embodiments, employ the cell lines, nucleic acid and vectors disclosed herein. Accordingly, in specific embodiments, the cell of step (a) is transfected with nucleic acid encoding the enzymes. In these embodiments, nucleic acid encoding POMGnT1, β1,4GalT, α2,6SialT and α2,3SialT should be operatively linked to a fungal-derived (in particular embodiments, that derived from *S. cerevisiae*, *Pichia pastoris*, or *Kluyveromyces lactis*) promoter and transcription terminator as well as a fungal-derived leader sequence having a signal sequence and an endoplasmic reticulum ("ER")- or Golgi-localizing transmembrane domain. The signal sequence serves to direct the nascent protein to the ER. The transmembrane domain localizes and anchors the protein to an ER or Golgi membrane. Nucleic acid encoding α-1,2-mannosidase should be operatively linked to a fungal-based (in particular embodiments, *Pichia*-derived) promoter and transcription terminator as well as a fungus-derived signal sequence (in specific embodiments wherein the signal sequence is the *Saccharomyces* alphaMAT pre-signal sequence). Nucleic acid encoding UDP-Gal epimerase, UDP-galactose transporter, GNE, SPS, SPP, CSS and CST should be operatively linked to a fungal-based (in particular embodiments, *Pichia*-derived) promoter and transcription terminator. In particular embodiments, the promoter utilized is the *Pichia* AOX1 promoter or the GAP promoter. In other embodiments, the promoter utilized is the *Pichia* TEF promoter or the PMA promoter; *see, e.g.,* Hamilton *et al.*, 2006 *Science* 303:1441-1443. In specific embodiments, the α2,6SialT gene is operatively linked to the *Pichia* TEF promoter. In specific embodiments, the CST gene is operatively linked to the PMA promoter.

Particular embodiments utilize a cell line as described within the present disclosure. The specific cell lines described herein form particular embodiments of the present invention.

[0082] Glycosylation in the resultant cells lines may, where desired, be regulated (e.g., reduced or eliminated) by treatment with one or more chemical inhibitors of fungal O-glycosylation, including but not limited to: 5-[[3,4-bis(phenylmethoxy) phenyl]methylene]-4-oxo-2-thioxo-3- thiazolidineacetic acid; 5-[[3-(l-phenylethoxy)-4-(2-phenylethoxy)] phenyl]methylene]-4-oxo-2-thioxo-3- thiazolidineacetic acid; 3-Hydroxy-4-(2-phenylethoxy)benzaldehyde; 3-(l-Phenylethoxy)-4-(2- phenylethoxy)-benzaldehyde; 5-[[3-(l-Phenyl-2-hydroxy)ethoxy)-4-(2- phenylethoxy)]phenyl] methylene]-4-oxo-2-thioxo-3-thiazolidineacetic acid (*See* WO 2007/061631; incorporated by reference in its entirety). In specific embodiments, the inhibitor is an inhibitor of protein O-mannosyltransferase ("Pmt"); *see* Orchard *et al.*, EP 1313471. In specific embodiments, the inhibitor is PMTi-3, WO 2007/061631. In specific embodiments, the Pmt inhibitors are those described within WO 09/143041 or U.S. application serial no. 61/369157, filed July 30, 2010 [Attorney Docket No. GF2010.7158L01US]. In specific embodiments, Pmt inhibitors are employed in the disclosed processes and provided to the disclosed cells (e.g., in the media) in order to control the overall levels of sialylated O-glycans and/or alter the percentage of different O-glycan species to favor the production of sialylated O-glycans. In particular embodiments, Pmt inhibitor is added in the induction phase as described, for example, in Bobrowicz *et al.*, WO 07/061631. In specific embodiments, Pmt inhibitor is added to the growing yeast with the highest doses added during the methanol induction phase. Administration of Pmt inhibitors can be used to impact the pharmacodynamic properties of the glycoprotein composition by bringing about certain levels of O-sialylation. The particular level desired depends on the specific glycoprotein, as the skilled artisan will appreciate. For instance, with certain therapeutic glycoproteins, too many sialylated O-glycans might interfere with activity. In this situation, lowering the overall numbers of sialylated O-glycans might improve activity. For particular glycoproteins, O-glycans with terminal GlcNAc or galactose reduce protein activity and/or serum half-life. In this latter situation, effectuating a very high percentage of sialylated O-glycans may be required. Alternatively, the maximum number of sialylated O-glycans might be desired for optimal half-life, thus we might add little or no Pmt inhibitor. The skilled artisan can employ these teachings as appropriate in the desired situation.

[0083] The present invention additionally encompasses methods for preparing the above modified lower eukaryotic cells and use of same for preparing recombinant glycoproteins, disclosed nucleic acid, and vectors and host cells comprising same.

[0084] In other embodiments, the present invention comprises recombinant glycoprotein compositions produced from the lower eukaryotic host cells of the invention. These recombinant glycoprotein compositions comprise predominantly a human-like O-glycan, which may preferably be selected from: O-Man-GlcNAc; O-Man-GlcNAc-Gal; or O-Man-GlcNAc-Gal-Sia. In specific embodiments, the present invention produces recombinant glycoprotein compositions comprising greater than 50% (and, preferably, greater than 60%, 70%, 80%, 85%, 90%, 95% and 97%, respectively) human-like O-glycans as described above and less than 15% (and preferably, less than 10%, and less than 5%) O-Man2 glycans, which are glycans possessing two mannose residues attached to the O-glycosylated serine/threonine residue. In particular embodiments, the present invention produces recombinant glycoprotein compositions comprising greater than 30% (and, preferably, greater than 40%, 50%, 60%, 70%, 80% and 83%, respectively) O-Man-GlcNAc-Gal. In particular embodiments, the present invention produces recombinant glycoprotein compositions comprising greater than 30% (and, preferably, greater than 40%, 50% or 55%, respectively) O-Man-GlcNAc-Gal-Sia. In specific embodiments, the present invention relates to the above compositions as described and, in addition, those embodiments that comprise less than 30%, 20%, or 15% (and, preferably, less than 10% and less than 5%) O-Man2 glycans, as well as methods of making the compositions as disclosed herein. O-glycan determination can be performed using methods readily available in the art, including but not limited to methods using a Dionex-HPLC (HPAEC-PAD) or other appropriate methods depending on the protein of interest. Recombinant glycoproteins and compositions comprising the glycoproteins produced using the methods and materials (vectors, host cells, etc.) disclosed herein can be formulated with other pharmaceutically acceptable active agents and/or inactive excipients to form glycoprotein compositions.

[0085] The present invention further relates to cell lines as described herein and as utilized in the methods disclosed herein.

[0086] The present invention can also be used in combination with recent developments allowing the production of therapeutic glycoproteins with human-like N-glycosylation in lower eukaryotic host organisms, yeast and filamentous fungi, such as *Pichia pastoris*; *see, e.g.,* Gerngross, US Patent 7,029,872, the disclosure of which is hereby incorporated by reference. Lower eukaryotes, particularly yeast, can be genetically modified so that they express glycoproteins in which the N-

glycosylation pattern is human-like or humanized. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or supplying exogenous enzymes as described by Gerngross et al., US Patent No. 7,449,308 , the disclosure of which is hereby incorporated herein by reference. If desired, additional genetic engineering of the glycosylation can be performed, such that the glycoprotein can be produced with or without core fucosylation. Use of lower eukaryotic host cells such as yeast are further advantageous in that these cells are able to produce relatively homogenous compositions of glycoprotein, such that the predominant glycoform of the glycoprotein may be present as greater than thirty mole percent of the glycoprotein in the composition. In particular aspects, the predominant glycoform may be present in greater than forty mole percent, fifty mole percent, sixty mole percent, seventy mole percent and, most preferably, greater than eighty mole percent of the glycoprotein present in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or supplying exogenous enzymes as described by Gerngross *et al.*, U.S. Patent No. 7,029,872 and U.S. Patent No. 7,449,308, the disclosures of which are incorporated herein by reference. For example, a host cell can be selected or engineered to be depleted in 1,6-mannosyl transferase activities, which would otherwise add mannose residues onto the *N*-glycan on a glycoprotein.

[0087] In specific embodiments, the host cells described herein includes an *N*-acetylglucosaminyltransferase I (GlcNAc transferase I or GnT I) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase I activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAcMan_5GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872, U.S. Patent No. 7,449,308, and U.S. Published Patent Application No. 2005/0170452, the disclosures of which are all incorporated herein by reference, disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_5GlcNAc_2$ glycoform.

[0088] In a further embodiment, the immediately preceding host cell further includes a mannosidase II catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target mannosidase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant

glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAcMan_3GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Patent No. 7,625,756, the disclosures of which are all incorporated herein by reference, discloses lower eukaryote host cells that express mannosidase II enzymes and are capable of producing glycoproteins having predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

[0089] In a further embodiment, the immediately preceding host cell further includes *N*-acetylglucosaminyltransferase II (GlcNAc transferase II or GnT II) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform. U.S. Patent Nos, 7,029,872 and 7,449,308 and U.S. Published Patent Application No. 2005/0170452, the disclosures of which are all incorporated herein by reference, disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

[0090] In a further embodiment, the immediately preceding host cell further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GalGlcNAc_2Man_3GlcNAc_2$ or $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform, or mixture thereof for example a recombinant glycoprotein composition comprising predominantly a $GalGlcNAc_2Man_3GlcNAc_2$ glycoform or $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. U.S. Patent No, 7,029,872 and U.S. Published Patent Application No. 2006/0040353, the disclosures of which are incorporated herein by reference, discloses lower eukaryote host cells capable of producing a glycoprotein comprising a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells

can be treated *in vitro* with a galactosidase to produce a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform.

[0091] In a further embodiment, the immediately preceding host cell further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly a $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or $NANAGal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. For lower eukaryote host cells such as yeast and filamentous fungi, it is useful that the host cell further include a means for providing CMP-sialic acid for transfer to the *N*-glycan. U.S. Published Patent Application No. 2005/0260729, the disclosure of which is incorporated herein by reference, discloses a method for genetically engineering lower eukaryotes to have a CMP-sialic acid synthesis pathway and U.S. Published Patent Application No. 2006/0286637, the disclosure of which is incorporated herein by reference, discloses a method for genetically engineering lower eukaryotes to produce sialylated glycoproteins. The glycoprotein produced in the above cells can be treated *in vitro* with a neuraminidase to produce a recombinant glycoprotein comprising predominantly a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or $GalGlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof.

[0092] Any one of the preceding host cells can include one or more GlcNAc transferase selected from the group consisting of GnT III, GnT IV, GnT V, GnT VI, and GnT IX to produce glycoproteins having bisected (GnT III) and/or multiantennary (GnT IV, V, VI, and IX) *N*-glycan structures such as disclosed in U.S. Patent No. 7,598,055 and U.S. Published Patent Application No. 2007/0037248, the disclosures of which are all incorporated herein by reference.

[0093] In further embodiments, the host cell that produces glycoproteins that have predominantly $GlcNAcMan_5GlcNAc_2$ *N*-glycans further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly the $GalGlcNAcMan_5GlcNAc_2$ glycoform.

[0094] In a further embodiment, the immediately preceding host cell that produced glycoproteins that have predominantly the $GalGlcNAcMan_5GlcNAc_2$ *N*-glycans further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialytransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $NANAGalGlcNAcMan_5GlcNAc_2$ glycoform.

[0095] In further aspects, any one of the aforementioned host cells, the host cell is further modified to include a fucosyltransferase and a pathway for producing fucose and transporting fucose into the ER or Golgi. Examples of methods for modifying *Pichia pastoris* to render it capable of producing glycoproteins in which one or more of the *N*-glycans thereon are fucosylated are disclosed in Published International Application No. WO 2008112092, the disclosure of which is incorporated herein by reference. In particular aspects of the invention, the *Pichia pastoris* host cell is further modified to include a fucosylation pathway comprising a GDP-mannose-4,6-dehydratase, GDP-keto-deoxy-mannose-epimerase/GDP-keto-deoxy-galactose-reductase, GDP-fucose transporter, and a fucosyltransferase. In particular aspects, the fucosyltransferase is selected from the group consisting of α1,2-fucosyltransferase, α1,3-fucosyltransferase, α1,4-fucosyltransferase, and α1,6-fucosyltransferase.

[0096] Various of the preceding host cells further include one or more sugar transporters such as UDP-GlcNAc transporters (for example, *Kluyveromyces lactis* and *Mus musculus* UDP-GlcNAc transporters), UDP-galactose transporters (for example, *Drosophila melanogaster* UDP-galactose transporter), and CMP-sialic acid transporter (for example, human sialic acid transporter). Because lower eukaryote host cells such as yeast and filamentous fungi lack the above transporters, it is preferable that lower eukaryote host cells such as yeast and filamentous fungi be genetically engineered to include the above transporters.

[0097] In a specific embodiment, a host cell can be selected or engineered to be depleted in 1,6-mannosyl transferase activities, which would otherwise add mannose residues onto the N-glycan on a glycoprotein. For example, a glycoengineered *P. pastoris* yeast strain may be generated in which the typical yeast-type N-glycosylation is modified to instead produce fully sialylated human-like N-glycans. First, deletion of the yeast gene *OCH1* eliminates the enzyme activity responsible for 'outer chain' glycosylation (Choi et al, Proc Natl Acad Sci US; 100:5022-7 (2003)). A mannosidase I (MNSI) gene and GlcNAc transferase I (GnTI) gene may then be engineered into the strain and properly localized to the secretory pathway to

efficiently generate mammalian hybrid-type N-glycans (Choi et al, 2003). In a further step, a mannosidase II (MNSII) gene and GlcNAc transferase II (GnTII) genes may then be engineered into the strain and properly localized to the secretory pathway to efficiently generate mammalian complex-type N-glycans (Hamilton et al, Science; 301:1244-6. (2003)). Furthermore, by engineering into this strain enzymes to generate a pool of UDP-galactose, appropriate Golgi membrane transporters and a gene encoding galactosyltransferase (GalT), a yeast strain can be generated that is capable of transferring a complex-type human N-glycan with terminal ß-1,4-galactose (Bobrowicz et al., Glycobiology; 14:757-66 (2004)). Finally, by introducing the enzymes required for *in vivo* synthesis and transfer of sialic acid (Hamilton et al, Science, 313(5792):1441-3 (2006)), sialylated glycans can be obtained. The disclosure of each of the references cited above is hereby incorporated herein by reference. Where no N-glycosylation is present on the recombinant protein, it is not necessary to engineer the lower eukaryotic host cells' glycosylation machinery in this manner. Moreover, one of skill in the art can add and/or eliminate one or more N-glycosylation sites from the sequence encoding the recombinant glycoprotein. In such embodiments, the glycoprotein compositions of the present invention comprise a predominantly mammalian-like N-glycan or human-like N-glycan. The lower eukaryotic host cells of the present invention may optionally be engineered to produce glycoproteins having a predominantly human-like N-glycan. In preferred embodiments, the host cells produce glycoproteins having a predominant N-glycan selected from: Man5GlcNAc2; GlcNAcMan5GlcNAc2; GalGlcNAcMan5GlcNAc2; SiaGalGlcNAcMan5GlcNAc2; Man3GlcNAc2; GlcNAcMan3GlcNAc2; GalGlcNAcMan3GlcNAc2; SiaGalGlcNAcMan3GlcNAc2; GlcNAc2Man3GlcNAc2; GalGlcNAc2Man3GlcNAc2; Gal2GlcNAc2Man3GlcNAc2; SiaGal2GlcNAc2Man3GlcNAc2; or Sia2Gal2GlcNAc2Man3GlcNAc2. In a preferred embodiment, the predominant N-glycan comprises at least 30 mole percent, preferably at least 40 mole percent and more preferably at least 50 mole percent of the N-glycans stated above present on the glycoprotein in the composition.

[0098] Therefore, the methods disclose herein can use any host cell that has been genetically modified to produce glycoproteins wherein the predominant *N*-glycan is selected from the group consisting of complex *N*-glycans, hybrid *N*-glycans, and high mannose *N*-glycans wherein complex *N*-glycans may be selected from the group consisting of $GlcNAc_{(2-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(2-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}$ $GlcNAc_{(2-4)}Man_3GlcNAc_2$; hybrid *N*-glycans maybe selected from the group consisting of $GlcNAcMan_3GlcNAc_2$; $GalGlcNAcMan_3GlcNAc_2$; $NANAGalGlcNAcMan_3GlcNAc_2$ $GlcNAcMan_5GlcNAc_2$,

$GalGlcNAcMan_5GlcNAc_2$, and $NANAGalGlcNAcMan_5GlcNAc_2$; and high Mannose *N*-glycans maybe selected from the group consisting of $Man_5GlcNAc_2$, $Man_6GlcNAc_2$, $Man_7GlcNAc_2$, $Man_8GlcNAc_2$, and $Man_9GlcNAc_2$. Further included are glycoproteins having *N*-glycans consisting of the *N*-glycan structure $Man_3GlcNAc_2$, for example, as shown in U.S. Published Application No. 20050170452.

[0099] The present invention, as that of the previously described N-glycosylation in lower eukaryotic host organisms, can be used for commercial scale production of recombinant glycoprotein compositions. In addition, the methods and materials of the present invention can be used for veterinary application as well.

[0100] As described herein, there are many attributes of the methods and materials of the present invention which provide unobvious advantages for the disclosed recombinantly produced glycoproteins, host cells, nucleic acids, vectors and expression processes over prior known expression processes.

[0101] Most notably, the present invention can be used to improve the current state of the art of recombinantly produced glycoprotein therapeutic products. A glycosylated protein produced according to the present invention can be expected to be less immunogenic to humans than glycoproteins produced from wild type yeast or other lower eukaryotic host cells. The glycosylation of recombinant glycoproteins produced according to the present invention can be expected to be much more uniform than glycosylation of glycoprotein compositions produced from mammalian cells. Furthermore, as stated above, humanizing the O-glycans can enhance the bioactivity of therapeutic proteins by improving pharmacokinetic properties, hence, facilitating better control over *in vivo* drug activity.

[0102] Unless otherwise defined herein, scientific and technical terms and phrases used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include the plural and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al. Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel

et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Taylor and Drickamer, *Introduction to Glycobiology*, Oxford Univ. Press (2003); *Worthington Enzyme Manual*, Worthington Biochemical Corp., Freehold, NJ; *Handbook of Biochemistry: Section A Proteins*, Vol I, CRC Press (1976); *Handbook of Biochemistry: Section A Proteins*, Vol II, CRC Press (1976); *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999).

[0103] All publications, patents and other references mentioned herein are hereby incorporated by reference in their entireties.

[0104] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0105] As used herein, the terms "O-glycan" and "O-glycoform" are used interchangeably and refer to an O-linked oligosaccharide, e.g., a glycan that is attached to a peptide chain via the hydroxyl group of either a serine or threonine residue. In fungal cells, native O-glycosylation occurs through attachment of a first mannosyl residue transferred from a dolichol monophosphate mannose (Dol-P-Man) to the protein in the endoplasmic reticulum, and additional mannosyl residues may be attached via transfer from GDP-Man in the Golgi apparatus. Higher eukaryotic cells, such as human or mammalian cells, undergo O-glycosylation through covalent attachment of N-acetyl-galactosamine (GalNAc) to the serine or threonine residue, to form mucin-type glycans. This is in addition to α-dystroglycan-type glycans described above.

[0106] As used herein, the terms "*N*-glycan" and "*N* glycoform" are used interchangeably and refer to an *N*-linked oligosaccharide, e.g., one that is attached by an asparagine-*N*-acetylglucosamine linkage to an asparagine residue of a polypeptide. *N*-linked glycoproteins contain an *N*-acetylglucosamine residue linked to the amide nitrogen of an asparagine residue in the protein. The predominant sugars found on glycoproteins are glucose, galactose, mannose, fucose, *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc) and sialic acid (e.g., *N*-acetyl-neuraminic acid (NANA)). The processing of the sugar groups occurs cotranslationally in the lumen of the ER and continues in the Golgi apparatus for *N*-linked glycoproteins.

[0107] *N*-glycans have a common pentasaccharide core of $Man_3GlcNAc_2$ ("Man" refers to mannose; "Glc" refers to glucose; and "NAc" refers to *N*-acetyl; GlcNAc refers to *N*-acetylglucosamine). *N*-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (e.g., GlcNAc, galactose, fucose and sialic acid) that are added to the $Man_3GlcNAc_2$ ("Man3") core structure which is

also referred to as the "trimannose core", the "pentasaccharide core" or the "paucimannose core". *N*-glycans are classified according to their branched constituents (e.g., high mannose, complex or hybrid). A "high mannose" type *N*-glycan has five or more mannose residues. A "complex" type *N*-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of a "trimannose" core. Complex *N*-glycans may also have galactose ("Gal") or *N*-acetylgalactosamine ("GalNAc") residues that are optionally modified with sialic acid or derivatives (e.g., "NANA" or "NeuAc", where "Neu" refers to neuraminic acid and "Ac" refers to acetyl). Complex *N*-glycans may also have intrachain substitutions comprising "bisecting" GlcNAc and core fucose ("Fuc"). Complex *N*-glycans may also have multiple antennae on the "trimannose core," often referred to as "multiple antennary glycans." A "hybrid" *N*-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core. The various *N*-glycans are also referred to as "glycoforms."

[0108] As used herein, the term "human-like O-glycosylation" will be understood to mean that fungal-specific phosphorylated and beta-linked mannose structures are reduced or eliminated, resulting in reduction or elimination of charge and beta-mannose structures, or that the predominant O-glycan species present on a glycoprotein or in a composition of glycoprotein comprises a glycan which capped with a terminal residue selected from GlcNAc; Gal or Sia. In this manner, the recombinant glycoprotein bearing predominantly human-like O-glycosylation may be recognized by a human or mammalian cell as if it were a natively produced glycoprotein, resulting in improved therapeutic properties of the recombinant glycoprotein.

[0109] Abbreviations used herein are of common usage in the art, see, e.g., abbreviations of sugars, above. Other common abbreviations include "PNGase", or "glycanase" or "glucosidase" which all refer to peptide *N*-glycosidase F (EC 3.2.2.18).

[0110] An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA, various derivatives thereof (*e.g., c*DNA) or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases and genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is

operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[0111] However, "isolated" does not necessarily require that the nucleic acid or polynucleotide so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed "isolated" herein if a heterologous sequence is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. In this context, a heterologous sequence is a sequence that is not naturally adjacent to the endogenous nucleic acid sequence, whether or not the heterologous sequence is itself endogenous (originating from the same host cell or progeny thereof) or exogenous (originating from a different host cell or progeny thereof). By way of example, a promoter sequence can be substituted (*e.g.*, by homologous recombination) for the native promoter of a gene in the genome of a host cell, such that this gene has an altered expression pattern. This gene would now become "isolated" because it is separated from at least some of the sequences that naturally flank it.

[0112] A nucleic acid is also considered "isolated" if it contains any modifications that do not naturally occur to the corresponding nucleic acid in a genome. For instance, an endogenous coding sequence is considered "isolated" if it contains an insertion, deletion or a point mutation introduced artificially, *e.g.*, by human intervention. An "isolated nucleic acid" also includes a nucleic acid integrated into a host cell chromosome at a heterologous site and a nucleic acid construct present as an episome. Moreover, an "isolated nucleic acid" can be substantially free of other cellular material, or substantially free of culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0113] The term "vector" as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*,

vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain preferred vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

[0114] As used herein, the term "sequence of interest" or "gene of interest" refers to a nucleic acid sequence, typically encoding a protein, that is not normally produced in the host cell. The methods disclosed herein allow efficient expression of one or more sequences of interest or genes of interest stably integrated into a host cell genome. Non-limiting examples of sequences of interest include sequences encoding one or more polypeptides having an enzymatic activity, *e.g.*, an enzyme which affects *N*-glycan synthesis in a host such as mannosyltransferases, *N*-acetyl-glucosaminyltransferases, UDP-*N*-acetylglucosamine transporters, galactosyltransferases, UDP-*N*-acetyl-galactosyltransferase, sialyltransferases and fucosyltransferases. The protein so produced is referred to as a "recombinant" protein.

[0115] The term "derived" means the sequence or protein is the same as that from which it was derived ("the reference sequence or protein"), or it has minor differences which do not negatively impact the function thereof as compared to the reference sequence or protein.

[0116] The term "marker sequence" or "marker gene" refers to a nucleic acid sequence capable of expressing an activity that allows either positive or negative selection for the presence or absence of the sequence within a host cell. For example, the *P. pastoris URA5* gene is a marker gene because its presence can be selected for by the ability of cells containing the gene to grow in the absence of uracil. Its presence can also be selected against by the inability of cells containing the gene to grow in the presence of 5-FOA. Marker sequences or genes do not necessarily need to display both positive and negative selectability. Non-limiting examples of marker sequences or genes from *P. pastoris* include *ADE1*, *ARG4*, *HIS4* and *URA3*. For antibiotic resistance marker genes, kanamycin, neomycin, geneticin (or G418), paromomycin and hygromycin resistance genes are commonly used to allow for growth in the presence of these antibiotics.

[0117] "Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

[0118] The term "expression control sequence" or "regulatory sequences" are used interchangeably and as used herein refer to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

[0119] The term "recombinant host cell" ("expression host cell", "expression host system", "expression system" or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A recombinant host cell may be an isolated cell or cell line grown in culture or may be a cell which resides in a living tissue or organism.

[0120] The term "eukaryotic" refers to a nucleated cell or organism, and includes insect cells, plant cells, mammalian cells, animal cells and lower eukaryotic cells.

[0121] The term "lower eukaryotic cells" includes yeast, fungi, collar-flagellates, microsporidia, alveolates (e.g., dinoflagellates), stramenopiles (e.g, brown algae, protozoa), rhodophyta (e.g., red algae), plants (e.g., green algae, plant cells, moss) and other protists. Yeast and filamentous fungi include, but are not limited to: *Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia minuta (Ogataea minuta, Pichia lindneri), Pichia opuntiae, Pichia thermotolerans, 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, Physcomitrella patens* and *Neurospora crassa.*

[0122] The term "peptide" as used herein refers to a short polypeptide, *e.g.*, one that is typically less than about 50 amino acids long and more typically less than about 30 amino acids long. The term as used herein encompasses analogs and mimetics that mimic structural and thus biological function.

[0123] The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins, and fragments, mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different domains each of which has one or more distinct activities.

[0124] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) exists in a purity not found in nature, where purity can be adjudged with respect to the presence of other cellular material (*e.g.*, is free of other proteins from the same species), (3) is expressed by a cell from a different species, or (4) does not occur in nature (*e.g.*, it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art. As thus defined, "isolated" does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

[0125] As used herein, the term "predominantly" or variations such as "the predominant" or "which is predominant" will, depending on the context used, be understood to mean the glycan species that has the highest mole percent (%) of total O-glycans or *N*-glycans after the glycoprotein has been treated with enzymes and released glycans analyzed by mass spectroscopy, for example, MALDI-TOF MS. In other words, the phrase "predominantly" is defined as an individual entity, such that a specific "predominant" glycoform is present in greater mole percent than any other individual entity. For example, if a composition consists of species A in 40 mole percent, species B in 35 mole percent and species C in 25 mole percent, the composition comprises predominantly species A.

[0126] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice of the present invention and will be apparent to those of skill in the art. All publications and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting in any manner.

[0127] The following examples illustrate the operation of the invention with respect to certain preferred embodiments, and are in no way limiting of the scope of the invention described in the specification and claims. As will be readily apparent to the skilled artisan, having read the specification, numerous modifications to the methods and materials described herein are possible without deviating from the scope of the invention described and claimed herein.

EXAMPLES

[0128] In the following examples, recombinant glycoproteins are expressed in host cells of the species *Pichia pastoris*. The host cells have been genetically engineered to produce glycoproteins with human-like O-glycosylation, as well as human-like N-glycosylation. These examples demonstrate the invention with respect to specific preferred embodiments of the invention, and are not limiting in any manner.

[0129] As will be apparent to the skilled practitioner, having read the disclosure and examples herein, numerous modifications, substitutions, adaptations and improvements within the skill of the art can be made without deviating from the scope of the specification and the appended claims. Accordingly, the examples are non-limiting and, with appropriate modifications, the methods and materials described herein may be used for the practice of the invention in other lower eukaryotic organisms.

EXAMPLE 1

Cloning of ER and Golgi Targeting Sequences

[0130] Most of the glycosyltransferases in the ER and Golgi of mammals and yeast are type II membrane proteins (Gleeson, P.A. Targeting of proteins to the Golgi apparatus. Histochem. Cell. Biol. 109, 517-532 (1998)) that consist of a short amino-terminal cytoplasmic tail followed by a transmembrane domain [TMD], a short stem

region and a large carboxy-terminal catalytic domain in the lumen of the ER or the Golgi. Another group of glycosyltransferases represented by the Pmts (protein O-mannosyltransferases; Strahl-Bolsinger et al., Protein O-mannosylation, Biochim Biophys Acta 1426:297 (1999); Tanner et al., US Patent 5,714,377) contain multiple membrane-spanning domains. Because we wanted to test a large variety of targeting sequences, we amplified DNA fragments encoding membrane targeting regions of varying length of ER and Golgi residing membrane proteins of *S. cerevisiae*, *P. pastoris* and *K. lactis*. For type II membrane proteins, the shortest fragments (designated '-s') contained only the cytoplasmic tail and the TMD. The medium length fragments (designated '-m') also contained parts of, or the entire stem region. The longest fragments (designated '-l') contained the entire stem region and also parts of the catalytic domain. A detailed description of targeting domains (TD) from type II membrane proteins can be found in Gerngross et al., US 7,449,308 (e.g., TABLE 5) and Choi *et al.*, 2003 *PNAS* 100:5022. For other integral membrane proteins, we took either the shortest membrane spanning region (designated '-s') or a region encompassing multiple membrane spanning domains (designated '-m').

[0131] In the present example, the ER targeting domains utilized were fragments of the *S. cerevisiae* alpha-glucosidase I encoded by *GLS1*, the *S. cerevisiae* alpha 1,2-mannosidase encoded by *MNS1*, the *S. cerevisiae* nucleotide exchange factor *SEC12* which cycles between the ER and Golgi, fragments from the *S. cerevisiae* and *P. pastoris PMT1, 2, 3, 4, 5* and *6* that encode integral ER membrane spanning domains, and domains from the SNARE proteins encoded by *BOS1, BET1*, and *SEC22* from *S. cerevisiae* and *P. pastoris*.

[0132] As targeting domains for the early or cis-Golgi we chose amino-terminal fragments of the *P. pastoris* Och1p, and of the proteins that make up the *S. cerevisiae* mannan polymerases M-Pol I and M-Pol II, namely Mnn9p, Van1p, Anp1p, Hoc1p, Mnn10p and Mnn11p. Targeting signals for the medial Golgi were derived from *S. cerevisiae* Kre2p, Ktr1p, Mnn2p, Mnn5p, and Yur1p, from *K. lactis* Gnt1p and from *P. pastoris* proteins with homology to Ktr1p, Ktr3p and Kre2p (unpublished results).

[0133] As targeting domains for the late Golgi we included amino-terminal regions of *S. cerevisiae* Mnn1p and Mnn6p. The complete list of targeting domains is given in Table 1, and DNA and protein sequences provided (SEQ IDs).

Table 1. ER/Golgi leaders fused to POMGnT1

Leader#	Gene	Size [Nuc]		Leader#	Gene	Size [Nuc]
1	ScGLS1-s	102		33	ScPMT1-m	330
2	ScMNS1-s	90		34	ScPMT2-m	360
3	ScMNS1-m	246		35	ScPMT3-m	330
4	ScSEC12-m	309		36	ScPMT4-m	405
5	PpSEC12-s	87		37	ScPMT5-m	255
6	PpOCH1-s	150		38	ScPMT6-m	585
7	ScMNN9-s	120		39	PpPMT1-s	255
8	ScVAN1-m	294		40	PpPMT2-s	594
9	ScANP1-s	180		41	PpPMT4-s	156
10	ScHOC1-s	102		42	PpPMT5-s	120
11	ScMNN10-s	219		43	PpPMT6-s	207
12	ScMNN11-s	156		44	PpPMT1-m	390
13	ScKRE2-s	174		45	PpPMT2-m	630
14	ScKRE2-l	306		46	PpPMT4-m	195
15	PpKTR1-s	216		47	PpPMT5-m	240
16	PpKTR3-s	93		48	PpPMT6-m	330
17	PpKRE2-s	93		49	ScBOS1-s	90
18	ScKTR1-s	117		50	ScBET1-s	99
19	ScKTR2-s	120		51	ScSEC22-s	105
20	KlGNT1-s	93		52	PpBOS1-s	87
21	ScMNN2-s	108		53	PpBET1-s	102
22	ScMNN2-m	291		54	PpSEC22-s	105
23	ScMNN5-s	105				
24	ScYUR1-s	108				
25	ScMNN1-s	126				
26	ScMNN6-s	90				
27	ScPMT1-s	213				
28	ScPMT2-s	237				
29	ScPMT3-s	213				
30	ScPMT4-s	357				
31	ScPMT5-s	195				
32	ScPMT6-s	543				

Sequences of ER/Golgi transmembrane domains (leaders)

DNA followed by amino acid sequences

ScGLS1-s

ATGCTTATTTCAAAATCTAAGATGTTTAAAACATTTTGGATACTAACCAGC
ATAGTTCTCCTGGCATCTGCCACCGTTGATATTAGTAAACTACAAGAATTC
GGGCGCGCC [SEQ ID NO: 1]

MLISKSKMFKTFWILTSIVLLASATVDISKLQEFGRA [SEQ ID NO: 2]

ScMNS1-s
ATGAAGAACTCTGTCGGTATTTCAATTGCAACCATTGTTGCTATCATAGCA
GCTATATACTATGTGCCATGGTACGAACACTTTGAGAGAGGGCGCGCC
[SEQ ID NO: 3]

MKNSVGISIATIVAIIAAIYYVPWYEHFERGRA [SEQ ID NO: 4]

ScMNS1-m
ATGAAGAACTCTGTCGGTATTTCAATTGCAACCATTGTTGCTATCATAGCA
GCTATATACTATGTGCCATGGTACGAACACTTTGAGAGAAAGTCACCGGG
GGCCGGAGAAATGAGAGATCGGATTGAAAGCATGTTCTTGGAATCGTGGA
GAGACTATTCCAAGCATGGCTGGGGATACGATGTGTATGGACCTATTGAG
CACACTTCCCATAATATGCCTCGTGGCAACCAGCCGTTAGGCTGGGGGCG
CGCC [SEQ ID NO: 5]

MKNSVGISIATIVAIIAAIYYVPWYEHFERKSPGAGEMRDRIESMFLESWRDYS
KHGWGYDVYGPIEHTSHNMPRGNQPLGWGRA [SEQ ID NO: 6]

ScSEC12-m
ATGAACACTATCCACATAATAAAATTACCGCTTAACTACGCCAACTACAC
CTCAATGAAACAAAAAAATCTCTAAATTTTTCACCAACTTCATCCTTATTGT
GCTGCTTTCTTACATTTTACAGTTCTCCTATAAGCACAATTTGCATTCCATG
CTTTTCAATTACGCGAAGGACAATTTTCTAACGAAAAGAGACACCATCTCT
TCGCCCTACGTAGTTGATGAAGACTTACATCAAACAACTTTGTTTGGCAAC
CACGGTACAAAAACATCTGTACCTAGCGTAGATTCCATAAAAGTGCATGG
CGTGGGGCGCGCC [SEQ ID NO: 7]

MNTIHIIKLPLNYANYTSMKQKISKFFTNFILIVLLSYILQFSYKHNLHSMLFNY
AKDNFLTKRDTISSPYVVDEDLHQTTLFGNHGTKTSVPSVDSIKVHGVGRA
[SEQ ID NO: 8]

PpSEC12-s
ATGCCCAGAAAAATATTTAACTACTTCATTTTGACTGTATTCATGGCAATT
CTTGCTATTGTTTTACAATGGTCTATAGAGAATGGACATGGGCGCGCC
[SEQ ID NO: 9]

MPRKIFNYFILTVFMAILAIVLQWSIENGHGRA [SEQ ID NO: 10]

PpOCH1-s
ATGGCGAAGGCAGATGGCAGTTTGCTCTACTATAATCCTCACAATCCACC
CAGAAGGTATTACTTCTACATGGCTATATTCGCCGTTTCTGTCATTTGCGTT
TTGTACGGACCCTCACAACAATTATCATCTCCAAAAATAGACTATGATGG
GCGCGCC [SEQ ID NO: 11]

MAKADGSLLYYNPHNPPRRYYFYMAIFAVSVICVLYGPSQQLSSPKIDYDGR
A [SEQ ID NO: 12]

ScMNN9-s
ATGTCACTTTCTCTTGTATCGTACCGCCTAAGAAAGAACCCGTGGGTTAAC
ATTTTTCTACCTGTTTTGGCCATATTTCTAATATATATAATTTTTTTCCAGA
GAGATCAATCTCTGTTGGGGCGCGCC [SEQ ID NO: 13]

MSLSLVSYRLRKNPWVNIFLPVLAIFLIYIIFFQRDQSLLGRA [SEQ ID NO: 14]

ScVAN1-m
ATGGGCATGTTTTTTAATTTAAGGTCAAATATAAAGAAGAAAGCCATGGA
CAATGGACTAAGCCTGCCCATTTCAAGGAACGGTAGCTCGAACAACATCA
AGGACAAACGCTCAGAGCATAACTCCAACTCATTAAAGGGCAAATACAG
GTACCAGCCGCGCTCCACACCGTCTAAATTCCAGCTTACGGTGAGTATCAC
ATCTCTTATTATTATCGCCGTTCTGTCGTTATATCTCTTTATATCATTTCTCT
CCGGAATGGGCATTGGTGTATCCACGCAAAATGGTAGGTCGGGGCGCGCC
[SEQ ID NO: 15]

MGMFFNLRSNIKKKAMDNGLSLPISRNGSSNNIKDKRSEHNSNSLKGKYRYQ
PRSTPSKFQLTVSITSLIIIAVLSLYLFISFLSGMGIGVSTQNGRSGRA [SEQ ID NO: 16]

ScANP1-s
ATGAAGTATAATAACAGAAAACTCTCGTTCAACCCTACCACAGTAAGTAT
CGCTGGAACGTTGCTTACGGTGTTCTTTCTCACAAGACTCGTGCTTTCGTT
CTTCTCGATATCGCTATTCCAGCTGGTAACTTTCCAAGGAATCTTCAAGCC
CTATGTTCCAGATTTTAAAAATACTCCCGGGCGCGCC [SEQ ID NO: 17]

MKYNNRKLSFNPTTVSIAGTLLTVFFLTRLVLSFFSISLFQLVTFQGIFKPYVPD
FKNTPGRA [SEQ ID NO: 18]

ScHOC-s
ATGGCCAAAACAACAAAAAGAGCCTCCAGTTTCAGGAGGTTGATGATATT
CGCCATAATAGCCCTCATCTCATTAGCATTTGGAGTTAGATACCTATTTCA
CGGGCGCGCC [SEQ ID NO: 19]

MAKTTKRASSFRRLMIFAIIALISLAFGVRYLFHGRA [SEQ ID NO: 20]

ScMNN10-s
ATGTCTAGTGTACCTTATAATTCCCAACTTCCTATATCCAACCATCTAGAG
TACGATGAAGATGAAAAGAAGAGCAGAGGCTCAAAACTAGGCCTGAAAT
ATAAAATGATATACTGGAGGAAAACTTTATGCAGTTCGCTAGCGAGATGG
AGAAAGCTAATACTATTAATATCTTTAGCTTTGTTTTTATTCATATGGATA
AGCGATTCCACCATAAGCGGGCGCGCC [SEQ ID NO: 21]

MSSVPYNSQLPISNHLEYDEDEKKSRGSKLGLKYKMIYWRKTLCSSLARWRK
LILLISLALFLFIWISDSTISGRA [SEQ ID NO: 22]

ScMNN11-s
ATGGCAATCAAACCAAGAACGAAGGGCAAAACGTACTCCTCAAGATCGG
TGGGTTCGCAGTGGTTCAACAGGCTTGGTTTCAAGCAGAACAAGTACGGA
ACTTGTAAATTTTTGTCGATAATAACGGCCTTTGTTTTTATCCTCTATTTCT
TCTCCGGGCGCGCC [SEQ ID NO: 23]

MAIKPRTKGKTYSSRSVGSQWFNRLGFKQNKYGTCKFLSIITAFVFILYFFSGR
A [SEQ ID NO: 24]

ScKRE2-s
ATGGCCCTCTTTCTCAGTAAGAGACTGTTGAGATTTACCGTCATTGCAGGT
GCGGTTATTGTTCTCCTCCTAACATTGAATTCCAACAGTAGAACTCAGCAA
TATATTCCGAGTTCCATCTCCGCTGCATTTGATTTTACCTCAGGATCTATAT
CCCCTGAACAACAAGTCATCGGGCGCGCC [SEQ ID NO: 25]

MALFLSKRLLRFTVIAGAVIVLLLTLNSNSRTQQYIPSSISAAFDFTSGSISPEQQ
VIGRA [SEQ ID NO: 26]

ScKRE2-l
ATGGCCCTCTTTCTCAGTAAGAGACTGTTGAGATTTACCGTCATTGCAGGT
GCGGTTATTGTTCTCCTCCTAACATTGAATTCCAACAGTAGAACTCAGCAA
TATATTCCGAGTTCCATCTCCGCTGCATTTGATTTTACCTCAGGATCTATAT
CCCCTGAACAACAAGTCATCTCTGAGGAAAATGATGCTAAAAAAATTAGAG
CAAAGTGCTCTGAATTCAGAGGCAAGCGAAGACTCCGAAGCCATGGATGA
AGAATCCAAGGCTCTGAAAGCTGCCGCTGAAAAGGCAGATGCCCCGATCG
ACGGGCGCGCC [SEQ ID NO: 27]

MALFLSKRLLRFTVIAGAVIVLLLTLNSNSRTQQYIPSSISAAFDFTSGSISPEQQ
VISEENDAKKLEQSALNSEASEDSEAMDEESKALKAAAEKADAPIDGRA
[SEQ ID NO: 28]

PpKTR1-s
ATGGAATTAGTGCGCCTGGCCAATCTTGTCAACGTCAACCACCCTTTCGAG
CAAAGCAATATATATCGCGTTCCACTTTTCTTCCTTCTCTCAACTACCAGA
CCAGACAGGACAACGGTACAAATGGCAGGTGCAACTAGGATCAATTCAC
GAGTAGTTCGGTTTGCTATTTTCGCATCAATCCTGGTACTGTTAGGATTCA
TCCTATCAAGAGGGGGGCGCGCC [SEQ ID NO: 29]

MELVRLANLVNVNHPFEQSNIYRVPLFFLLSTTRPDRTTVQMAGATRINSRVV
RFAIFASILVLLGFILSRGGRA [SEQ ID NO: 30]

PpKTR3-s
ATGATGCGAGCAAGATTAAGCCTTGAACGAGTTAACTTGAGCTTTATTAC
GTCCGTATTTTTGGCTTCAGTTGCAGTTCTTTTCATCTCTTTGGGGCGCGCC
[SEQ ID NO: 31]

MMRARLSLERVNLSFITSVFLASVAVLFISLGRA [SEQ ID NO: 32]

PpKRE2-s
ATGGTACACATAGGGTTCAGAAGCTTGAAAGCGGTGTTCATTTTGGCCCTT
TCGTCATTGATTCTGTACGGGATTGTCACGACCTTTGACGGGGGGCGCGCC
[SEQ ID NO: 33]

MVHIGFRSLKAVFILALSSLILYGIVTTFDGGRA [SEQ ID NO: 34]

ScKTR1-s
ATGGCGAAGATTATGATCCCAGCTAGCAAGCAGCCTGTTTACAAAAAATT
AGGACTTCTTCTGGTCGCCGTGTTTACTGTGTATGTGTTCTTTCATGGAGCT
CAGTATGCGAGAGGCGGGCGCGCC [SEQ ID NO: 35]

MAKIMIPASKQPVYKKLGLLLVAVFTVYVFFHGAQYARGGRA
[SEQ ID NO: 36]

ScKTR2-s
ATGCAAATCTGCAAGGTATTTCTTACACAGGTTAAAAAACTACTTTTTGTT
AGTCTTCTATTTTGCTTGATAGCTCAAACATGTTGGCTTGCACTTGTACCAT
ATCAGAGACAGCTGAGCGGGCGCGCC [SEQ ID NO: 37]

MQICKVFLTQVKKLLFVSLLFCLIAQTCWLALVPYQRQLSGRA
[SEQ ID NO: 38]

KIGNT1-s
ATGGCTTTTGGATCTAGAAGGAAAATCAAGGCCATTTTGGTCGCTGCTTCT
GCTATGGTCTTTATTTCTCTACTTGGAACGTTTGGATCCGACGGGCGCGCC
[SEQ ID NO: 39]

MAFGSRRKIKAILVAASAMVFISLLGTFGSDGRA [SEQ ID NO: 40]

ScMNN2-s
ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCTGACGTTCATAGTT
TTGATATTGTGCGGGCTGTTCGTCATTACAAACAAATACATGGATGAGAA
CACGTCGGGGCGCGCC [SEQ ID NO: 41]

MLLTKRFSKLFKLTFIVLILCGLFVITNKYMDENTSGRA [SEQ ID NO: 42]

ScMNN2-m
ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCTGACGTTCATAGTT
TTGATATTGTGCGGGCTGTTCGTCATTACAAACAAATACATGGATGAGAA
CACGTCGGTCAAGGAGTACAAGGAGTACTTAGACAGATATGTCCAGAGTT
ACTCCAATAAGTATTCATCTTCCTCAGACGCCGCCAGCGCTGACGATTCAA
CCCCATTGAGGGACAATGATGAGGCAGGCAATGAAAAGTTGAAAAGCTTC

TACAACAACGTTTTCAACTTTCTAATGGTTGATTCGCCCGGGCGCGCC
[SEQ ID NO: 43]

MLLTKRFSKLFKLTFIVLILCGLFVITNKYMDENTSVKEYKEYLDRYVQSYSN
KYSSSSDAASADDSTPLRDNDEAGNEKLKSFYNNVFNFLMVDSPGRA
[SEQ ID NO: 44]

ScMNN5-s
ATGCTGATTAGGTTAAAGAAGAGAAAAATCCTGCAGGTCATCGTGAGCGC
AGTAGTGCTAATTTTATTTTTTTGTTCTGTGCATAATGATGTGTCTTCTAGT
TGGGGGCGCGCC [SEQ ID NO: 45]

MLIRLKKRKILQVIVSAVVLILFFCSVHNDVSSSWGRA [SEQ ID NO: 46]

ScYUR1-s
ATGGCAAAAGGAGGCTCGCTATACATCGTTGGCATATTCTTACCAATATG
GACCTTTATGATCTATATTTTTGGCAAAGAGTTATTCCTCATACGAAAATA
CCAAAAGGGGCGCGCC [SEQ ID NO: 47]

MAKGGSLYIVGIFLPIWTFMIYIFGKELFLIRKYQKGRA [SEQ ID NO: 48]

ScMNN1-s
ATGTTGGCACTCCGGAGATTTATATTAAACCAAAGGTCTTTGAGATCGTGT
ACCATACCGATTCTAGTCGGAGCCTTGATCATTATTCTCGTGCTATTCCAA
CTAGTTACCCACCGAAATGATGCGGGGCGCGCC [SEQ ID NO: 49]

MLALRRFILNQRSLRSCTIPILVGALIIILVLFQLVTHRNDAGRA
[SEQ ID NO: 50]

ScMNN6-s
ATGCACGTACTGCTGAGCAAAAAAATAGCACGCTTTCTGTTGATTTCGTTT
GTTTTCGTGCTTGCGCTAATGGTGACAATAAATCATCCAGGGCGCGCC
[SEQ ID NO: 51]

MHVLLSKKIARFLLISFVFVLALMVTINHPGR [SEQ ID NO: 52]

ScPMT1-s
ATGTCGGAAGAGAAAACGTACAAACGTGTAGAGCAGGATGATCCCGTGC
CCGAACTGGATATCAAGCAGGGCCCCGTAAGACCCTTTATTGTTACCGAT
CCGAGTGCCGAATTGGCCTCGTTACGAACCATGGTCACTCTTAAAGAGAA
GCTGTTAGTGGCCTGTCTTGCTGTCTTTACAGCGGTCATTAGATTGCATGG
CTTGGCATGGCCTGGGCGCGCC [SEQ ID NO: 53]

MSEEKTYKRVEQDDPVPELDIKQGPVRPFIVTDPSAELASLRTMVTLKEKLLV
ACLAVFTAVIRLHGLAWPGRA [SEQ ID NO: 54]

ScPMT2-s
ATGTCCTCGTCTTCGTCTACCGGGTACAGCAAAAACAATGCCGCCCACATT
AAGCAAGAGAATACACTGAGACAAAGAGAATCGTCTTCCATCAGCGTCAG
TGAGGAACTTTCGAGCGCTGATGAGAGAGACGCGGAAGATTTCTCGAAGG
AAAAGCCCGCTGCACAAAGCTCACTGTTACGCCTGGAATCCGTTGTAATG
CCGGTGATCTTTACTGCATTGGCGTTGTTTACCAGGGGGCGCGCC
[SEQ ID NO: 55]

MSSSSSTGYSKNNAAHIKQENTLRQRESSSISVSEELSSADERDAEDFSKEKPA
AQSSLLRLESVVMPVIFTALALFTRGRA [SEQ ID NO: 56]

ScPMT3-s
ATGCCGTACAGAGTGGCGACGGGCTACAGTGAAAAAAAGTACTGACGATG
ATTTGATATGGAGAACGCCAATAGTAAAAGAGGAACTCGAGGATGCTGAC
AACTTTTTAAAGGATGATGCCGAGTTGTATGATAAAGTCAAGAACGAGAG
TGCAGTATCACACCTGGATACCATAGTTATGCCGATCATTTTCACGGTACT
GGGCATGTTCACTGGGCGCGCC [SEQ ID NO: 57]

MPYRVATGYSEKSTDDDLIWRTPIVKEELEDADNFLKDDAELYDKVKNESAV
SHLDTIVMPIIFTVLGMFTGRA [SEQ ID NO: 58]

ScPMT4-s
ATGTCTGTGCCCAAAAAAACGTAACCATGGGAAGTTACCACCTTCCACTAA
GGACGTAGACGATCCTTCGTTGAAGTACACGAAGGCCGCGCCTAAATGTG
AACAAGTTGCTGAACATTGGCTCTTGCAGCCACTACCGGAACCGGAATCA
CGTTATAGCTTTTGGGTAACAATTGTTACCTTATTAGCGTTTGCTGCTAGA
TTTTATAAGATCTGGTATCCAAAAGAAGTTGTTTTTGATGAGGTACATTTC

GGGAAATTTGCATCGTATTACTTAGAAAGGTCTTATTTCTTTGACGTTCAT
CCCCCTTTTGCTAAGATGATGATTGCCTTCATTGGTTGGTTATGTGGCTAT
GATGGGCGCGCC
[SEQ ID NO: 59]

MSVPKKRNHGKLPPSTKDVDDPSLKYTKAAPKCEQVAEHWLLQPLPEPESRY
SFWVTIVTLLAFAARFYKIWYPKEVVFDEVHFGKFASYYLERSYFFDVHPPFA
KMMIAFIGWLCGYDGRA [SEQ ID NO: 60]

ScPMT5-s
ATGAATAAAGAGCATTTGCTGAAGGTGGATCCCATCCCCGATGTGACTAT
TAAACGCGGCCCTTTGAGGTCTTTTCTCATAACAAAACCCTGTGATAATTT
GAGTTCATTACGAACAGTTACTTCATCTAAGGAAAAGCTTCTAGTTGGCTG
TTTGCTGATATTTACTGCCATCGTAAGGCTACACAATATCTCCGGGCGCGC
C [SEQ ID NO: 61]

MNKEHLLKVDPIPDVTIKRGPLRSFLITKPCDNLSSLRTVTSSKEKLLVGCLLIF
TAIVRLHNISGRA [SEQ ID NO: 62]

ScPMT6-s
ATGAGTAAAGCCAAGGGAACGGGATTTTCATCAATTGATACTGAAGATGA
AAACTTACGCGAACGTTATGTTAATCAACCAAAAGCTAATGCCTCCGATA
TTCAAGATGAACAATTAGATTGCTTTGAGCAACTAGAAGAAAAACATAGG
ACAAAAAAAAAATGAAGAATACACTGCGTTGAAAATTTTAAGGGATGTCAT
AGGTCCCCTTTTATTAACTATAACTTCGTTTTATCTAAGATTCCAACATATA
GATCAGAACAATTATGTTGTCTGGGATGAGGCTCATTTTGGGAAATTCGG
ATCATACTACATCAAACATGAGTACTACCACGATGTCCACCCTCCACTTGG
TAAAATGCTTATTGCATTGAGCGAATGGATGGCAGGATTTGACGGTCAAT
TTGACTTTTCCTCTAATAATGCATATCCGGAAAACGTAAACTTTAAACTAA
TGAGACAATTTAATGCCACATTTGGAGCTCTATGTACACCAGTAGCTTTCT
TTACAGCCAAATGGATGGGGTTCAATTATTTTACTGTTGGGCGCGCC
[SEQ ID NO: 63]

MSKAKGTGFSSIDTEDENLRERYVNQPKANASDIQDEQLDCFEQLEEKHRTK
KNEEYTALKILRDVIGPLLLTITSFYLRFQHIDQNNYVVWDEAHFGKFGSYYI
KHEYYHDVHPPLGKMLIALSEWMAGFDGQFDFSSNNAYPENVNFKLMRQFN
ATFGALCTPVAFFTAKWMGFNYFTVGRA [SEQ ID NO: 64]

ScPMT1-m

ATGTCGGAAGAGAAAACGTACAAACGTGTAGAGCAGGATGATCCCGTGC
CCGAACTGGATATCAAGCAGGGCCCCGTAAGACCCTTTATTGTTACCGAT
CCGAGTGCCGAATTGGCCTCGTTACGAACCATGGTCACTCTTAAAGAGAA
GCTGTTAGTGGCCTGTCTTGCTGTCTTTACAGCGGTCATTAGATTGCATGG
CTTGGCATGGCCTGACAGCGTGGTGTTTGATGAAGTACATTTCGGTGGGTT
TGCCTCGCAATACATTAGGGGGACTTACTTCATGGATGTGCATCCTCCTCT
TGCAAAGATGTTGTATGCTGGTGTGGCAGGGCGCGCC [SEQ ID NO: 65]

MSEEKTYKRVEQDDPVPELDIKQGPVRPFIVTDPSAELASLRTMVTLKEKLLV
ACLAVFTAVIRLHGLAWPDSVVFDEVHFGGFASQYIRGTYFMDVHPPLAKML
YAGVAGRA [SEQ ID NO: 66]

ScPMT2-m

ATGTCCTCGTCTTCGTCTACCGGGTACAGCAAAAACAATGCCGCCCACATT
AAGCAAGAGAATACACTGAGACAAAGAGAATCGTCTTCCATCAGCGTCAG
TGAGGAACTTTCGAGCGCTGATGAGAGAGACGCGGAAGATTTCTCGAAGG
AAAAGCCCGCTGCACAAAGCTCACTGTTACGCCTGGAATCCGTTGTAATG
CCGGTGATCTTTACTGCATTGGCGTTGTTTACCAGGATGTACAAAATCGGC
ATCAACAACCATGTTGTTTGGGATGAGGCGCACTTTGGTAAATTTGGTTCT
TATTACTTGAGACACGAATTTTACCACGATGTCCATCCTCCCCTAGGAAAA
ATGCTGGGGCGCGCC [SEQ ID NO: 67]

MSSSSSTGYSKNNAAHIKQENTLRQRESSSISVSEELSSADERDAEDFSKEKPA
AQSSLLRLESVVMPVIFTALALFTRMYKIGINNHVVWDEAHFGKFGSYYLRH
EFYHDVHPPLGKMLGRA [SEQ ID NO: 68]

ScPMT3-m

ATGCCGTACAGAGTGGCGACGGGCTACAGTGAAAAAAGTACTGACGATG
ATTTGATATGGAGAACGCCAATAGTAAAAGAGGAACTCGAGGATGCTGAC
AACTTTTTAAAGGATGATGCCGAGTTGTATGATAAAGTCAAGAACGAGAG
TGCAGTATCACACCTGGATACCATAGTTATGCCGATCATTTTCACGGTACT
GGGCATGTTCACTAGAATGTACAAGATTGGTCGTAATAATCATGTGGTCT
GGGATGAAGCTCATTTTGGTAAGTTCGGCTCTTACTATCTGAGACACGAAT
TTTACCATGATGTTCATCCACCTTTAGGTGGGCGCGCC [SEQ ID NO: 69]

MPYRVATGYSEKSTDDDLIWRTPIVKEELEDADNFLKDDAELYDKVKNESAV
SHLDTIVMPIIFTVLGMFTRMYKIGRNNHVVWDEAHFGKFGSYYLRHEFYHD
VHPPLGGRA [SEQ ID NO: 70]

ScPMT4-m
ATGTCTGTGCCCAAAAAACGTAACCATGGGAAGTTACCACCTTCCACTAA
GGACGTAGACGATCCTTCGTTGAAGTACACGAAGGCCGCGCCTAAATGTG
AACAAGTTGCTGAACATTGGCTCTTGCAGCCACTACCGGAACCGGAATCA
CGTTATAGCTTTTGGGTAACAATTGTTACCTTATTAGCGTTTGCTGCTAGA
TTTTATAAGATCTGGTATCCAAAAGAAGTTGTTTTTGATGAGGTACATTTC
GGGAAATTTGCATCGTATTACTTAGAAAGGTCTTATTTCTTTGACGTTCAT
CCCCCTTTTGCTAAGATGATGATTGCCTTCATTGGTTGGTTATGTGGCTAT
GATGGTTCCTTTAAGTTTGATGAGATTGGGTATTCTTATGAAACTCATCCA
GGGCGCGCC
[SEQ ID NO: 71]

MSVPKKRNHGKLPPSTKDVDDPSLKYTKAAPKCEQVAEHWLLQPLPEPESRY
SFWVTIVTLLAFAARFYKIWYPKEVVFDEVHFGKFASYYLERSYFFDVHPPFA
KMMIAFIGWLCGYDGSFKFDEIGYSYETHPGRA [SEQ ID NO: 72]

ScPMT5-m
ATGAATAAAGAGCATTTGCTGAAGGTGGATCCCATCCCCGATGTGACTAT
TAAACGCGGCCCTTTGAGGTCTTTTCTCATAACAAAACCCTGTGATAATTT
GAGTTCATTACGAACAGTTACTTCATCTAAGGAAAAGCTTCTAGTTGGCTG
TTTGCTGATATTTACTGCCATCGTAAGGCTACACAATATCTCCCTGCCAAA
TAGTGTTGTTTTTGGTGAAAATGAAGTTGGTACATTTGTTTCTCAATACGT
GGGGCGCGCC
[SEQ ID NO: 73]

MNKEHLLKVDPIPDVTIKRGPLRSFLITKPCDNLSSLRTVTSSKEKLLVGCLLIF
TAIVRLHNISLPNSVVFGENEVGTFVSQYVGRA [SEQ ID NO: 74]

ScPMT6-m
ATGAGTAAAGCCAAGGGAACGGGATTTTCATCAATTGATACTGAAGATGA
AAACTTACGCGAACGTTATGTTAATCAACCAAAAGCTAATGCCTCCGATA
TTCAAGATGAACAATTAGATTGCTTTGAGCAACTAGAAGAAAAACATAGG
ACAAAAAAAAAATGAAGAATACACTGCGTTGAAAATTTTAAGGGATGTCAT

AGGTCCCCTTTTATTAACTATAACTTCGTTTTATCTAAGATTCCAACATATA
GATCAGAACAATTATGTTGTCTGGGATGAGGCTCATTTTGGGAAATTCGG
ATCATACTACATCAAACATGAGTACTACCACGATGTCCACCCTCCACTTGG
TAAAATGCTTATTGCATTGAGCGAATGGATGGCAGGATTTGACGGTCAAT
TTGACTTTTCCTCTAATAATGCATATCCGGAAAACGTAAACTTTAAACTAA
TGAGACAATTTAATGCCACATTTGGAGCTCTATGTACACCAGTAGCTTTCT
TTACAGCCAAATGGATGGGGTTCAATTATTTTACTGTTTATTTGATTGCTA
CGATGGTAACGTTGGAACATTCATATATTGGGCGCGCC [SEQ ID NO: 75]

MSKAKGTGFSSIDTEDENLRERYVNQPKANASDIQDEQLDCFEQLEEKHRTK
KNEEYTALKILRDVIGPLLLTITSFYLRFQHIDQNNYVVWDEAHFGKFGSYYI
KHEYYHDVHPPLGKMLIALSEWMAGFDGQFDFSSNNAYPENVNFKLMRQFN
ATFGALCTPVAFFTAKWMGFNYFTVYLIATMVTLEHSYIGRA
[SEQ ID NO: 76]

PpPMT1-s
ATGTGCCAGATATTTCTCCCGCAAAACGTAACACGTTGTTCTGTTTCCCTT
TTGACAATGAGTAAAACAAGTCCTCAAGAGGTGCCAGAAAACACTACTGA
GCTTAAAATCTCAAAAGGAGAGCTCCGTCCTTTTATTGTGACCTCTCCATC
TCCTCAATTGAGCAAGTCTCGTTCTGTGACTTCAACCAAGGAGAAGCTGAT
ATTGGCTAGTTTGTTCATATTTGCAATGGTCATCAGGTTCCACAACGTCGC
CGGGCGCGCC [SEQ ID NO: 77]

MCQIFLPQNVTRCSVSLLTMSKTSPQEVPENTTELKISKGELRPFIVTSPSPQLS
KSRSVTSTKEKLILASLFIFAMVIRFHNVAGRA [SEQ ID NO: 78]

PpPMT2-s
ATGACAGGCCGTGTCGACCAGAAATCTGATCAGAAGGTGAAGGAATTGAT
CGAAAAGATCGACTCCGAATCCACTTCCAGAGTTTTTCAGGAAGAACCAG
TCACTTCGATCTTGACACGTTACGAACCCTATGTCGCCCCAATTATATTCA
CGTTGTTGTCCTTTTTCACTCGTATGTACAAAATTGGGATCAACAACCACG
TCGTTTGGGATGAAGCTCACTTCGGAAAGTTTGGCTCCTACTATCTCAGAC
ACGAGTTCTACCACGATGTCCACCCTCCGTTGGGTAAGATGTTGGTCGGTC
TATCTGGCTACATTGCCGGTTACAATGGCTCCTGGGATTTCCCCTCCGGTC
AAGAGTACCCTGACTATATTGATTACGTTAAAATGAGGTTATTCAATGCCA
CCTTCAGTGCCTTATGTGTGCCATTCGCCTATTTCACCATGAAGGAGATTG
GATTTGATATCAAGACAACTTGGCTATTCACACTGATGGTCTTGTGTGAAA

CAAGTTATTGTACGTTAGGAAAATTCATCTTGCTGGATTCAATGCTGCTGC
TATTCACTGTGACTACGGTTTTCACCTTTGTTAGGGGGCGCGCC
[SEQ ID NO: 79]

MTGRVDQKSDQKVKELIEKIDSESTSRVFQEEPVTSILTRYEPYVAPIIFTLLSF
FTRMYKIGINNHVVWDEAHFGKFGSYYLRHEFYHDVHPPLGKMLVGLSGYI
AGYNGSWDFPSGQEYPDYIDYVKMRLFNATFSALCVPFAYFTMKEIGFDIKT
TWLFTLMVLCETSYCTLGKFILLDSMLLLFTVTTVFTFVRGRA
[SEQ ID NO: 80]

PpPMT4-s
ATGATAAAATCAAGAAAGAGATCGAGAAAAGTTTCTTTGAACACTGAAAA
GGAGCTGAAAAATAGCCATATTTCTCTTGGAGATGAAAGATGGTACACTG
TGGGTCTTCTCTTGGTGACAATCACAGCTTTCTGTACTCGATTCTATGCTAT
CAACGGGCGCGCC [SEQ ID NO: 81]

MIKSRKRSRKVSLNTEKELKNSHISLGDERWYTVGLLLVTITAFCTRFYAING
RA [SEQ ID NO: 82]

PpPMT5-s
ATGACATTCTTCTTATTAGACTGCCTAGTTTTGTATAATCTTACAGAAATTC
TAGCTCAAGCCCTCTTACTTGTTCTTCTTCTATGTCAACTGATTCCTCAATA
TATGTGGTTGGTGGCCGGGCGCGCC [SEQ ID NO: 83]

MTFFLLDCLVLYNLTEILAQALLLVLLLCQLIPQYMWLVAGRA
[SEQ ID NO: 84]

PpPMT6-s
ATGGCAACAGAGGAAGAGAGAAATGAACTGAGAAGTCGGATGGACGCCA
ATAATTCAAAAGTTTCCACGTTCACTACGAACAATTCAGATGATCCTTCTG
TTGATAGCCAGGGTAAGGTGAAAATTAAGTCATGGGTTTGGAGCCTTGAA
TCTTTAATTGGCCCTCTGGTGATCACTGCCTTGGCAATTTTTCTTCGAGTTT
ACCAAGGGCGCGCC [SEQ ID NO: 85]

MATEEERNELRSRMDANNSKVSTFTTNNSDDPSVDSQGKVKIKSWVWSLESL
IGPLVITALAIFLRVYQGRA [SEQ ID NO: 86]

PpPMT1-m
ATGTGCCAGATATTTCTCCCGCAAAACGTAACACGTTGTTCTGTTTCCCTT
TTGACAATGAGTAAAACAAGTCCTCAAGAGGTGCCAGAAAACACTACTGA
GCTTAAAATCTCAAAAGGAGAGCTCCGTCCTTTTATTGTGACCTCTCCATC
TCCTCAATTGAGCAAGTCTCGTTCTGTGACTTCAACCAAGGAGAAGCTGAT
ATTGGCTAGTTTGTTCATATTTGCAATGGTCATCAGGTTCCACAACGTCGC
CCACCCTGACAGCGTTGTGTTTGATGAAGTTCACTTTGGGGGGTTTGCCAG
AAAGTACATTTTGGGAACCTTTTTCATGGATGTTCATCCGCCATTGGCCAA
GCTATTATTTGCTGGTGTTGGCAGTCTTGGTGGAGGGCGCGCC
[SEQ ID NO: 87]

MCQIFLPQNVTRCSVSLLTMSKTSPQEVPENTTELKISKGELRPFIVTSPSPQLS
KSRSVTSTKEKLILASLFIFAMVIRFHNVAHPDSVVFDEVHFGGFARKYILGTF
FMDVHPPLAKLLFAGVGSLGGGRA [SEQ ID NO: 88]

PpPMT2-m
ATGACAGGCCGTGTCGACCAGAAATCTGATCAGAAGGTGAAGGAATTGAT
CGAAAAGATCGACTCCGAATCCACTTCCAGAGTTTTTCAGGAAGAACCAG
TCACTTCGATCTTGACACGTTACGAACCCTATGTCGCCCCAATTATATTCA
CGTTGTTGTCCTTTTTCACTCGTATGTACAAAATTGGGATCAACAACCACG
TCGTTTGGGATGAAGCTCACTTCGGAAAGTTTGGCTCCTACTATCTCAGAC
ACGAGTTCTACCACGATGTCCACCCTCCGTTGGGTAAGATGTTGGTCGGTC
TATCTGGCTACATTGCCGGTTACAATGGCTCCTGGGATTTCCCCTCCGGTC
AAGAGTACCCTGACTATATTGATTACGTTAAAATGAGGTTATTCAATGCCA
CCTTCAGTGCCTTATGTGTGCCATTCGCCTATTTCACCATGAAGGAGATTG
GATTTGATATCAAGACAACTTGGCTATTCACACTGATGGTCTTGTGTGAAA
CAAGTTATTGTACGTTAGGAAAATTCATCTTGCTGGATTCAATGCTGCTGC
TATTCACTGTGACTACGGTTTTCACCTTTGTTAGGTTCCATAACGAAAACA
GTAAACCAGGAAACTCGTTTGGGCGCGCC [SEQ ID NO: 89]

MTGRVDQKSDQKVKELIEKIDSESTSRVFQEEPVTSILTRYEPYVAPIIFTLLSF
FTRMYKIGINNHVVWDEAHFGKFGSYYLRHEFYHDVHPPLGKMLVGLSGYI
AGYNGSWDFPSGQEYPDYIDYVKMRLFNATFSALCVPFAYFTMKEIGFDIKT
TWLFTLMVLCETSYCTLGKFILLDSMLLLFTVTTVFTFVRFHNENSKPGNSFG
RA [SEQ ID NO: 90]

PpPMT4-m
ATGATAAAATCAAGAAAGAGATCGAGAAAAGTTTCTTTGAACACTGAAAA
GGAGCTGAAAAATAGCCATATTTCTCTTGGAGATGAAAGATGGTACACTG
TGGGTCTTCTCTTGGTGACAATCACAGCTTTCTGTACTCGATTCTATGCTAT
CAACTATCCAGATGAGGTTGTTTTTGACGAAGTTCATTTCGGAGGGCGCGC
C [SEQ ID NO: 91]

MIKSRKRSRKVSLNTEKELKNSHISLGDERWYTVGLLLVTITAFCTRFYAINYP
DEVVFDEVHFGGRA [SEQ ID NO: 92]

PpPMT5-m
ATGACATTCTTCTTATTAGACTGCCTAGTTTTGTATAATCTTACAGAAATTC
TAGCTCAAGCCCTCTTACTTGTTCTTCTTCTATGTCAACTGATTCCTCAATA
TATGTGGTTGGTGGCCCGCGAAATGACTCCTGAGATATTTGGTCAAACCTA
CCAAAGGACACCACACCACAGTACTATAGCACAACAATACATGGCCGCCT
TTGAGTACAAAAAGGGCATTCAAAGACCCTATTTTGGGCGCGCC
[SEQ ID NO: 93]

MTFFLLDCLVLYNLTEILAQALLLVLLLCQLIPQYMWLVAREMTPEIFGQTYQ
RTPHHSTIAQQYMAAFEYKKGIQRPYFGRA [SEQ ID NO: 94]

PpPMT6-m
ATGGCAACAGAGGAAGAGAGAAATGAACTGAGAAGTCGGATGGACGCCA
ATAATTCAAAAGTTTCCACGTTCACTACGAACAATTCAGATGATCCTTCTG
TTGATAGCCAGGGTAAGGTGAAAATTAAGTCATGGGTTTGGAGCCTTGAA
TCTTTAATTGGCCCTCTGGTGATCACTGCCTTGGCAATTTTTCTTCGAGTTT
ACCAAATAGGAAAAGCTGATAGGGTTGTTTGGGATGAAGCTCATTTCGGA
AAGTTTGGGTCATTCTACTTGAAGCACCAGTTCTATTTTGATGTCCATCCT
CCCCTGGGAAAACTTCTTACAGGTTTGGGGCGCGCC [SEQ ID NO: 95]

MATEEERNELRSRMDANNSKVSTFTTNNSDDPSVDSQGKVKIKSWVWSLESL
IGPLVITALAIFLRVYQIGKADRVVWDEAHFGKFGSFYLKHQFYFDVHPPLGK
LLTGLGRA [SEQ ID NO: 96]

ScBOS1-s
ATGAACGCTCTTTACAACCATGCTGTGAAGCAAAAAAAATCAACTACAACA
AGAGTTGGCCAGGTTTGAAAAGAATTCTGTGACCGCCCCTGGGCGCGCC
[SEQ ID NO: 97]

MNALYNHAVKQKNQLQQELARFEKNSVTAPGRA [SEQ ID NO: 98]

ScBET1-s
ATGAGTTCAAGATTTGCAGGGGGAAACGCTTATCAACGTGATACTGGTAG
AACACAGTTATTCGGACCGGCTGATGGATCAAATAGTCTCGATGACAATG
GGCGCGCC [SEQ ID NO: 99]

MSSRFAGGNAYQRDTGRTQLFGPADGSNSLDDNGRA [SEQ ID NO: 100]

ScSEC22-s
ATGTCCGCGCAAAAGATCAACTTCGATCTCTTGATCAGTCAATATGCTCCT
ATTGTCATTGTCGCTTTCTTTTTCGTCTTTCTCTTCTGGTGGATCTTCCTCAA
AGGGCGCGCC [SEQ ID NO: 101]

MSAQKINFDLLISQYAPIVIVAFFFVFLFWWIFLKGRA [SEQ ID NO: 102]

PpBOS1-s
ATGAAGGCATTTGAAGACAAGTGGATTTTTTATGGTGGCGCTATAAGTGTT
TTTGTTATTTTCTATTTGGCGGTCAAATATTTAAGAGGGCGCGCC
[SEQ ID NO: 103]

MKAFEDKWIFYGGAISVFVIFYLAVKYLRGRA [SEQ ID NO: 104]

PpBET1-s
ATGAGGATGATGGTAATGGCTAAGAAAACAGGTATTTCATGGAAGTTATG
GCTGCTGTTCTTCTTCCTCGTCTGGCTTTGGTTCTTTTTTGTGTGGCTTAGA
GGGCGCGCC [SEQ ID NO: 105]

MRMMVMAKKTGISWKLWLLFFFLVWLWFFFVWLRGRA [SEQ ID NO: 106]

PpSEC22-s
ATGGCAGCTCGAAGAATCAATTTGGAGGCTCTGATAAAACAGTACGTTCC
GGTTGCAATGGTGGGGATTTTCTTCGTATTTATAATATGGTGGATATTCTT
GCGCGGGCGCGCC [SEQ ID NO: 107]

MAARRINLEALIKQYVPVAMVGIFFVFIIWWIFLRGRA [SEQ ID NO: 108]

EXAMPLE 2

Construction of Targeting Domain ("TD") Libraries

[0134] All TDs were amplified by PCR using ExTaq (Takara) and the respective fungal genomic DNA as template, cloned into plasmids pCR2.1 (Invitrogen) and sequenced. The 5'-oligo introduced a *Not*I site and the CACC Kozak consensus sequence just upstream of the native ATG. In the case of the leaders derived from PpSEC12 or ScSEC12 the codon for the first amino acid of the TD was changed to ATG. The 3'-oligo introduced an *Asc*I site and an additional C, which resulted in a fusion linker encoding GlyArgAla. After amplification of the targeting domain containing plasmids, 36 µg of each was digested with 50 units *Asc*I for 4 hr. The linearized DNA was then ethanol precipitated, digested with 60 units *Not*I for 15 hr and the resulting inserts were purified by two consecutive rounds of separation on agarose gels. For fragments ranging from 87 to 219 nucleotides we used 2 % agarose, for fragments from 240 to 426 nucleotides 1.5 % agarose, and for fragments from 435 to 1098 nucleotides 1.0 % agarose. 1.2 pmole of each was diluted to 100 µl, arranged in a 96 well plate and stored at –80° C.

EXAMPLE 3

Generation of POMGnT1 Expression Vectors

[0135] Nucleic acids (codon-optimized for *P. pastoris*) encoding the catalytic domains of the human, mouse, chicken, fish, and frog POMGnT1 proteins were synthesized by GeneArt AG (Regensburg, Germany) using amino acid sequences taken from GenBank deposits listed below. Non-codon optimized sequences for human, mouse, chicken, zebrafish and xenopus POMGnT1 are disclosed in the prior art; *see, e.g.,* Q8WZA1 for human; NP_080927 for mouse; XP_426653 for chicken; NP_001036152 for zebrafish; and AAH84747 for xenopus. Nucleotide and amino acid sequences (SEQ ID NOs: 109-118) encoding C-terminal catalytic domains of POMGnT1 are provided. The sequences lack the native *N*-terminal signal sequence and transmembrane anchoring regions. The nucleic acid sequences are codon-

optimized for expression in *Pichia pastoris*. In all cases, an in-frame *Asc*1 site was added to the 5' end, and a *Pac*1 site added just 3' to the STOP codon.

human POMGnT1

CGCGCCATTTCTGAAGCTAACGAGGACCCTGAACCAGAACAAGATTACGA
CGAGGCTTTGGGAAGATTGGAACCACCAAGAAGAAGAGGTTCCGGTCCA
AGAAGAGTTTTGGACGTTGAGGTTTACTCTTCCAGATCCAAGGTTTACGTT
GCTGTTGACGGTACTACTGTTTTGGAGGACGAGGCTAGAGAACAAGGTAG
AGGTATCCACGTTATCGTTTTGAACCAGGCTACTGGTCATGTTATGGCTAA
GAGAGTTTTCGACACTTACTCTCCACACGAAGATGAGGCTATGGTTTTGTT
CTTGAACATGGTTGCTCCAGGTAGAGTTTTGATTTGTACTGTTAAGGACGA
GGGATCCTTCCATTTGAAGGACACTGCTAAGGCTTTGTTGAGATCCTTGGG
TTCTCAAGCTGGTCCAGCTTTGGGATGGAGAGATACTTGGGCTTTCGTTGG
TAGAAAGGGTGGTCCAGTTTTCGGTGAAAAGCACTCTAAGTCCCCAGCTT
TGTCCTCTTGGGGTGACCCAGTTTTGTTGAAAACTGACGTTCCATTGTCCT
CTGCTGAAGAGGCTGAATGTCACTGGGCTGACACTGAGTTGAACAGAAGA
AGAAGAAGATTCTGTTCCAAGGTTGAGGGTTACGGTTCTGTTTGTTCCTGT
AAGGACCCAACTCCAATTGAATTCTCCCCAGACCCATTGCCAGATAACAA
GGTTTTGAACGTTCCAGTTGCTGTTATCGCTGGTAACAGACCAAACTACTT
GTACAGAATGTTGAGATCTTTGTTGTCCGCTCAGGGAGTTTCTCCACAGAT
GATCACTGTTTTCATCGACGGTTACTACGAAGAACCAATGGACGTTGTTGC
TTTGTTCGGATTGAGAGGTATTCAGCACACTCCAATCTCCATCAAGAACGC
TAGAGTTTCCCAACACTACAAGGCTTCCTTGACTGCTACTTTCAACTTGTT
CCCAGAGGCTAAGTTCGCTGTTGTTTTGGAAGAGGACTTGGACATTGCTGT
TGATTTCTTCTCCTTCTTGTCCCAATCCATCCACTTGTTGGAAGAGGATGA
CTCCTTGTACTGTATCTCTGCTTGGAACGACCAAGGTTACGAACACACTGC
TGAGGATCCAGCTTTGTTGTACAGAGTTGAGACTATGCCAGGATTGGGAT
GGGTTTTGAGAAGATCCTTGTACAAAGAAGAGTTGGAGCCAAAGTGGCCA
ACTCCAGAAAAGTTGTGGGATTGGGACATGTGGATGAGAATGCCAGAGCA
GAGAAGAGGTAGAGAGTGTATCATCCCAGACGTTTCCAGATCTTACCACT
TCGGTATTGTTGGATTGAACATGAACGGTTACTTCCACGAGGCTTACTTCA
AGAAGCACAAGTTCAACACTGTTCCAGGTGTTCAGTTGAGAAACGTTGAC
TCCTTGAAGAAAGAGGCTTACGAGGTTGAGGTTCACAGATTGTTGTCTGA
GGCTGAGGTTTTGGACCATTCCAAGAACCCATGTGAGGACTCATTCTTGCC
AGATACTGAGGGTCATACTTACGTTGCTTTCATCAGAATGGAAAAGGACG
ACGACTTCACTACTTGGACTCAGTTGGCTAAGTGTTTGCACATTTGGGACT
TGGATGTTAGAGGTAACCACAGAGGATTGTGGAGATTGTTCAGAAAGAAG

AACCACTTCTTGGTTGTTGGTGTTCCAGCTTCTCCATACTCCGTTAAGAAG
CCACCATCCGTTACTCCAATTTTCTTGGAGCCACCACCAAAGGAAGAAGG
TGCTCCTGGTGCTCCAGAGCAAACTTAATAGTTAATTAAA
[SEQ ID NO: 109]

RAISEANEDPEPEQDYDEALGRLEPPRRRGSGPRRVLDVEVYSSRSKVYVAV
DGTTVLEDEAREQGRGIHVIVLNQATGHVMAKRVFDTYSPHEDEAMVLFLN
MVAPGRVLICTVKDEGSFHLKDTAKALLRSLGSQAGPALGWRDTWAFVGRK
GGPVFGEKHSKSPALSSWGDPVLLKTDVPLSSAEEAECHWADTELNRRRRRF
CSKVEGYGSVCSCKDPTPIEFSPDPLPDNKVLNVPVAVIAGNRPNYLYRMLRS
LLSAQGVSPQMITVFIDGYYEEPMDVVALFGLRGIQHTPISIKNARVSQHYKA
SLTATFNLFPEAKFAVVLEEDLDIAVDFFSFLSQSIHLLEEDDSLYCISAWNDQ
GYEHTAEDPALLYRVETMPGLGWVLRRSLYKEELEPKWPTPEKLWDWDMW
MRMPEQRRGRECIIPDVSRSYHFGIVGLNMNGYFHEAYFKKHKFNTVPGVQL
RNVDSLKKEAYEVEVHRLLSEAEVLDHSKNPCEDSFLPDTEGHTYVAFIRME
KDDDFTTWTQLAKCLHIWDLDVRGNHRGLWRLFRKKNHFLVVGVPASPYS
VKKPPSVTPIFLEPPPKEEGAPGAPEQT [SEQ ID NO: 110]

Mouse POMGnT1

CGCGCCATTTCTGAAGCTAACGAGGACCCTGAACCAGAACAAGATTACGA
CGAGGCTTTGGGAAGATTGGAATCCCCAAGAAGAAGAGGATCCTCCCCTA
GAAGAGTTTTGGACGTTGAGGTTTACTCTTCCAGATCCAAGGTTTACGTTG
CTGTTGACGGTACTACTGTTTTGGAGGACGAGGCTAGAGAACAAGGTAGA
GGTATCCACGTTATCGTTTTGAACCAGGCTACTGGTCATGTTATGGCTAAG
AGAGTTTTCGACACTTACTCTCCACACGAAGATGAGGCTATGGTTTTGTTC
TTGAACATGGTTGCTCCAGGTAGAGTTTTGATTTGTACTGTTAAGGACGAG
GGATCCTTCCATTTGAAGGACACTGCTAAGGCTTTGTTGAGATCCTTGGGT
TCTCAAGCTGGTCCAGCTTTGGGATGGAGAGATACTTGGGCTTTCGTTGGT
AGAAAGGGTGGTCCAGTTTTGGGTGAAAAGCACTCTAAGTCCCCAGCTTT
GTCCTCTTGGGGTGACCCAGTTTTGTTGAAAACTGACGTTCCATTGTCCTC
TGCTGAAGAGGCTGAATGTCACTGGGCTGACACTGAGTTGAACAGAAGAA
GAAGAAGATTCTGTTCCAAGGTTGAGGGTTACGGTTCTGTTTGTTCCTGTA
AGGACCCAACTCCAATTGAATTCTCCCCAGACCCATTGCCAGATAACAAG
GTTTTGAACGTTCCAGTTGCTGTTATCGCTGGTAACAGACCAAACTACTTG
TACAGAATGTTGAGATCTTTGTTGTCCGCTCAGGGAGTTTCTCCACAGATG
ATCACTGTTTTCATCGACGGTTACTACGAAGAACCAATGGACGTTGTTGCT
TTGTTCGGATTGAGAGGTATTCAGCACACTCCAATCTCCATCAAGAACGCT

AGAGTTTCCCAACACTACAAGGCTTCCTTGACTGCTACTTTCAACTTGTTC
CCAGAGGCTAAGTTCGCTGTTGTTTTGGAAGAGGACTTGGACATTGCTGTT
GATTTCTTCTCCTTCTTGTCCCAATCCATCCACTTGTTGGAAGAGGATGAC
TCCTTGTACTGTATCTCTGCTTGGAACGACCAAGGTTACGAACACACTGCT
GAGGATCCAGCTTTGTTGTACAGAGTTGAGACTATGCCAGGATTGGGATG
GGTTTTGAGAAAGTCCTTGTACAAAGAGGAGTTGGAGCCAAAGTGGCCAA
CTCCAGAAAAGTTGTGGGATTGGGACATGTGGATGAGAATGCCAGAGCAG
AGAAGAGGTAGAGAGTGTATCATCCCAGACGTTTCCAGATCTTACCACTT
CGGTATTGTTGGATTGAACATGAACGGTTACTTCCACGAGGCTTACTTCAA
GAAGCACAAGTTCAACACTGTTCCAGGTGTTCAGTTGAGAAACGTTGACT
CCTTGAAGAAAGAGGCTTACGAGGTTGAGATCCACAGATTGTTGTCTGAG
GCTGAGGTTTTGGATCACTCCAAGGATCCATGTGAGGACTCATTCTTGCCA
GATACTGAGGGTCATACTTACGTTGCTTTCATCAGAATGGAAACTGACGA
CGACTTTGCTACTTGGACTCAGTTGGCTAAGTGTTTGCACATTTGGGACTT
GGATGTTAGAGGTAACCACAGAGGATTGTGGAGATTGTTCAGAAAGAAG
AACCACTTCTTGGTTGTTGGTGTTCCAGCTTCTCCATACTCCGTTAAGAAG
CCACCATCCGTTACTCCAATTTTCTTGGAGCCACCACCAAAGGAAGAAGG
TGCTCCTGGAGCTGCTGAACAAACTTAGTAGTTAA [SEQ ID NO: 111]

RAISEANEDPEPEQDYDEALGRLESPRRRGSSPRRVLDVEVYSSRSKVYVAVD
GTTVLEDEAREQGRGIHVIVLNQATGHVMAKRVFDTYSPHEDEAMVLFLNM
VAPGRVLICTVKDEGSFHLKDTAKALLRSLGSQAGPALGWRDTWAFVGRKG
GPVLGEKHSKSPALSSWGDPVLLKTDVPLSSAEEAECHWADTELNRRRRRFC
SKVEGYGSVCSCKDPTPIEFSPDPLPDNKVLNVPVAVIAGNRPNYLYRMLRSL
LSAQGVSPQMITVFIDGYYEEPMDVVALFGLRGIQHTPISIKNARVSQHYKAS
LTATFNLFPEAKFAVVLEEDLDIAVDFFSFLSQSIHLLEEDDSLYCISAWNDQG
YEHTAEDPALLYRVETMPGLGWVLRKSLYKEELEPKWPTPEKLWDWDMW
MRMPEQRRGRECIIPDVSRSYHFGIVGLNMNGYFHEAYFKKHKFNTVPGVQL
RNVDSLKKEAYEVEIHRLLSEAEVLDHSKDPCEDSFLPDTEGHTYVAFIRMET
DDDFATWTQLAKCLHIWDLDVRGNHRGLWRLFRKKNHFLVVGVPASPYSV
KKPPSVTPIFLEPPPKEEGAPGAAEQT [SEQ ID NO: 112]

Chicken POMGnT1

CGCGCCGCTTATGAAGAAGAGGAAGAGTCCGCTCAAGATTACGACGACG
AGATGTTGAATGTTGAGGCTCCAAGACACCCAGTTTCCAACAAGAAGGTT
TTGGACGTTGAGGTTTACTCTTCCAGATCCAAGGTTTACGTTGCTGTTGAC
GGTACTACTGTTTTGGAGGACGAGGCTAGAGAACAAGGTAGAGGTATCCA

CGTTATCGTTTTGAACCAGGCTACTGGTCATGTTATGGCTAAGAGAGTTTT
CGACACTTACTCTCCACACGAAGATGAGGCTATGGTTTTGTTCTTGAACAT
GGTTGCTAGAGGTAGAATCTTGATCTTCACTATCAAGGACGAGGGATCCT
TCCACTTGAAAGAGACTGCTAAGAACGTTTTGAAGTCCTTGGGTTCCCAA
GTTGCTCCATTCTTGTCTTGGAGAGACATGTGGACTTTTGTTGGAAAGAAG
GGTGGAGAAGTTTACGGTGAAAAGCACGCTAAGTCTCCAGCTTTGTCTAC
TTGGGGTGACCCAGTTTTGTTGAAAACTGAGGTTCACTTGACTTCCGTTGA
GGATGCTGAATGTCACTGGCCAGACACTGAGTTGAACAGAAGAAGAAGA
AGATTCTGTTCCAAGGTTGAGGGTTACGGTTCTGTTTGTTCCTGTAAGGAC
CCAACTCCAATCGAATTCAACCCAGACCCATTGAAGGACAACAAGGTTTT
CGATGTTCCAGTTGCTGTTATCGCTGGTAACAGACCAAACTACTTGTACAG
AATGTTGAGATCCTTGTTGTCCGCTCAAGGTGTTAACCCACAGATGATCAC
TGTTTTCATCGACGGTTACTACGAAGAACCAATGGACGTTGTTGAGTTGTT
CGGTTTGTCCGGTATTCAACACACTCCAATCTCCATCAAGAACGCTAGAGT
TTCCCAACACTACAAGGCTTCCTTGACTGCTACTTTCAACTTGTTCCCAGA
CGCTAAGTTCGCTGTTGTTTTGGAAGAGGACTTGGACATTTCCGTTGATTT
CTTCTCCTTCTTGTCCCAATCCATCCACTTGTTGGAAGAGGATGAGTCCTT
GTACTGTATCTCTGCTTGGAACGACCAAGGTTACGAACACACTGCTGAGG
ATCCATCCTTGTTGTACAGAGTTGAGACTATGCCAGGATTGGGATGGGTTT
TGAGAAAGTCATTGTATAAGGACGAATTGGAACCAAAGTGGCCAACTCCA
GAAAAGTTGTGGGATTGGGACATGTGGATGAGAATGCCAGAGCAGAGAA
AGGGTAGAGAGTGTATCATTCCAGACATCTCCAGATCTTACCACTTCGGTA
TTGTTGGATTGAACATGAACGGTTACTTCCACGAGGCTTACTTCAAGAAGC
ACAAGTTCAACACTGTTCCAAACGTTCAGTTGAAGAACGTTGAGTCCTTG
AGAAAGGACGCTTACGAAGCTGAGATCCACAGATTGTTGGGTGAAGCTGA
GGTTTTGGACCACTCCAAGAACCCATGTGAGGATTCTTTCGTTCCTGACAC
TGAGGGTAAAGTTTACGTTATGTTCATCAAGATGGAACAAGAGGCTGACT
TCACTACTTGGACTCAGTTGGCTAAAGAATTGATGGCTTAGTAGTTAATTA
A [SEQ ID NO: 113]

RAAYEEEEESAQDYDDEMLNVEAPRHPVSNKKVLDVEVYSSRSKVYVAVDG
TTVLEDEAREQGRGIHVIVLNQATGHVMAKRVFDTYSPHEDEAMVLFLNMV
ARGRILIFTIKDEGSFHLKETAKNVLKSLGSQVAPFLSWRDMWTFVGKKGGE
VYGEKHAKSPALSTWGDPVLLKTEVHLTSVEDAECHWPDTELNRRRRRFCS
KVEGYGSVCSCKDPTPIEFNPDPLKDNKVFDVPVAVIAGNRPNYLYRMLRSL
LSAQGVNPQMITVFIDGYYEEPMDVVELFGLSGIQHTPISIKNARVSQHYKAS
LTATFNLFPDAKFAVVLEEDLDISVDFFSFLSQSIHLLEEDESLYCISAWNDQG

YEHTAEDPSLLYRVETMPGLGWVLRKSLYKDELEPKWPTPEKLWDWDMW
MRMPEQRKGRECIIPDISRSYHFGIVGLNMNGYFHEAYFKKHKFNTVPNVQL
KNVESLRKDAYEAEIHRLLGEAEVLDHSKNPCEDSFVPDTEGKVYVMFIKME
QEADFTTWTQLAKELMA
[SEQ ID NO: 114]

Zebrafish POMGnT1

CGCCGCTTCTGAAGATGATGCTGCTCAAGAATACGATGACGCTTTGCCAA
ACATGGAAACTCCAAGAAGACCAGCTTCCGGTAGAAAGGTTTTGGACATC
GAGGTTTACTCTTCCAGATCCAAGGTTTACGTTGCTGTTGACGGTACTACT
GTTTTGGAGGACGAGATTAGAGAACAGGGTAGAGGTATCCACGTTATCGT
TTTGAACCAGGCTACTGGTCATGTTATGGCTAAGAGAGTTTTCGACACTTA
CTCTCCACACGAAGATGAGGCTATGATCTTGTTCTTGAACATGGTTACTAG
AGGTAGAATCTTGATCTTCACTATCAAGGACGAGGGAACTTTCCATTTGA
AGGACGCTGCTAAGAACTTGTTGAAGGGATTGGGTTCCCAAGTTGCTGTT
ACTTTGGGATGGAGAGACATGTGGACTTTGGTTGTTAAGAAGGGTGGACA
GGTTTACGGTGAAAAGCACTCTAAGTCCCCAGCTTTGTCTACTTGGGGTGA
CCCAGTTTTGTTGAAAACTGAGGTTCAGTTGACTGCTTCTGAAGAGGCTGA
ATGTCACTGGGCTGACACTGAGTTGAACAGAAGAAGAAAGTTGTTCTGTT
CCAAGGTTGAAGGTTACGGTTCTATCTGTTCCTGTAAGGACCCAGCTCCAA
TTGAATTCAACCCAGATCCATTGTCCAACAACAACGTTTACAACATCCCTG
TTGCTGTTATCGCTGGTAACAGACCAAACTACTTGTACAGAATGTTGAGAT
CCTTGTTGTCCTCTCACGGTGTTAACCCACAGATGATCACTGTTTTCATCG
ACGGTTACTACGAAGAACCAATGGACGTTGTTGACTTGTTCGGATTGAAG
GGTGTTCAACACACTCCAATCTCCATCAAGAACGCTAGAGTTTCCCAACA
CTACAAGGCTTCCTTGACTGCTACTTTCAACTTGCACCCAGATGCTGACTT
CGCTATCGTTTTGGAAGAGGACTTGGACATTTCCATCGATTTCTTCTCATT
CTTGGGACAGACTATCCACTTGTTGCACGAGGACGATTCCTTGTACTGTAT
CTCCGCTTGGAACGACCAAGGTTACGAACACACTGCTGAGGATCCATCCT
TGTTGTACAGAGTTGAGTCCATGCCAGGATTGGGATGGGTTTTGAAGAAG
TCATTGTATAAGGACGAATTGGAACCAAAGTGGCCAACTCCAGAAAAGTT
GTGGGATTGGGACATGTGGATGAGAATGCCAGAGCAGAGAAAGGGAAGA
GAGTGTGTTATTCCAGACGTTTCCAGATCTTACCACTTCGGTATCATCGGA
TTGAACATGAACGGTTACTTCCACGAGGTTTACTTCAAGAAGCACAAGTT
CAACACTATCCCAAACGTTCAGATGAAGAACGTTGAGAACTTGAAGAAGG
ACCCATACGAGATTGAGATCCAAAACTTGTTGAGAGAGGCTGAAGTTTTG
GACCACTCCAAGAACCCATGTGAGGATTCCTTCATCCCAGACACTGAGGG

AAAGACTTTCGTTATGTTCATCAAGATGGAACAAGAGACTGACACTAACA
CTTGGACTGAGTTGGCTAAGTGTTTGCATGTTTGGGACTTGGATGTTAGAG
GTTACCACAAGGGTTTGTGGAGATTGTTCAGAAAGAAGAACCACATCTTG
GTTGTTGCTTTCCCAATTTCCCCATACTCCGTTAAGAAGCCATCCAACGTT
ACTCCAATCCACTTGGAACCAGCTCCAAAAGAAGAAGGTCCACCAGTTGA
GCAGATGTAGTAGTTAA
[SEQ ID NO: 115]

RAASEDDAAQEYDDALPNMETPRRPASGRKVLDIEVYSSRSKVYVAVDGTT
VLEDEIREQGRGIHVIVLNQATGHVMAKRVFDTYSPHEDEAMILFLNMVTRG
RILIFTIKDEGTFHLKDAAKNLLKGLGSQVAVTLGWRDMWTLVVKKGGQVY
GEKHSKSPALSTWGDPVLLKTEVQLTASEEAECHWADTELNRRRKLFCSKVE
GYGSICSCKDPAPIEFNPDPLSNNNVYNIPVAVIAGNRPNYLYRMLRSLLSSHG
VNPQMITVFIDGYYEEPMDVVDLFGLKGVQHTPISIKNARVSQHYKASLTATF
NLHPDADFAIVLEEDLDISIDFFSFLGQTIHLLHEDDSLYCISAWNDQGYEHTA
EDPSLLYRVESMPGLGWVLKKSLYKDELEPKWPTPEKLWDWDMWMRMPE
QRKGRECVIPDVSRSYHFGIIGLNMNGYFHEVYFKKHKFNTIPNVQMKNVEN
LKKDPYEIEIQNLLREAEVLDHSKNPCEDSFIPDTEGKTFVMFIKMEQETDTNT
WTELAKCLHVWDLDVRGYHGLWRLFRKKNHILVVAFPISPYSVKKPSNVTPI
HLEPAPKEEGPPVEQM
[SEQ ID NO: 116]

Xenopus POMGnT1

CGCGCCGTTAACGAAGAAGAGATCGACCAAGACTACGACGAATCCTTGCA
ACAAGCTGACTCTCCAAGAAGACCAGCTAACTCCAAGAAGGTTTTGGACA
CTGAGATCTACTCTTCCAGATCCAAGGTTTACATTGCTGTTGACGGTACTA
CTGTTTTGGAGGACGAGGTTCACGAACAAGGTAGAGGTATCCACGTTATC
GTTTTGAACCAGGCTACTGGTCATGTTATGGCTAAGAGAGTTTTCGACACT
TACTCTCCACACGAAGATGAGGCTATGGTTTTGTTCTTGAACATGGTTGCT
AGAGGTAGAATCTTGATCTTCACTATCAAGGACGAGGGATCCTTTCACTTG
AAGGACACTGCTAAGAACTTGTTGAAGTCCTTGGGTTCCCAAATTGCTCCA
TCCTTGGGATGGAGAGACATGTGGACTTTCGTTGTTAAGAAGGGTGGACA
GGTTTACGGTGAAAAGCACTCTAAGTCCCCAGCTTTGTCTACTTGGGGTGA
CCCAATCTTGTTGAAAACTGACATCCAGTTGGTTCCACCAGAGGATGCTG
AATGTCACTGGCCAGACACTGAGTTGAACAGAAGAAGAAAGAGATTCTGT
TCCAAGGTTGAGGGTTACGGTTCTGTTTGTTCCTGTAAGGACCCAACTCCA
ATCGAATTCAACCCAATGCCATTGAAAGAGAACAAGGTTACAACTGTTCC

AGTTGCTGTTATCGCTGGTAACAGACCAAACTACTTGTACAGAATGTTGA
GATCCTTGTTGTCCGCTCAGGGAGTTTCTCCACAGATGATCACTGTTTTCA
TCGACGGTTACTACGAAGAACCAATGGACGTTGTTGAGTTGTACGGATTG
AAGGGTATTCAGCACACTCCAATCTCCATCAAGAACGCTAGAGTTTCCCA
ACACTACAAGGCTTCCTTGACTGCTACTTTCAACTTGCACCCAGACGCTAA
GTTCGCTATCGTTTTGGAAGAGGACTTGGACATTTCCGTTGATTTCTTCTCC
TTCTTGTCCCAGACTATCCACTTGTTGGAAGAGGATGAGTCCTTGTACTGT
ATCTCTGCTTGGAACGACCAAGGTTACGAACACACTGCTGAGGATTCTTCC
TTGTTGTACAGAGTTGAGTCCATGCCAGGATTGGGATGGGTTTTGAGAAA
GAACTTGTACAAGGACGAGTTGGAACCAAAATGGCCAACTCCAGAGAAG
TTGTGGGATTGGGACATGTGGATGAGAATGCCAGAGCAGAGAAAGGACA
GAGAGTGTTTGATTCCAGACGTTTCCAGATCTTACCACTTCGGTATTGTTG
GATTGAACATGAACGGTTACTTCCACGAGGCTTACTTCAAGAAGCACAAG
TTCAACACTGTTCCAAACGTTCAGTTGTCCAACGTTAAGTCCTTGCAGAAG
GACGCTTACGAGATTGAGATCCACAGAATCTTGTCTGAGGCTGAGGTTTT
GGACCATTCCAAGAACCCATGTGAGGATTCCTTCATCCCAGACACAGAGG
GAAAGACTTACATCATGTACATCAAGATGGAACAAGAGGCTGACTTCACT
ACTTGGACTCAGTTGGCTAAGTGTTTGCACATTTGGGACTTGGATGTTAGA
GGTAACCACAAGGGTTTGTGGAGATTGTTCAGAAAGAAGAACCACTTCTT
GGTTGTTGGTTTCCCATTCTCCCCATACGCTGTTAAGAAGCCAGCTTCCGT
TACTCCAATCTACTTGGAGCCACCACCAAAAGAAGAAGCTGCTGTTGCTG
GTATTGACCAGTCCTAGTAGTTAA [SEQ ID NO: 117]

RAVNEEEIDQDYDESLQQADSPRRPANSKKVLDTEIYSSRSKVYIAVDGTTVL
EDEVHEQGRGIHVIVLNQATGHVMAKRVFDTYSPHEDEAMVLFLNMVARGR
ILIFTIKDEGSFHLKDTAKNLLKSLGSQIAPSLGWRDMWTFVVKKGGQVYGE
KHSKSPALSTWGDPILLKTDIQLVPPEDAECHWPDTELNRRRKRFCSKVEGYG
SVCSCKDPTPIEFNPMPLKENKVTTVPVAVIAGNRPNYLYRMLRSLLSAQGVS
PQMITVFIDGYYEEPMDVVELYGLKGIQHTPISIKNARVSQHYKASLTATFNL
HPDAKFAIVLEEDLDISVDFFSFLSQTIHLLEEDESLYCISAWNDQGYEHTAED
SSLLYRVESMPGLGWVLRKNLYKDELEPKWPTPEKLWDWDMWMRMPEQR
KDRECLIPDVSRSYHFGIVGLNMNGYFHEAYFKKHKFNTVPNVQLSNVKSLQ
KDAYEIEIHRILSEAEVLDHKNPCEDSFIPDTEGKTYIMYIKMEQEADFTTWTQ
LAKCLHIWDLDVRGNHKGLWRLFRKKNHFLVVGFPFSPYAVKKPASVTPIYL
EPPPKEEAAVAGIDQS [SEQ ID NO: 118]

EXAMPLE 4

Generation of pGLY579 and pGLY4863

[0136] The vectors that received the TD-POMGnT1 fusions were pGLY579 (places the TD-POMGnT1 fusion under the *P. pastoris* GAP promoter (Cereghino and Cregg, FEMS Microbiol Rev. 24:45-66 (2000)) and pGLY4863 (under the *P. pastoris* AOX1 promoter (Cereghino and Cregg, FEMS Microbiol Rev. 24:45-66 (2000)). pGLY579 is a double-crossover integration vector which contains a *PpHIS3* ORF and, separately, nucleotides located immediately 3' to the *PpHIS3* ORF. The *PpHIS3* ORF (the 5' arm) was generated by PCR using primers PpHIS3 1 (SEQ ID NO:119) and PpHIS3 2 (SEQ ID NO: 120); the resulting DNA is shown (SEQ ID NO: 121). The *PpHIS3* 3' fragment (3' arm) was generated by PCR using primers PpHIS3 3 (SEQ ID NO: 122) and PpHIS3 4 (SEQ ID NO: 123); the resulting DNA is shown (SEQ ID NO: 124). The template was *P. pastoris* genomic DNA from wild-type strain NRRL-Y11430 (from Northern Regional Research Center, Peoria, IL). The PCR fragments were first cloned into pCR2.1 (Invitrogen) and sequenced. The *PpHIS3* integration arms were then sub-cloned successively into pGLY566 using enzymes *Fse*1 and *Sac*I for the 5' arm, and *Swa*I and *Sal*I for the 3' arm to generate pGLY579; *see, e.g.,* WO 07/136865. Situated just downstream of the PpHIS3 5' arm is a DNA fragment encompassing the *PpALG3* transcriptional terminator (TT) sequence, which was cloned by PCR using primers PpALG3TT-f (SEQ ID NO: 125) and PpALG3TT-rev (SEQ ID NO: 126) resulting in the DNA fragment shown (SEQ ID NO: 127). The *PpALG3TT* was sub-cloned using flanking *Fse*1 and *Pme*1 restriction sites. Situated between the *PpHIS3* fragments in pGLY579 are the *URA5* marker (Nett and Gerngross, Yeast 20:1279 (2003), and an expression cassette consisting of the *PpGAP* promoter and *ScCYC1* transcriptional terminator separated by *Not*1 and *Pac*1 restriction sites. The *PpGAP* promoter sequence (SEQ ID NO: 128) and the *ScCYC1* transcriptional terminator sequence (SEQ ID NO: 129) are shown. The *PpGAP* promoter was replaced with the methanol-inducible *PpAOX1* promoter using *Xho*I and *Not*1 sites to generate pGLY4863. The *PpAOX1* promoter sequence is shown (SEQ ID NO: 130).

PpHIS3 1

GAGCTCGGCCACGGTGGCCCTGTGAGTCTGGCT [SEQ ID NO: 119]

PpHIS3 2

GGCCGGCCTCAGAAAAGAACACCCTTCGTACT [SEQ ID NO: 120]

PpHIS3 5' arm

GAGCTCGGCCACGGTGGCCCTGTGAGTCTGGCTCAATCACTTTTCAAAGAT
AAGGACTATTCTGCAGAACATGCAGCCCAGGCAACATCATCCCAGTTCAT
CTCTGTGAACACAGGAATAGGATTCCTGGACCATATGTTACACGCACTTG
CTAAGCACGGCGGCTGGTCTGTCATTATCGAATGTGTAGGTGATTTGCACA
TTGATGACCATCATTCAGCAGAAGATACTGGAATCGCATTGGGGATGGCA
TTCAAAGAAGCCTTGGGCCATGTTCGTGGTATCAAAAGATTCGGGTCCGG
ATTTGCTCCACTAGACGAAGCTCTCAGTCGGGCTGTTATTGATATGTCTAA
CAGGCCCTATGCTGTTGTCGATCTGGGTTTGAAAAGAGAGAAGATTGGAG
ACCTATCGTGTGAGATGATTCCCCATGTTTTGGAAAGTTTTGCCCAAGGAG
CCCATGTAACCATGCACGTAGATTGTTTGCGAGGTTTCAACGACCATCATC
GTGCCGAGAGTGCATTCAAAGCTTTGGCTATAGCTATCAAAGAGGCCATT
TCAAGCAACGGCACGGACGACATTCCAAGTACGAAGGGTGTTCTTTTCTG
AGGCCGGCC [SEQ ID NO: 121]

PpHIS3 3

ATTTAAATGTCTGGAAGGTGTCTACATCTGTGA [SEQ ID NO: 122]

PpHIS3 4

GTCGACGGCCAGTCTGGCCAAGTAATCATTGTCT [SEQ ID NO: 123]

PpHIS3 3' arm

ATTTAAATGTCTGGAAGGTGTCTACATCTGTGAAATCCGTATTTATTTAAG
TAAAACAATCAGTAATATAAGATCTTAGTTGGTTTACCACATAGTCGGTAC
CGGTCGTGTGAACAATAGTTCAATGCCTCCGATTGTGCCTTATTGTTGTGG
TCTGCATTTTCGCGGCGAAATTTCTACTTCAGATCGGGGCTGAGATGACCT
TAGTACTCACATCAACCAGCTCGTTGAAAGTTCCCACATGACCACTCAATG
TTTAATAGCTTGGCACCCATGAGGTTGAAGAAACTACTTAAGGTGTTTTGT
GCCTCAGTAGTGCTGTTAGCGGCGACATCTGTGGTGTTATTTTTCCACTTT
GGAGGTCAGATCATAATCCCCATACCGGAACGCACTGTGACCTTAAGTAC
TCCTCCCGCAAACGATACTTGGCAGTTTCAACAGTTCTTCAACGGCTATTT
AGACGCCCTGTTAGAGAATAACCTGTCGTATCCGATACCAGAAAGGTGGA
ATCATGAAGTTACAAATGTAAGATTCTTCAATCGCATAGGTGAATTGCTCT
CGGAGAGTAGGCTACAGGAGCTGATTCATTTTAGTCCTGAGTTCATAGAG
GATACCAGTGACAAATTCGACAATATTGTTGAACAAATTCCAGCAAAATG
GCCTTACGAAAACATGTACAGAGGAGATGGATACGTTATTGTTGGTGGTG
GCAGACACACCTTTTTGGCACTGCTGAATATCAACGCTTTGAGAAGAGCA

GGCAATAAACTGCCAGTTGAGGTCGTGTTGCCAACTTACGACGACTATGA
GGAAGATTTCTGTGAAAATCATTTTCCACTTTTGAATGCAAGATGCGTAAT
CTTAGAAGAACGATTTGGTGACCAAGTTTATCCCCGGTTACAACTAGGAG
GCTACCAGTTTAAAATATTTGCGATAGCAGCAAGTTCATTCAAAAACTGCT
TTTTGTTAGATTCAGATAATATACCCTTGCGAAAGATGGATAAGATATTCT
CAAGCGAACTATACAAGAATAAGACAATGATTACTTGGCCAGACTGGCCG
TCGAC [SEQ ID NO: 124]

PpALG3TT-f
GGCCGGCCATTTACAATTAGTAATATTAAGGT [SEQ ID NO: 125]

PpALG3TT-rev
GTTTAAACCTACTAAGCGACGAAAACGGGA [SEQ ID NO: 126]

PpALG3 transcription terminator
GGCCGGCCATTTACAATTAGTAATATTAAGGTGGTAAAAACATTCGTAGA
ATTGAAATGAATTAATATAGTATGACAATGGTTCATGTCTATAAATCTCCG
GCTTCGGTACCTTCTCCCCAATTGAATACATTGTCAAAATGAATGGTTGAA
CTATTAGGTTCGCCAGTTTCGTTATTAAGAAAACTGTTAAAATCAAATTCC
ATATCATCGGTTCCAGTGGGAGGACCAGTTCCATCGCCAAAATCCTGTAA
GAATCCATTGTCAGAACCTGTAAAGTCAGTTTGAGATGAAATTTTTCCGGT
CTTTGTTGACTTGGAAGCTTCGTTAAGGTTAGGTGAAACAGTTTGATCAAC
CAGCGGCTCCCGTTTTCGTCGCTTAGTAGGTTTAAAC [SEQ ID NO: 127]

PpGAP promoter
CTCGAGAGATCTTTTTTGTAGAAATGTCTTGGTGTCCTCGTCCAATCAGGT
AGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCGAACGACCTGCTGGC
AACGTAAAATTCTCCGGGGTAAAACTTAAATGTGGAGTAATGGAACCAGA
AACGTCTCTTCCCTTCTCTCTCCTTCCACCGCCCGTTACCGTCCCTAGGAAA
TTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTTC
CCAGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGC
CCAGGGATGGAAAAGTCCCGGCCGTCGCTGGCAATAATAGCGGGCGGAC
GCATGTCATGAGATTATTGGAAACCACCAGAATCGAATATAAAAGGCGAA
CACCTTTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTA
TTTGTCCCTATTTCAATCAATTGAACAACTATCAAAACACAGCGGCCGC
[SEQ ID NO: 128]

ScCYC transcription terminator
TTAATTAAACAGGCCCCTTTTCCTTTGTCGATATCATGTAATTAGTTATGTC
ACGCTTACATTCACGCCCTCCTCCCACATCCGCTCTAACCGAAAAGGAAG
GAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTTAATAGTTAT
GTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTAC
AAACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGT
TTTGGGACGCTCGAAGGCTTTAATTTGCAAGCTGCCGGCTCTTAA
[SEQ ID NO: 129]

PpAOX1 promoter
CTCGAGAGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGC
CATCCGACATCCACAGGTCCATTCTCACACATAAGTGCCAAACGCAACAG
GAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCCACTCCT
CTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGG
CTTGATTGGAGCTCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCA
TGACTTTATTAGCCTGTCTATCCTGGCCCCCCTGGCGAGGTTCATGTTTGTT
TATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCAGAT
GAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGGCCC
AAAACTGACAGTTTAAACGCTGTCTTGGAACCTAATATGACAAAAGCGTG
ATCTCATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTAACGGCCAGT
TGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTA
TTGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCTCT
CTATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGCAAATGGGGAAA
CACCCGCTTTTTGGATGATTATGCATTGTCTCCACATTGTATGCTTCCAAG
ATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAAC
TGTTCTAACCCCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCT
GTCTTAAACCTTTTTTTTTATCATCATTATTAGCTTACTTTCATAATTGCGA
CTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAACGACAACTTG
AGAAGATCAAAAAACAACTAATTATTCGAAACGGCGGCCGC
[SEQ ID NO: 130]

EXAMPLE 5

Generation of *TD-POMGnT1* Library

[0137] Construction of TD-POMGnT1 expression/integration vectors were carried out as described below. The POMGnT1 catalytic domains were isolated from GeneArt vectors following digestion with *Asc*I and *Pac*I and cloned into integration plasmids pGLY579 (for GAPp-driven expression) and pGLY4863 (for AOX1p-

driven expression). The agarose gel purified targeting domain (TD) fragments were then in a high throughput format ligated in frame into these plasmids to create the fusion libraries. Thirty μg of each POMGnT1 catalytic domain-containing vector was digested with 10 units *Asc*I overnight. In the morning another 10 units were added and the incubation was continued for another 1 hr. The linearized DNA was ethanol precipitated and incubated with 50 units *Not*I for 4 hr. Then 20 units calf intestinal alkaline phosphatase were added and the incubation was continued for another 1 hr. After purification by agarose gel extraction, the DNA was diluted to 3 nM and 2.5 μl was loaded into a 96 well PCR plate. Then 2.5 μl of a 12 nM solution of the isolated targeting domain *Not*I/*Asc*I fragments (see above) was added. After addition of 5 μl 2x ligation buffer and 0.5 μl Quick ligase (NEB) the ligation was allowed to proceed for 5 min at room temperature ("RT"), and 2 μl of the ligation mix was added to 50 μl *E. coli* strain DH5α made competent according to the method of Hanahan et al. (Methods Enzymol 204: 63-113 (1991)) that were arranged in a 96 well plate. The mixture was incubated for 20 min on ice, heat shocked at 42° C for 1 min, 200 μl Super Optimal broth with catabolite repression ("SOC") were added to each well of the 96 well plate and the cells were allowed to recover at 37° C for 1 hr. Of each transformation mix 200 μl were then plated on a single LB Amp plate and incubated overnight. We routinely obtained 10 to 1000 colonies per plate as compared to 0 to 10 colonies on the no insert controls. To assess whether the ligation reaction had resulted in in-frame fusions, plasmid DNA was isolated and an *Asc*I restriction digest was performed. This was based on the fact that only if the *Asc*I site had been recreated a genuine in-frame fusion had occurred. This proved to be true in over 99 % of all cases. An example of a TD-POMGnT1 expression/integration vector, pGLY3136, is shown in Figure11.

[0138] Before transformation into yeast strains, the plasmids were digested with *Sfi*I which cuts at sites flanking the HIS3 sequences, thus freeing a DNA fragment containing the promoter-TD-POMGnT1-transcriptional terminator plus URA5 marker flanked by HIS3 5' and 3' integration sequences.

EXAMPLE 6

Generation of Strain GFI 2.0

[0139] The TD-POMGnT1 library was transformed by electroporation (method below) into strain YGLY7877 (GFI 2.0, FIG. 1) which was derived from strain yGLY14 (*och1*Δ::*lacZ*, *bmt2*Δ::*lacZ*/*KlMNN2-2*, *mnn4b*Δ::*lacZ*/ *MmSLC35A3*, *pno1*Δ*mnn4a*Δ::*lacZ*) and constructed using methods described earlier (Nett and Gerngross, Yeast 20:1279 (2003); Choi *et al.*, PNAS USA 100:5022 (2003); Hamilton

et al., Science 301:1244 (2003), and WO 07/136752). Briefly, genes encoding the phosphomannose transferases Pno1, Mnn4a and Mnn4b (Li et al., Nat Biotechnol. 24:210-5 (2006)) and the beta-mannose transferase *BMT2* (Mille et al., J Biol Chem. 283:9724-36 (2008)) were deleted in wild-type strain NRRL-Y11430 to generate YGLY-14 (Strain GFI 1.0, FIG 1). In order to minimize the number of downstream steps in our glyco-engineering process, we included two UDP-GlcNAc transporters into the knock-out vectors. The *K. lactis* UDP-GlcNAc transporter was inserted into the *BMT2* deletion vector and the mouse UDP-GlcNAc transporter inserted into the *MNN4B* deletion vector. This allowed for the stable integration of UDP-GlcNAc transporter genes at sites of gene deletions. In the initial step, YGLY14 was counterselected using 5-Fluoroorotic Acid (5-FOA) (Nett and Gerngross, Yeast 20:1279 (2003) to generate the uracil-auxotrophic strain YGLY-16, which was the recipient of the *Trichoderma reesei* alpha 1,2-mannosidase (Bobrowicz et al., WO 2007/061631) expression vector pGLY1896. The resulting strain YGLY6361 was counterselected again with 5-FOA to generate the uracil-auxotroph YGLY2004. Genes encoding the beta-mannose transferases *BMT1,3,4* (Mille et al., J Biol Chem. 283:9724-36 (2008)) were then deleted using the recyclable URA5 marker (*see* Nett and Gerngross, Yeast 20:1279 (2003) to generate YGLY7827 (GFI 2.0, FIG. 1). YGLY7827 was then transformed with expression vector pGLY3465 (method below), which encodes the TNFRII-Fc reporter protein, to generate YGLY7877. For the above transformation and also those with the TD-POMGnT1 library, transformants were selected on minimal media lacking uracil to select for incorporation of the URA5 marker.

EXAMPLE 7

Generation of TNFRII-Fc Expression Vector pGLY3465

[0140] Expression plasmid vector pGLY3465 contains an expression cassette under the control of the methanol-inducible *P. pastoris* AOX1 promoter that encodes the TNFR-IgG1 fusion protein. The TNFR-IgG1 fragment was codon-optimized by GeneArt with 5' *Pvu*II and 3' *Fse*I cloning sites (SEQ ID NO: 131), and the GeneArt vector designated pGLY3431. The TNFR domain from pGLY3431 was fused to an alternative IgG1 Fc domain from pGLY1477 (also synthesized by GeneArt) by PCR to give a DNA sequence (SEQ ID NO: 132). Specifically primers with SEQ ID NOs: 133 & 134 were used to amplify the TNFR domain from pGLY3431, while primers with SEQ ID NOs: 135 & 136 were used to amplify the IgG1 Fc domain from pGLY1477. Both of these fragments were then fused together by PCR using primers with SEQ ID NOs: 133 & 136. The TNFR-IgG1 was fused at the N-terminus to a

DNA sequence (SEQ ID NO: 137) encoding the human serum albumin ("HSA") pre signal peptide. The DNA encoding the HSA signal sequence (ss) was generated using oligonucleotides purchased from Integrated DNA Technologies (Coralville, IA). The fusion of TNFR-IgG1 to HSAss created a DNA fragment (SEQ ID NO: 138, encoding a protein with SEQ ID NO: 139) with unique 5' *Eco*R1 and 3' *Fse*1 sites. The nucleic acid fragment encoding the HSAss-TNFR-IgG1 fusion protein was subcloned using the 5' *Eco*R1 and 3' *Fse*1 unique sites into an expression plasmid vector pGLY2198, which contains the *P. pastoris TRP2* targeting nucleic acid and the Zeocin-resistance marker and generates expression cassettes under the control of the AOX1 promoter and *Saccharomyces cerevisiae CYC* terminator, to form plasmid pGLY3465. Following transformation of pGLY3465 into *Pichia pastoris*, methanol induction results in the secretion of TNFR-IgG1 with the protein sequence of SEQ ID NO: 140. Transformants were selected on rich media containing Zeocin.

SEQ ID NO: 131 [TNFR-IgG1]
CAGCTGCCAGCTCAAGTTGCTTTTACTCCATACGCTCCAGAACCAGGTTCT
ACTTGTAGATTGAGAGAGTACTACGACCAAACTGCTCAGATGTGTTGTTCC
AAGTGTTCTCCAGGTCAACACGCTAAGGTTTTCTGTACTAAGACTTCCGAC
ACTGTTTGTGACTCTTGTGAGGACTCCACTTACACTCAATTGTGGAACTGG
GTTCCAGAATGTTTGTCCTGTGGTTCCAGATGTTCTTCCGACCAAGTTGAG
ACTCAGGCTTGTACTAGAGAGCAGAACAGAATCTGTACTTGTAGACCTGG
TTGGTACTGTGCTTTGTCCAAGCAAGAGGGTTGTAGATTGTGTGCTCCATT
GAGAAAGTGTAGACCAGGTTTCGGTGTTGCTAGACCAGGTACAGAAACTT
CCGACGTTGTTTGTAAGCCATGTGCTCCAGGAACTTTCTCCAACACTACTT
CCTCCACTGACATCTGTAGACCACACCAAATCTGTAACGTTGTTGCTATCC
CAGGTAACGCTTCTATGGACGCTGTTTGTACTTCTACTTCCCCAACTAGAT
CCATGGCTCCAGGTGCTGTTCATTTGCCACAGCCAGTTTCCACTAGATCCC
AACACACTCAACCAACTCCAGAACCATCTACTGCTCCATCCACTTCCTTTT
TGTTGCCAATGGGACCATCTCCACCTGCTGAAGGTTCTACTGGTGACGAAC
CAAAGTCCTGTGACAAGACTCATACTTGTCCACCATGTCCAGCTCCAGAAT
TGTTGGGTGGTCCATCCGTTTTTTTGTTCCCACCAAAGCCAAAGGACACTT
TGATGATCTCCAGAACTCCAGAGGTTACATGTGTTGTTGTTGACGTTTCTC
ACGAGGACCCAGAGGTTAAGTTCAACTGGTACGTTGACGGTGTTGAAGTT
CACAACGCTAAGACTAAGCCAAGAGAAGAGCAGTACAACTCCACATACA
GAGTTGTTTCCGTTTTGACTGTTTTGCACCAGGATTGGTTGAACGGAAAGG
ACTACAAGTGTAAGGTTTCCAACAAGGCTTTGCCAGCTCCAATGCAAAAG
ACTATCTCCAAGGCTAAGGGTCAACCAAGAGAGCCACAGGTTTACACTTT

GCCACCATCCAGAGATGAGTTGACTAAGAATCAGGTTTCCTTGACTTGTTT
GGTTAAGGGATTCTACCCAAGACACATCGCTGTTGAATGGGAATCTAACG
GACAGCCAGAGAACAACTACAAGACTACTCCACCAGTTTTGGACTCTGAC
GGTTCCTTCTTCTTGTACTCCAAGTTGACTGTTGACAAGTCCAGATGGCAA
CAGGGTAACGTTTTCTCCTGTTCCGTTATGCATGAGGCTTTGCACAACCAC
TACACTCAAAAGTCCTTGTCTTTGTCCCCTGGTAAGTAGGGCCGGCC

SEQ ID NO: 132 [TNFR-Fc]
CAGCTGCCAGCTCAAGTTGCTTTTACTCCATACGCTCCAGAACCAGGTTCT
ACTTGTAGATTGAGAGAGTACTACGACCAAACTGCTCAGATGTGTTGTTCC
AAGTGTTCTCCAGGTCAACACGCTAAGGTTTTCTGTACTAAGACTTCCGAC
ACTGTTTGTGACTCTTGTGAGGACTCCACTTACACTCAATTGTGGAACTGG
GTTCCAGAATGTTTGTCCTGTGGTTCCAGATGTTCTTCCGACCAAGTTGAG
ACTCAGGCTTGTACTAGAGAGCAGAACAGAATCTGTACTTGTAGACCTGG
TTGGTACTGTGCTTTGTCCAAGCAAGAGGGTTGTAGATTGTGTGCTCCATT
GAGAAAGTGTAGACCAGGTTTCGGTGTTGCTAGACCAGGTACAGAAACTT
CCGACGTTGTTTGTAAGCCATGTGCTCCAGGAACTTTCTCCAACACTACTT
CCTCCACTGACATCTGTAGACCACACCAAATCTGTAACGTTGTTGCTATCC
CAGGTAACGCTTCTATGGACGCTGTTTGTACTTCTACTTCCCCAACTAGAT
CCATGGCTCCAGGTGCTGTTCATTTGCCACAGCCAGTTTCCACTAGATCCC
AACACACTCAACCAACTCCAGAACCATCTACTGCTCCATCCACTTCCTTTT
TGTTGCCAATGGGACCATCTCCACCTGCTGAAGGTTCTACTGGTGACGAGC
CAAAGTCCTGTGACAAGACACATACTTGTCCACCATGTCCAGCTCCAGAA
TTGTTGGGTGGTCCATCCGTTTTCTTGTTCCCACCAAAGCCAAAGGACACT
TTGATGATCTCCAGAACTCCAGAGGTTACATGTGTTGTTGTTGACGTTTCT
CACGAGGACCCAGAGGTTAAGTTCAACTGGTACGTTGACGGTGTTGAAGT
TCACAACGCTAAGACTAAGCCAAGAGAAGAGCAGTACAACTCCACTTACA
GAGTTGTTTCCGTTTTGACTGTTTTGCACCAGGATTGGTTGAACGGTAAAG
AATACAAGTGTAAGGTTTCCAACAAGGCTTTGCCAGCTCCAATCGAAAAG
ACAATCTCCAAGGCTAAGGGTCAACCAAGAGAGCCACAGGTTTACACTTT
GCCACCATCCAGAGAAGAGATGACTAAGAACCAGGTTTCCTTGACTTGTT
TGGTTAAAGGATTCTACCCATCCGACATTGCTGTTGAATGGGAATCTAACG
GTCAACCAGAGAACAACTACAAGACTACTCCACCAGTTTTGGATTCTGAC
GGTTCCTTCTTCTTGTACTCCAAGTTGACTGTTGACAAGTCCAGATGGCAA
CAGGGTAACGTTTTCTCCTGTTCCGTTATGCATGAGGCTTTGCACAACCAC
TACACTCAAAAGTCCTTGTCTTTGTCCCCAGGTAAGTAGGGCCGGCC

SEQ ID NO: 133
GGGTACCCAGCTGCCAGCTCAAGTTGCTTTTACTCCATAC

SEQ ID NO: 134
GTGTCTTGTCACAGGACTTTGGCTCGTCACCAGTAGAACCTTCAGCAGGTG
GAGATG

SEQ ID NO: 135
CATCTCCACCTGCTGAAGGTTCTACTGGTGACGAGCCAAAGTCCTGTGAC
AAGACAC

SEQ ID NO: 136
GGAGCTCGGCCGGCCCTACTTACCTGGGGACAAAGACAAGGACTTTTG

SEQ ID NO: 137 [HSA pre-signal sequence]
GAATTCGAAACGATGAAGTGGGTTACCTTTATCTCTTTGTTGTTTCTTTTCT
CTTCTGCTTACTCT

SEQ ID NO: 138 [HSAss-TNFR-Fc DNA sequence]
GAATTCGAAACGATGAAGTGGGTTACCTTTATCTCTTTGTTGTTTCTTTTCT
CTTCTGCTTACTCTCTGCCAGCTCAAGTTGCTTTTACTCCATACGCTCCAGA
ACCAGGTTCTACTTGTAGATTGAGAGAGTACTACGACCAAACTGCTCAGA
TGTGTTGTTCCAAGTGTTCTCCAGGTCAACACGCTAAGGTTTTCTGTACTA
AGACTTCCGACACTGTTTGTGACTCTTGTGAGGACTCCACTTACACTCAAT
TGTGGAACTGGGTTCCAGAATGTTTGTCCTGTGGTTCCAGATGTTCTTCCG
ACCAAGTTGAGACTCAGGCTTGTACTAGAGAGCAGAACAGAATCTGTACT
TGTAGACCTGGTTGGTACTGTGCTTTGTCCAAGCAAGAGGGTTGTAGATTG
TGTGCTCCATTGAGAAAGTGTAGACCAGGTTTCGGTGTTGCTAGACCAGG
TACAGAAACTTCCGACGTTGTTTGTAAGCCATGTGCTCCAGGAACTTTCTC
CAACACTACTTCCTCCACTGACATCTGTAGACCACACCAAATCTGTAACGT
TGTTGCTATCCCAGGTAACGCTTCTATGGACGCTGTTTGTACTTCTACTTCC
CCAACTAGATCCATGGCTCCAGGTGCTGTTCATTTGCCACAGCCAGTTTCC
ACTAGATCCCAACACACTCAACCAACTCCAGAACCATCTACTGCTCCATCC
ACTTCCTTTTTGTTGCCAATGGGACCATCTCCACCTGCTGAAGGTTCTACT
GGTGACGAGCCAAAGTCCTGTGACAAGACACATACTTGTCCACCATGTCC
AGCTCCAGAATTGTTGGGTGGTCCATCCGTTTTCTTGTTCCCACCAAAGCC
AAAGGACACTTTGATGATCTCCAGAACTCCAGAGGTTACATGTGTTGTTGT

TGACGTTTCTCACGAGGACCCAGAGGTTAAGTTCAACTGGTACGTTGACG
GTGTTGAAGTTCACAACGCTAAGACTAAGCCAAGAGAAGAGCAGTACAA
CTCCACTTACAGAGTTGTTTCCGTTTTGACTGTTTTGCACCAGGATTGGTTG
AACGGTAAAGAATACAAGTGTAAGGTTTCCAACAAGGCTTTGCCAGCTCC
AATCGAAAAGACAATCTCCAAGGCTAAGGGTCAACCAAGAGAGCCACAG
GTTTACACTTTGCCACCATCCAGAGAAGAGATGACTAAGAACCAGGTTTC
CTTGACTTGTTTGGTTAAAGGATTCTACCCATCCGACATTGCTGTTGAATG
GGAATCTAACGGTCAACCAGAGAACAACTACAAGACTACTCCACCAGTTT
TGGATTCTGACGGTTCCTTCTTCTTGTACTCCAAGTTGACTGTTGACAAGT
CCAGATGGCAACAGGGTAACGTTTTCTCCTGTTCCGTTATGCATGAGGCTT
TGCACAACCACTACACTCAAAAGTCCTTGTCTTTGTCCCCAGGTAAGTAGG
GCCGGCC

SEQ ID NO: 139 [HSAss-TNFR-Fc peptide sequence]

M K W V T F I S L L F L F S S A Y S L P A Q V A F T P Y A P
E P G S T C R L R E Y Y D Q T A Q M C C S K C S P G Q H A K
V F C T K T S D T V C D S C E D S T Y T Q L W N W V P E C L
S C G S R C S S D Q V E T Q A C T R E Q N R I C T C R P G W
Y C A L S K Q E G C R L C A P L R K C R P G F G V A R P G T
E T S D V V C K P C A P G T F S N T T S S T D I C R P H Q I C
N V V A I P G N A S M D A V C T S T S P T R S M A P G A V H
L P Q P V S T R S Q H T Q P T P E P S T A P S T S F L L P M G
P S P P A E G S T G D E P K S C D K T H T C P P C P A P E L L
G G P S V F L F P P K P K D T L M I S R T P E V T C V V V D
V S H E D P E V K F N W Y V D G V E V H N A K T K P R E E
Q Y N S T Y R V V S V L T V L H Q D W L N G K E Y K C K V
S N K A L P A P I E K T I S K A K G Q P R E P Q V Y T L P P S
R E E M T K N Q V S L T C L V K G F Y P S D I A V E W E S N
G Q P E N N Y K T T P P V L D S D G S F F L Y S K L T V D K
S R W Q Q G N V F S C S V M H E A L H N H Y T Q K S L S L
S P G K

SEQ ID NO: 140 [secreted TNFR-Fc peptide]

L P A Q V A F T P Y A P E P G S T C R L R E Y Y D Q T A Q M
C C S K C S P G Q H A K V F C T K T S D T V C D S C E D S T
Y T Q L W N W V P E C L S C G S R C S S D Q V E T Q A C T
R E Q N R I C T C R P G W Y C A L S K Q E G C R L C A P L R

K C R P G F G V A R P G T E T S D V V C K P C A P G T F S N
T T S S T D I C R P H Q I C N V V A I P G N A S M D A V C T
S T S P T R S M A P G A V H L P Q P V S T R S Q H T Q P T P
E P S T A P S T S F L L P M G P S P P A E G S T G D E P K S C
D K T H T C P P C P A P E L L G G P S V F L F P P K P K D T L
M I S R T P E V T C V V V D V S H E D P E V K F N W Y V D
G V E V H N A K T K P R E E Q Y N S T Y R V V S V L T V L
H Q D W L N G K E Y K C K V S N K A L P A P I E K T I S K A
K G Q P R E P Q V Y T L P P S R E E M T K N Q V S L T C L V
K G F Y P S D I A V E W E S N G Q P E N N Y K T T P P V L D
S D G S F F L Y S K L T V D K S R W Q Q G N V F S C S V M H
E A L H N H Y T Q K S L S L S P G K

EXAMPLE 8

Yeast Transformation

[0141] All yeast transformations were as follows. *Pichia pastoris* strains were grown in 50 mL YPD media (yeast extract (1%), peptone (2%), dextrose (2%)) overnight to an optical density ("OD") of between about 0.2 to 6. After incubation on ice for 30 minutes, cells were pelleted by centrifugation at 2500-3000 rpm for 5 minutes. Media was removed and the cells washed three times with ice cold sterile 1M sorbitol before resuspension in 0.5 ml ice cold sterile 1M sorbitol. Ten μL linearized DNA (5-20 μg) and 100 μL cell suspension was combined in an electroporation cuvette and incubated for 5 minutes on ice. Electroporation was in a Bio-Rad GenePulser Xcell following the preset *Pichia pastoris* protocol (2 kV, 25 μF, 200 Ω), immediately followed by the addition of 1 mL YPDS recovery media (YPD media plus 1 M sorbitol). The transformed cells were allowed to recover for four hours to overnight at room temperature (26°C) before plating the cells on selective media.

EXAMPLE 9

Isolation of Positive Transformants

[0142] PCR was used to confirm the double crossover integration of TD-POMGnT1 fusions using PCR primers matching *URA5* (SEQ ID NO: 141), *PpALG3TT* (SEQ ID NO: 142), and *PpHIS3* sequences 5' (SEQ ID NO: 143) and 3' (SEQ ID NO:144) to the integration arms. The PCR conditions were one cycle of 98°C for 2 minutes, 30 cycles of 98°C for 10 seconds, 50°C for 30 seconds, and 72°C for 1 minute, and followed by one cycle of 72°C for 7 minutes. The PCR products were analyzed by agarose gel electrophoresis following standard methods.

SEQ ID NO: 141: GGGAGAGTTGAAGGTTGTATTATTGCC
SEQ ID NO: 142: CTACTAAGCGACGAAAACGGGAGCCG
SEQ ID NO: 143: GTTCCCTCATTAAGAGGATCACAAACG
SEQ ID NO: 144: GATAATAGTGCGGGCTGGTACTTCG

EXAMPLE 10

Screening of TD-POMGnT1 Fusions by Lectin GS-II Staining

[0143] The TD-POMGnT1 fusions were screened in a 96-well format using staining by the alkaline phosphate conjugated GS-II lectin (from *Griffonia simplicifolia*, EY Labs) which binds terminal GlcNAc residues on glycans. Transformants of the TD-POMGnT1 library in strain YGLY 7877 (GFI 2.0) were grown in 0.6mL BMGY media (1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer pH 6.0, 1.34% yeast nitrogen base, 4 x 10-5 % biotin, and 1% glycerol) for 48 h with strong shaking, pelleted and washed 1X in BMMY media (same as BMGY except 1% methanol replaced the glycerol), and then grown for 48h in 0.2 mL BMMY. Supernatants were then harvested by centrifugation, and tested for levels of terminal GlcNAc on the TNFRII-Fc reporter protein by GS-II staining in an ELISA format.

EXAMPLE 11

96-Well Lectin Staining Assay

[0144] 96-well plates were coated with 100µl per well of 1µg/ml monoclonal Ab (anti-hs TNF RII from R& D Systems: MAB726) in coating buffer (Virolabs: RE-COAT-7-100 pH7.2). After incubation at room temperature for one hour, plates were washed 3X using 100µl of wash buffer (Virolabs: RE-WASH-T-100) per well. Wells were blocked using 100µl per well of 1x Tris ELISA Diluent/Blocking buffer (Virolabs: RE-DILU-T-100) followed by incubation at room temperature for one hour, and then washing as before. 100µl of sample diluted 1:10 were added to each well, followed by incubation at room temperature for one hour, then plates washed as before. GS-II lectin (from *Griffonia Simplicifolia,* EY Laboratories Inc., LA-2402) solution was prepared in 1x Tris ELISA diluent/blocking buffer at a dilution of 1:1000. 100µl of this lectin solution were added to each well of the plate, followed by incubation at room temperature for one hour. Plates were washed as before. 100µl of SuperPhos™ 4-MUP Solution (Virolabs: X-PHOS-100) were added to each well, followed by incubation at room temperature in the dark for 45 minutes. Plates were read using 360nm excitation and 450nm emission wavelengths on a Tecan (Mannedorf, Switzerland) Genius Pro plate reader with Magellan software.

[0145] Data from the library where TD-POMGnT1 expression was driven by the PpAOX1 promoter did not necessarily correlate with that of the library having PpGAP promoter-driven expression. Some of the AOX1p-driven TD-POMGnT1 fusions that gave strong GS-II binding gave weak binding with the PpGAP promoter. Additionally, the most active TD-POMGnT1 fusions, when linked to the PpGAP promoter, were toxic to the cells and/or inhibited protein secretion. These included human, mouse, or frog POMGnT1 fused to ScMNN2-s and –m, ScPMT5-m and PpPMT1-m. Similar findings have been made with other enzymes, e.g., MNS1 [Choi et al., PNAS 100:5022 (2003)]. The skilled artisan, with the benefit of the present disclosure, will be able to assess the best combination for the desired outcome and employ same as described herein. As in the present application, the outcome to be assessed is the level of O-linked GlcNAc (as determined by lectin GS-II staining), expression levels of the particular protein of interest and/or cell growth of the production cell. Where the glycoform is O-linked GlcNAc-Gal or O-linked GlcNAc-Gal-Sia, ECA or SNA1, respectively, should be used in the place of lectin GS-II in the above lectin-based assay

[0146] We used the data obtained for the AOX1p-TD-POMGnT1 library for the initial analysis. We determined that the chicken POMGnT1 catalytic domains were inactive and zebrafish POMGnT1 displayed very weak activity. Results for the AOXp-TD-POMGnT1 library screening are shown in Table 2. The data indicated that the most active POMGnT1 catalytic domains were from human, mouse and frog, and the most active TDs were ScMNS1-s, ScMNN9-s, PpKre2-s, KlGnt1-s, ScMNN2-s, ScMNN2-m, ScMNN5-s, ScMNN6-s, ScPMT5-m, PpPMT1-m, and PpBET1-s.

Table 2. Lectin staining of POMGnT1-leader fusions.

		Hs	Mm	Xen			Hs	Mm	Xen
1	ScGLS1-s	2	3	3	**28**	ScPMT2-s	1	1	1
2	ScMNS1-s	2	3	5	**29**	ScPMT3-s	1	1	1
3	ScMNS1-m	2	0	0	**30**	ScPMT4-s	1	1	1
4	ScSEC12-m	3	3	4	**31**	ScPMT5-s	1	1	1
5	PpSEC12-s	4	3	4	**32**	ScPMT6-s	1	1	1
6	PpOCH1-s	1	2	1	**33**	ScPMT1-m	1	0	0
7	ScMNN9-s	4	4	5	**34**	ScPMT2-m	1	1	1
8	ScVAN1-m	1	0	3	**35**	ScPMT3-m	2	3	1
9	ScANP1-s	1	3	4	**36**	ScPMT4-m	3	3	1
10	ScHOC1-s	2	1	4	**37**	ScPMT5-m	5	4	4
11	ScMNN10-s	3	2	4	**38**	ScPMT6-m	3	4	2
12	ScMNN11-s	2	0	4	**39**	PpPMT1-s	2	4	2
13	ScKRE2-s	2	0	3	**40**	PpPMT2-s	2	4	2
14	ScKRE2-l	2	2	3	**41**	PpPMT4-s	2	4	2
15	PpKTR1-s	2	1	2	**42**	PpPMT5-s	2	2	2
16	PpKTR3-s	2	1	2	**43**	PpPMT6-s	2	2	2
17	PpKRE2-s	3	4	5	**44**	PpPMT1-m	5	2	4
18	ScKTR1-s	2	3	3	**45**	PpPMT2-m	2	2	2
19	ScKTR2-s	2	4	4	**46**	PpPMT4-m	2	0	2
20	KlGNT1-s	4	4	5	**47**	PpPMT5-m	2	0	2
21	ScMNN2-s	5	4	5	**48**	PpPMT6-m	2	0	N/A
22	ScMNN2-m	5	5	4	**49**	ScBOS1-s	2	0	1
23	ScMNN5-s	2	5	2	**50**	ScBET1-s	2	1	1
24	ScYUR1-s	1	1	1	**51**	ScSEC22-s	2	0	N/A
25	ScMNN1-s	2	1	4	**52**	PpBOS1-s	2	1	2
26	ScMNN6-s	3	5	4	**53**	PpBET1-s	5	5	5
27	ScPMT1-s	1	1	1	**54**	PpSEC22-s	2	N/A	1

EXAMPLE 12

O-Glycan Analysis by HPAEC-PAD

[0147] We next analyzed the O-glycans in GFI SO-1 strains (*GFI 2.0 strains further expressing POMGnT1; see, e.g.,* WO 07/136752) expressing the most active TD-POMGnT1 fusions. Strains showing strong GS-II staining were grown in shake flasks containing 100mL of BMGY for 48 h, pelleted and washed 1X with BMMY, and then grown an additional 48h in 50 mL BMMY prior to harvest by centrifugation. Secreted TNFRII-Fc was purified from cleared supernatants using protein A chromatography (Li et al. Nat. Biotechnol. 24(2):210-5 (2006)), and the O-glycans released from and separated from protein by alkaline elimination (beta-elimination) (Harvey, Mass Spectrometry Reviews 18: 349- 451 (1999), Stadheim et al., Nat. Protoc. 3:1026-31 (2006)). This process also reduces the newly formed reducing terminus of the released O-glycan (either oligomannose or mannose) to mannitol.

The mannitol group thus serves as a unique indicator of each O- glycan. 0.5 nmole or more of protein, contained within a volume of 100 µL PBS buffer, was required for beta elimination. The sample was treated with 25 µL alkaline borohydride reagent and incubated at 50°C for 16 hours. About 20 µL arabitol internal standard was added, followed by 10 µL glacial acetic acid. The sample was then centrifuged through a Millipore filter containing both SEPABEADS and AG 50W-X8 resin and washed with water. The samples, including wash, were transferred to plastic autosampler vials and evaporated to dryness in a centrifugal evaporator. 150 µL 1% AcOH/MeOH was added to the samples and the samples evaporated to dryness in a centrifugal evaporator. This last step was repeated five more times. 200 µL of water was added and 100 µL of the sample was analyzed by high pH anion-exchange chromatography coupled with pulsed electrochemical detection- HPLC (HPAEC-PAD) according to the manufacturer (Dionex, Sunnyvale, CA).

[0148] Results: HPAEC-PAD traces for TNFRII-Fc glycans from strain YGLY7879 which harbors PpGAPp-driven ScMNN6s-mousePOMGnT1 fusion is shown in FIG 2, PANEL B. Comparison is to GFI 2.0 Strain YGLY6428 (PANEL A) which lacks POMGnT1. Results indicated a POMGnT1-dependent O-glycan in YGLY7879 that co-migrated with the arabitol standard. Arabitol was added to the samples at a concentration that typically gave a maximum reading of ~100 nano coulombs (nC) in strains lacking POMGnT1, but up to 450nC in GFI SO-1 strains that harbor POMGnT1. Shown in FIG 3 are HPAEC-PAD traces of glycans from strain YGLY7879 without (PANEL A) and with (PANEL B) hexosaminidase treatment that removes terminal GlcNAc residues. Hexosaminidase (β-N-acetyl-hexosaminidase, NE Biolabs, Ipswich, MA) treatment was as described by the manufacturer. Briefly, O-glycans released by beta-elimination were buffer-exchanged into 50mM NaCitrate, pH 6.0, and incubated at 37C overnight. The treatment results in the reduction of the putative arabitol + Man-GlcNAc peak, appearance of a new peak at T=23.909 which is free GlcNAc, and an increase in the manitol (man1) peak. This verifies that the sugar co-migrating with arabitol in POMGnT1-containing strains is Man-GlcNAc.

[0149] Figures 4 and 5 show generation of strains that efficiently transfer galactose to the O-linked Man-GlcNAc. The most active TD-POMGnT1 fusions were screened in a 96-well format using staining by the alkaline phosphate conjugated ECA lectin (from *Erythina cristagalli*, EY Labs LA-5901) which binds terminal galactose residues on glycans. The screening was done exactly as described above for lectin GS-II, except that the TD-POMGnT1 vectors were transformed into strain YGLY7875 which contains enzymes required for galactose synthesis and transfer. YGLY7875 was constructed by transforming YGLY5858 with the TNFRII-Fc

expression vector pGLY3465 as described above. Generation of YGLY5858 was as described for YGLY1703 (*see, e.g.,* WO 07/136752) except that it is a uracil auxotroph altered above (*see* Bobrowicz et al., 2004. *Glycobiology* 14:757-66. Li *et al.*, 2006 *Nat. Biotech* 24:210-215, Gerngross *et al.* WO 04/074499, and WO 07/136752). One of the most active strains identified by lectin screening was YGLY7880 which harbors the GAPp-ScMNN2s-humanPOMGnT1 fusion. YGLY7880 was cultured in shakeflasks and supernatant subjected to protein A purification as described for Figures 2 and 3. Figure 4 shows HPAEC-PAD traces of TNFRII-Fc O-glycans from GFI 2.0 strain YGLY6428 (PANEL A) which has no POMGnT1, and GFI SO-2 strain YGLY7880 (+ POMGnT1) (PANEL B). Results indicate a novel POMGnT1-dependent O-glycan in YGLY7880 migrating at ~T=21.38 that is Man-GlcNAc-Gal. Figure 5 shows HPAEC-PAD traces of TNFRII-Fc O-glycans from GFI SO-2 strainYGLY7880 untreated (PANEL A), or treated with galactosidase to remove terminal galactose plus hexosaminidase to remove terminal GlcNAc (PANEL B). Hexosamindase (described above) and galactosidase (β1,4-galactosidase, Calbiochem/EMD Biosciences, La Jolla, CA) treatment was as described above for hexosaminidase. The treatment results in the elimination of the putative Man-GlcNAc-Gal peak, appearance of a peak for free GlcNAc at T=23.909, an increase in the manitol (man1) peak, and the appearance of a galactose peak at T=38.35. These results indicate that the POMGnT1-dependent O-glycan in GFI 5.0 strains that migrates at ~T=21.38 is Man-GlcNAc-Gal.

[0150] Figure 6 shows TNFRII-Fc O-glycans in GFI SO-3 strain YGLY8750 which is GFI SO-2 strain YGLY7880 transformed with vector pGLY1758 that harbors the five genes (Figure 1) required for generating sialylated glycans (Hamilton et al., Science 213:1441 (2006)). Shown are HPAEC-PAD traces of TNFRII-Fc O-glycans from YGLY8750 untreated (PANEL A), or treated with neuraminidase to remove terminal sialic acid, galactosidase to remove terminal galactose, plus hexosaminidase to remove terminal GlcNAc (PANEL B). Hexosaminidase (described above), galactosidase (described above) and neuraminidase (NE Biolabs, Ipswich, MA) treatment was as described above for hexosaminidase. Sub-optimal doses of enzymes were added in order to generate all possible species. The treatment results in the reduction of the putative Man-GlcNAc-Gal-Sialic acid peak at T=45.8, the reduction of the Man-GlcNAc-Gal peak at T=21.7, appearance of a peak for free GlcNAc at T=23.5, an increase in the manitol (man1) peak, and the appearance of a galactose peak at T=37.9. These results indicate that the POMGnT1-dependent O-glycan in GFI SO-3 strains that migrates at ~T=45.8 is Man-GlcNAc-Gal-sialic acid.

[0151] Figure 7 shows Western blotting with alkaline phosphatase conjugated lectin SNA-1 to visualize high levels of terminal sialic acid on TNFRII-Fc. Shown is a coomassie stained SDS-PAGE gel (bottom) showing TNFRII-Fc protein levels in samples 1-12, and a Western blot (top) using detection by alkaline phosphatase-conjugated lectin SNA-1. Lectin SNA-1 binds to terminal sialic acid, and thus will detect terminal sialic acid on O-glycans in Strain GFI SO-3 that harbors genes for sialic acid synthesis and transfer. Results indicate strong lectin binding to TNFRII-Fc from samples 1-9 which are from GFI SO-3 strains, but no binding to TNFRII-Fc from samples 10-12 which are from GFI SO-2 strains which lack the ability to transfer sialic acid.

EXAMPLE 13

Western Blotting with Alkaline Phosphatase Conjugated Lectin SNA-1

[0152] Three separate colonies of YGLY7880 transformed with sialylation vector pGLY1758 were grown in BMGY and induced in BMMY as described above. Two, six, and ten μL of supernatants were separated by reducing polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli, U. K. (1970) Nature 227, 680–685 and then electroblotted onto nitrocellulose membranes (Schleicher & Schuell, now Whatman, Inc., Florham Park, NJ). The membrane was incubated at room temperature for 30 min. in blocking solution (Roche DIG Glycan Differentiation Kit 1-210-238) that was diluted in TBS (0.05 M Tris-HCl pH 7.5, 0.15 M NaCl), followed by three 5 min. washes in Lectin Binding Buffer (50 mM Tris-HCl pH 7.5, 0.15 M NaCl, 0.1 mM CaCl2, 0.1% Tween-20). Twenty micrograms/mL of lectin SNA-1 (from *Sambucus Nigra*, EY Labs. LA-6802) was diluted in Lectin Binding Buffer (total volume 2.5 mL) and added to the membrane at room temperature for 1 h. Following three 10 min. washes in Lectin Binding Buffer, NBT/BCIP (Roche DIG Glycan Differentiation Kit 1-210-238) in 10 mM Tris pH 9.5 was added until bands were visible.

[0153] Figure 8 shows results of efforts to optimize generation of O-sialylated glycans. We transformed our set of most active TD-POMGnT1 strains into TNFRII-Fc expressing strain YGLY11731 which contains all genes needed for sialylation but lacks POMGnT1. In addition, YGLY11731 contains a plasmid encoding an extra copy of each of the five genes required for sialylation (Hamilton et al., Science 213:1441 (2006)). POMGnT1 transformants were screened using lectin SNA-1 in the 96-well lectin staining assay described above. Colonies that generated the strongest SNA-1 staining were selected for shake flask expression studies. Growth and methanol induction was as described above, except that a synthetic inhibitor of

protein O-mannosyltransferase (Pmt) (Orchard et al, EP 1 313 471 B1) was added during the induction phase as described (Bobrowicz et al., WO2007/061631). Briefly, inhibitor PMTi-3 (Bobrowicz et al., WO2007/061631) was added to a final concentration of 2μM with BMMY at the start of the induction phase. The Pmt inhibitor reduces the transfer of mannose to serines and threonines, and thus reduces the overall levels of O-linked mannosylation. For TNFRII-Fc, the Pmt inhibitor lowers the number of serines and threonines with O-mannose glycans from ~80 to ~20 per protein molecule. Thus the overall level of O-linked glycans is significantly lowered by the drug. Importantly, this results in a higher percentage of the desired sialylated O-glycan compared to more undesired O-Man-Man (O-Man2) or asialylated O-man-GlcNAc or O-Man-GlcNAc-Gal. Shown in Figure 8 is the HPAEC-PAD trace (PANEL A) for TNFRII-Fc O-glycans from optimized GFI SO-3 strain YGLY11603 (GAPp-ScMNN6s-mouse POMGnT1). PANEL B lists percentages of each major O-glycan species which were quantified from the HPAEC-PAD trace as described by the manufacturer (Dionex, Sunnyvale, CA). Results show that the predominant O-glycan is Man-GlcNAc-Gal-Sialic acid, which makes up 56% of the TNFRII-Fc O-glycans.

EXAMPLE 14

Total Sialic Acid Determination

[0154] We next used an assay that detects total sialic acid content on glycoproteins as a ratio of moles sialic acid/mole protein. Sialic acid was released from glycoprotein samples by acid hydrolysis and analyzed by HPAEC-PAD using the following method: 10-15 μg of protein sample were buffer-exchanged into phosphate buffered saline. Four hundred μL of 0.1M hydrochloric acid was added, and the sample heated at 80°C for 1 h. After drying in a SpeedVac, the samples were reconstituted with 500 μL of water. One hundred uL was then subjected to HPAEC-PAD analysis. Total sialic content of TNFRII-Fc protein from GFI SO-3 strain YGLY11603 was compared to that from Chinese hamster ovary (CHO) cell produced TNFRII-Fc and also to that from the parent stain YGLY11731 which lacks POMGnT1. The total sialic acid on TNFRII-Fc is from both N- and O-glycans, and thus TNFRII-Fc sialic acid for CHO-produced and strain YGLY11603 comes from both N-and O-glycans, while that from strain YGLY11731 comes completely from N-glycans. TNFRII-Fc has six potential N-glycan sites per dimer. Results are shown in Table 3. The data indicate that roughly similar levels of sialylated O-glycans were generated as that produced in CHO cells.

Table 3. Total sialic acid content.

Sample	Sialic acid (mol Sial/mol protein)
MK-TNFR from CHO	29
MK-TNFR from *P. pastoris* YGLY11731	4
MK-TNFR from *P. pastoris* YGLY11603	22

EXAMPLE 15

Bioavailability & Serum Half-Life of Modified Glycoproteins

[0155] TNFRII-Fc from strain YGLY14252. YGLY14252 was constructed as described above for YGLY11603, except that it was selected for higher TNFRII-Fc expression. This was accomplished by extensive screening of strains using the 96-well lectin SNA-1 staining assay described above. The strains identified by strong SNA-1 staining were further screened by HPAEC-PAD analysis following growth in 50mL shakeflasks. YGLY14252 was chosen based on having the highest percentage of sialylated O-glycans and highest total sialic acid levels measured as described above. Partially purified TNFRII-Fc from YGLY14252 was separated by hydroxyapatite chromatography. Wash and bound fractions were collected and total sialic acid content ("TSA") as mol sialic acid/mol TNFRII-Fc was determined. For pooled wash fractions (Form 5-A), TSA=21, and for eluted bound fractions (Form 5-C), TSA=11. In addition, Forms 5-A and 5-C were mixed in equal ratios to generate Form 5-B which had an intermediate TSA=16.

[0156] For the bioavailability studies, B6 mice were dosed subcutaneously with glycovariants. Serum samples were collected 48 hours later. TNFRII-Fc concentration was determined with a Gyro-based Immunoassay (anti-TNFRII capture and anti-hFc detection). Figure 9 illustrates how increased O-sialylation of a glycoprotein results in increased bioavailability in B6 mice. Serum concentrations of mice 48 hours after subcutaneous administration of different doses of TNFRII-Fc are shown. Total sialic acid content (mol/mol) was as follows: Form 5-A: 21; Form 5-B: 16; and Form 5-C: 11.

[0157] For the serum half-life studies, rats were dosed subcutaneously or intravenously with glycovariants. Serum samples were collected at time intervals post-injection. TNFRII-Fc concentration was determined with a Gyro-based Immunoassay (anti-TNFRII capture and anti-hFc detection). Figures 10A-B illustrate how increased O-sialylation of a glycoprotein results in increased serum half-life in rats. Serum time-concentration curves following intravenous ("IV") [Figure 10A] and subcutaneous ("SC") [Figure 10B] administrations at 1 mg/kg are shown. Total sialic

acid content (mol/mol) was as follows: Form 5-A: 21; Form 5-B: 16; and Form 5-C: 11.

[0158] Forms A-C were generated by separating TNFRII-Fc by hydroxyapatite (HA) chromatography. Form A represents fractions eluting from the HA column with the wash, and Form C represents the bound fractions. Form B is a 1:1 mix of Forms A and C.

CLAIMS:

1. A method of producing a recombinant glycoprotein having O-Man-GlcNAc; O-Man-GlcNAc-Gal; or O-Man-GlcNAc-Gal-Sia as the predominant O-glycan thereof in a lower eukaryotic host cell which comprises:

(a) providing lower eukaryotic host cells which

(i) do not express functional beta-mannosyltransferase enzymes Bmt 1, 2, 3 and 4, and phospho-mannose transferase enzymes Mnn4a, Mnn4b and Pno1; and

(ii) do express functional protein O-linked mannose β-1,2-N-linked acetylglucosaminyltransferase 1 ("POMGnT1"), UDP-GlcNAc transporter and α-1,2-mannosidase enzymes;

(b) transfecting or transforming the lower eukaryotic host cells with nucleic acid encoding the glycoprotein of interest; and

(c) culturing the host cells so that the recombinant glycoprotein is produced.

2. The method of claim 1 wherein the cells of step (a) further do not express Ktr1.

3. The method of claim 1 wherein the enzymes of step (a)(ii) further include one of the following combinations of enzymes:

(a) β1,4GalT, UDP-Gal transporter and UDP-Gal epimerase;

(b) β1,4GalT, UDP-Gal transporter, UDP-Gal epimerase; α2,6SialT; GNE; SPS; SPP; CSS; and CST; or

(c) β1,4GalT, UDP-Gal transporter, UDP-Gal epimerase; α2,3SialT; GNE; SPS; SPP; CSS; and CST.

4. The method of claim 1 wherein the UDP-GlcNAc transporter of step (a)(ii) is a *Kluyveromyces lactis* UDP-GlcNAc transporter, or a murine UDP-GlcNAc transporter.

5. The method of claim 1 wherein the α-1,2-mannosidase is derived from *Trichoderma reesei*, *Saccharomyces sp.*, *Coccidiodes sp.*, or *Aspergillus sp.*

6. The method of claim 1 wherein the POMGnT1 enzyme of step (a)(ii) comprises SEQ ID NO: 110, SEQ ID NO: 112 or SEQ ID NO: 118.

7. The method of claim 1 wherein the POMGnT1 enzyme from step (a)(ii) is derived from human, murine, or frog POMGnT1 sequences.

8. The method of claim 1 wherein nucleic acid encoding the POMGnT1 enzyme from step (a)(ii) is codon-optimized for expression in the lower eukaryotic host cell.

9. The method of claim 1 wherein the POMGnT1 of step (a)(ii) is encoded by nucleic acid sequence encoding POMGnT1 operatively linked to an ER and/or Golgi leader sequence which is that of or derived from a leader sequence of S. *cerevisiae*, *Pichia pastoris*, or *Kluyveromyces lactis*.

10. The method of claim 9 wherein the ER and/or Golgi leader sequence is derived from sequence of one of the following proteins: ScMNS1-s, ScMNN9-s, PpKre-2s; KlGnt1-s, ScMNN2-s, ScMNN2-m, ScMNN5-s, ScMNN6-s, ScPMT5-m, PpPMT1-m or PpBET1-s

11. The method of claim 9 wherein the nucleic acid sequence encoding POMGnT1 is operatively linked to a sequence selected from: SEQ ID NOs: 1-108.

12. The method of claim 9 wherein the nucleic acid sequence encoding POMGnT1 is operatively linked to a sequence selected from: SEQ ID NO 3 (or sequence encoding SEQ ID NO 4); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 45 (or sequence encoding SEQ ID NO 46); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 93 (or sequence encoding SEQ ID NO 94); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88) or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

13. The method of claim 9 wherein the nucleic acid sequence encoding POMGnT1 is operatively linked to an ER and/or Golgi leader sequence in one of the following combinations:

(a) human POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a leader sequence which is that of, or derived

from that of, a protein selected from: PpSEC12-s; ScMNN9-s; K1GNT1-s; ScMNN2-s; ScMNN2-m; ScPMT5-m; PpPMT1-m or PpBET1-s;

(b) human POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a sequence selected from: SEQ ID NO 9 (or sequence encoding SEQ ID NO 10); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 73 (or sequence encoding SEQ ID NO 74); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88) or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106);

(c) murine POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a leader sequence which is that of, or derived from that of, a protein selected from: ScMNN9-s; PpKRE2-s; ScKTR2-s; K1GNT1-s; ScMNN2-s; ScMNN2-m; ScMNN5-s; ScMNN6-s; ScPMT5-m; ScPMT6-m; PpPMT1-s; PpPMT 2-s; PpPMT4-s; or PpBET1-s;

(d) murine POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a sequence selected from: SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 37 (or sequence encoding SEQ ID NO 38); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 45 (or sequence encoding SEQ ID NO 46); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 73 (or sequence encoding SEQ ID NO: 74); SEQ ID NO 75 (or sequence encoding SEQ ID NO 76); SEQ ID NO 77 (or sequence encoding SEQ ID NO 78); SEQ ID NO 79 (or sequence encoding SEQ ID NO 80); SEQ ID NO 81 (or sequence encoding SEQ ID NO 82); or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106);

(e) frog POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a leader sequence which is that of, or derived from that of, a protein selected from: ScMNS1-s; ScSEC12-m; PpSEC12-s; ScMNN9-s; ScANP1-s; ScHOC1-s; ScMNN10-s; ScMNN11-s; PpKRE2-s; ScKTR2-s; K1GNT1-s; ScMNN2-s;

ScMNN2-m; ScMNN1-s; ScMNN6-s; ScPMT5-m; PpPMT1-m; or PpBET1-s; or

(f) frog POMGnT1 sequence or catalytic domain sequence thereof operatively linked to a sequence selected from: SEQ ID NO 3 (or sequence encoding SEQ ID NO 4); SEQ ID NO 7 (or sequence encoding SEQ ID NO 8); SEQ ID NO 9 (or sequence encoding SEQ ID NO 10); SEQ ID NO 13 (or sequence encoding SEQ ID NO 14); SEQ ID NO 17 (or sequence encoding SEQ ID NO 18); SEQ ID NO 19 (or sequence encoding SEQ ID NO 20); SEQ ID NO 21 (or sequence encoding SEQ ID NO 22); SEQ ID NO 23 (or sequence encoding SEQ ID NO 24); SEQ ID NO 33 (or sequence encoding SEQ ID NO 34); SEQ ID NO 37 (or sequence encoding SEQ ID NO 38); SEQ ID NO 39 (or sequence encoding SEQ ID NO 40); SEQ ID NO 41 (or sequence encoding SEQ ID NO 42); SEQ ID NO 43 (or sequence encoding SEQ ID NO 44); SEQ ID NO 49 (or sequence encoding SEQ ID NO 50); SEQ ID NO 51 (or sequence encoding SEQ ID NO 52); SEQ ID NO 73 (or sequence encoding SEQ ID NO 74); SEQ ID NO 87 (or sequence encoding SEQ ID NO 88); or SEQ ID NO 105 (or sequence encoding SEQ ID NO 106).

14. The method of claim 1 wherein nucleic acid sequence encoding the POMGnT1 enzyme from step (a)(ii) is operatively linked to a Pichia AOX1, GAP, TEF or PMA promoter.

15. The method of claim 1 wherein nucleic acid sequence encoding the POMGnT1 enzyme from step (a)(ii) is operatively linked to a ScCYC1 transcription termination sequence.

16. The method of claim 1 wherein nucleic acid sequence encoding the POMGnT1 enzyme of step (a)(ii) is operatively linked to promoter and leader sequences; wherein the promoter, leader and POMGnT1 sequences utilized are one of the following promoter-leader-POMGnT1 combinations: AOX1-PpKRE2s-mouse POMGnT1; GAP-K1GNT1s-mouse POMGnT1; AOX1-K1GNT1s-mouse POMGnT1; GAP-K1GNT1s-frog POMGnT1; AOX1-ScMNN2s-mouse POMGnT1; GAP-ScMNN5s-mouse POMGnT1; AOX1-ScMNN5s-mouse POMGnT1; GAP-ScMNN6s-mouse POMGnT1; AOX1-ScPMT5m-mouse POMGnT1 or GAP-ScMNN6s-human POMGNT1.

17. The method of claim 1 wherein the lower eukaryotic host cell is a *Pichia sp.*

18. The method of claim 1 wherein the lower eukaryotic host cell is *Pichia pastoris.*

19. A lower eukaryotic host cell of any of claims 1-18.

20. A lower eukaryotic host cell capable of producing a recombinant glycoprotein having O-Man-GlcNAc; O-Man-GlcNAc-Gal; or O-Man-GlcNAc-Gal-Sia as the predominant O-glycan which:

(a) does not express functional beta-mannosyltransferase enzymes Bmt 1, 2, 3 and 4, and phospho-mannose transferase enzymes Mnn4a, Mnn4b and PNO1; and

(b) which is transformed or transfected with exogenous nucleic acid encoding POMGnT1, UDP-GlcNAc transporter, α-1,2-mannosidase, and the glycoprotein of interest;

wherein the predominant O-glycan is present at a level greater than 50% and O-Man2 is present at less than 30%.

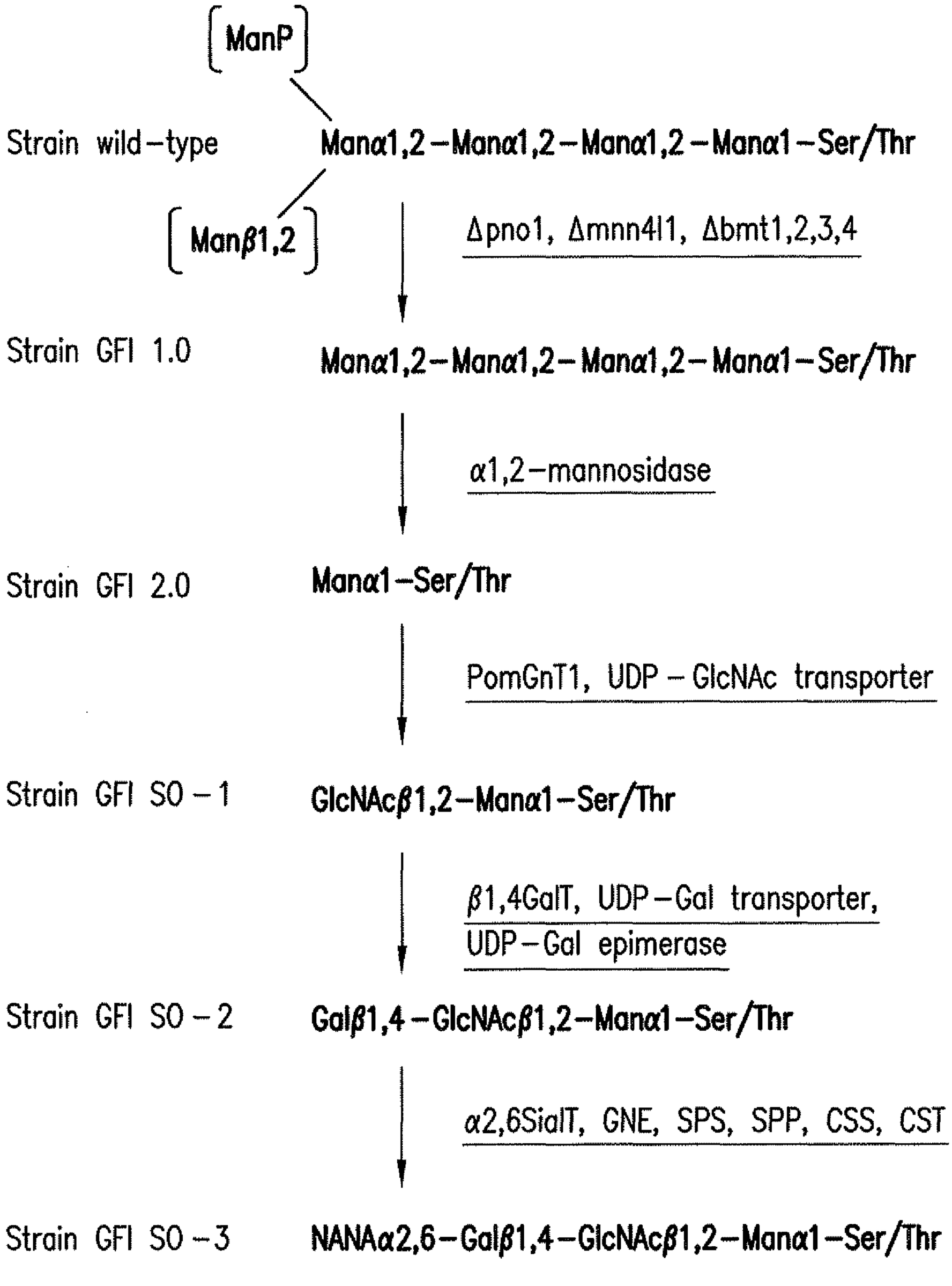

[ManP]
Strain wild-type
Manα1,2-Manα1,2-Manα1,2-Manα1-Ser/Thr
[Manβ1,2]
Δpno1, Δmnn4l1, Δbmt1,2,3,4
Strain GFI 1.0
Manα1,2-Manα1,2-Manα1,2-Manα1-Ser/Thr
α1,2-mannosidase
Strain GFI 2.0
Manα1-Ser/Thr
PomGnT1, UDP-GlcNAc transporter
Strain GFI SO-1
GlcNAcβ1,2-Manα1-Ser/Thr
β1,4GalT, UDP-Gal transporter,
UDP-Gal epimerase
Strain GFI SO-2
Galβ1,4-GlcNAcβ1,2-Manα1-Ser/Thr
α2,6SialT, GNE, SPS, SPP, CSS, CST
Strain GFI SO-3
NANAα2,6-Galβ1,4-GlcNAcβ1,2-Manα1-Ser/Thr

FIG. 1

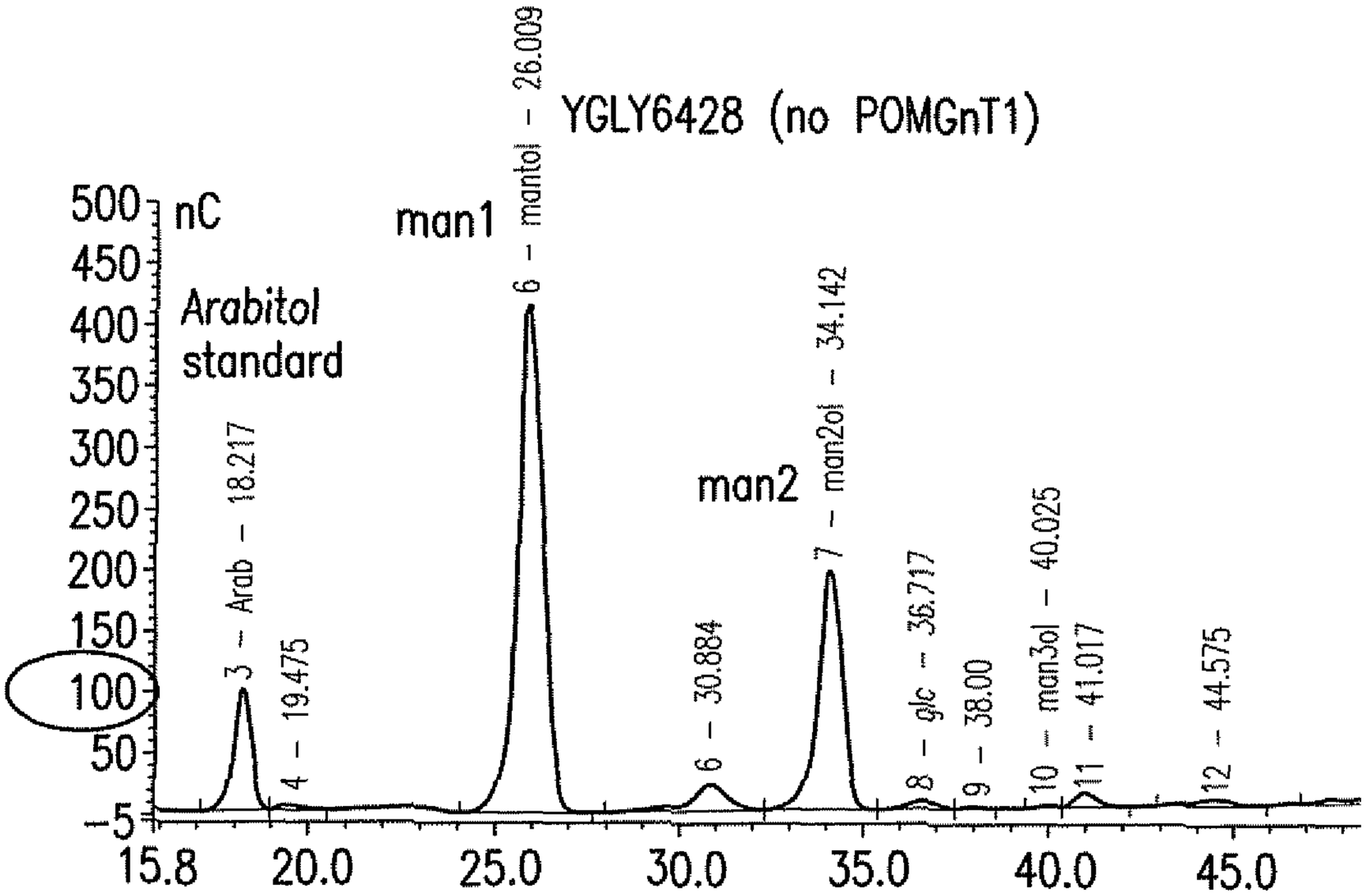
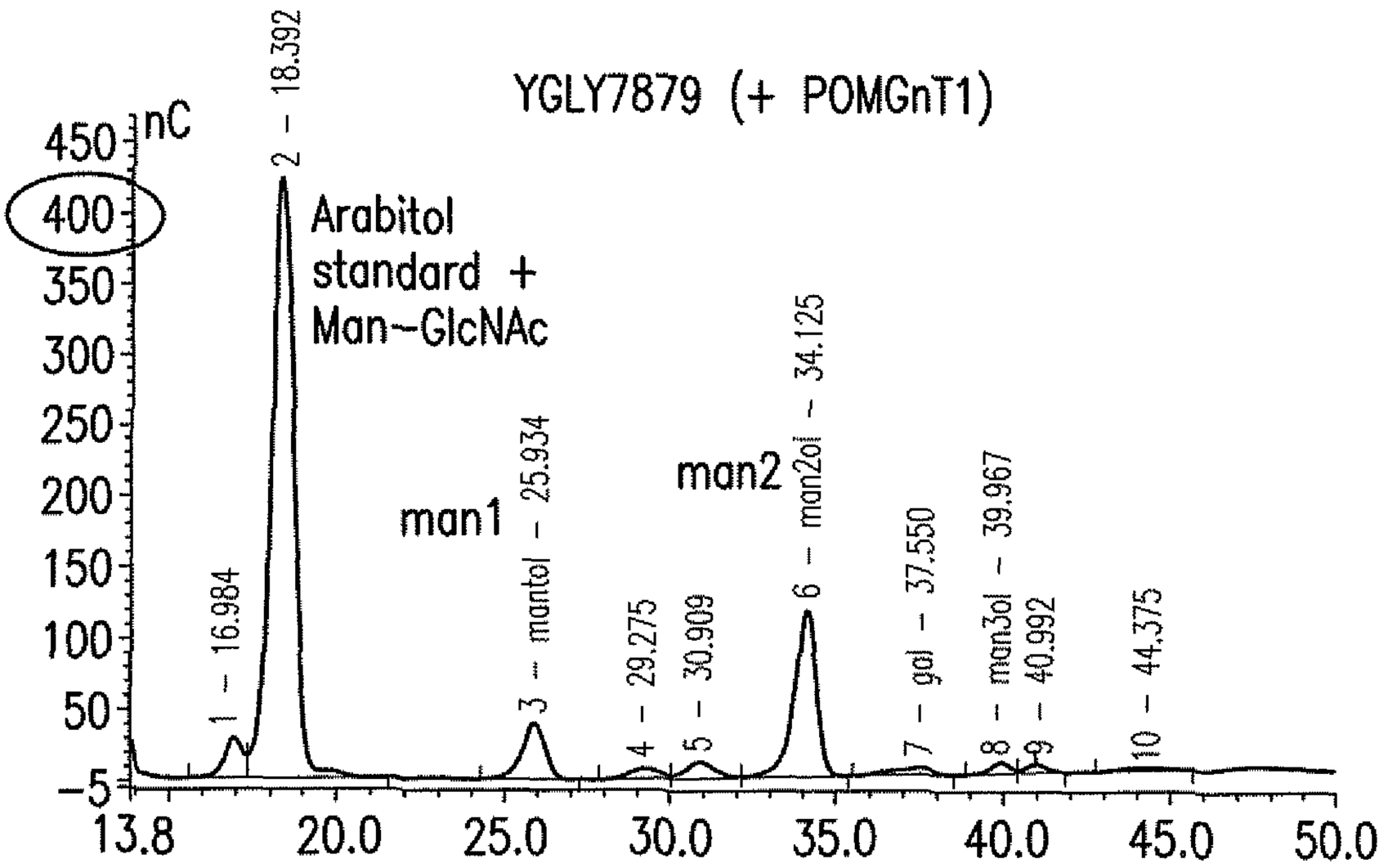


FIG.2A



• mantitol–GlcNAc co–migrates with arabitol standard

FIG.2B

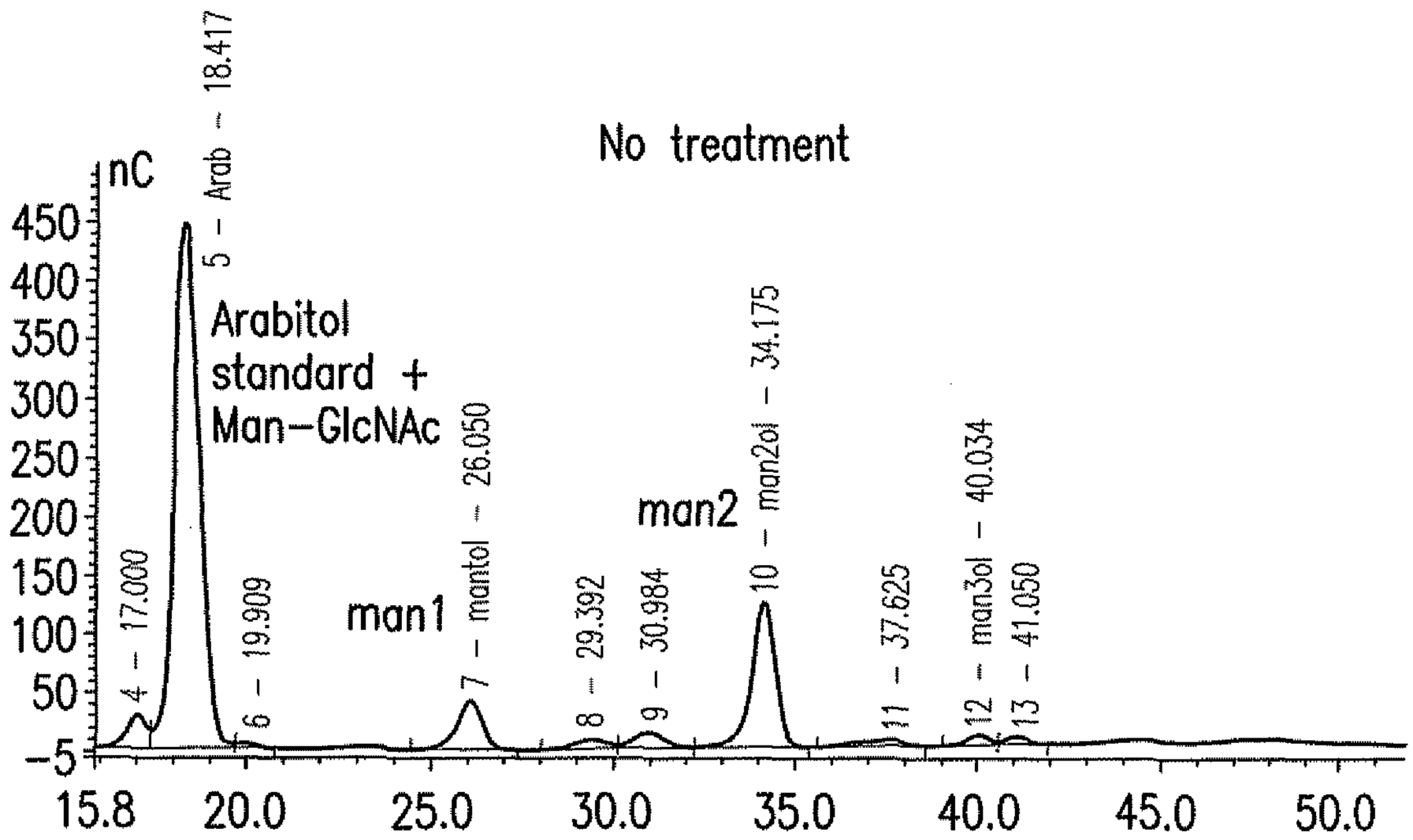

No treatment
nC
450
400
350
300
250
200
150
100
50
-5
4 – 17.000
5 – Arab – 18.417
Arabitol
standard +
Man–GlcNAc
6 – 19.909
man1
7 – mantol – 26.050
8 – 29.392
9 – 30.984
man2
10 – man2ol – 34.175
11 – 37.625
12 – man3ol – 40.034
13 – 41.050
15.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.3A

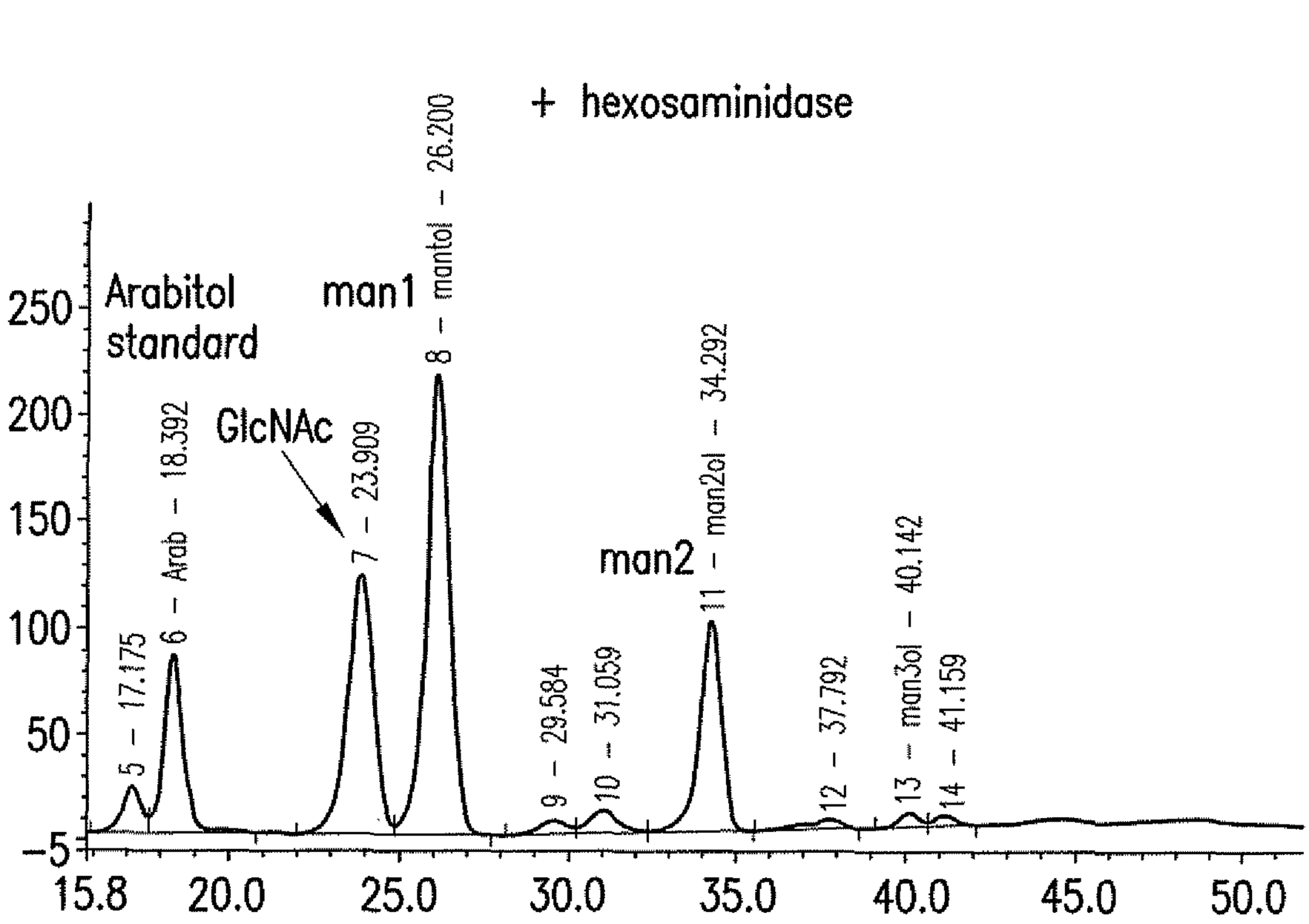

+ hexosaminidase
250
200
150
100
50
-5
Arabitol
standard
man1
8 – mantol – 26.200
GlcNAc
7 – 23.909
man2
11 – man2ol – 34.292
5 – 17.175
6 – Arab – 18.392
9 – 29.584
10 – 31.059
12 – 37.792
13 – man3ol – 40.142
14 – 41.159
15.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.3B

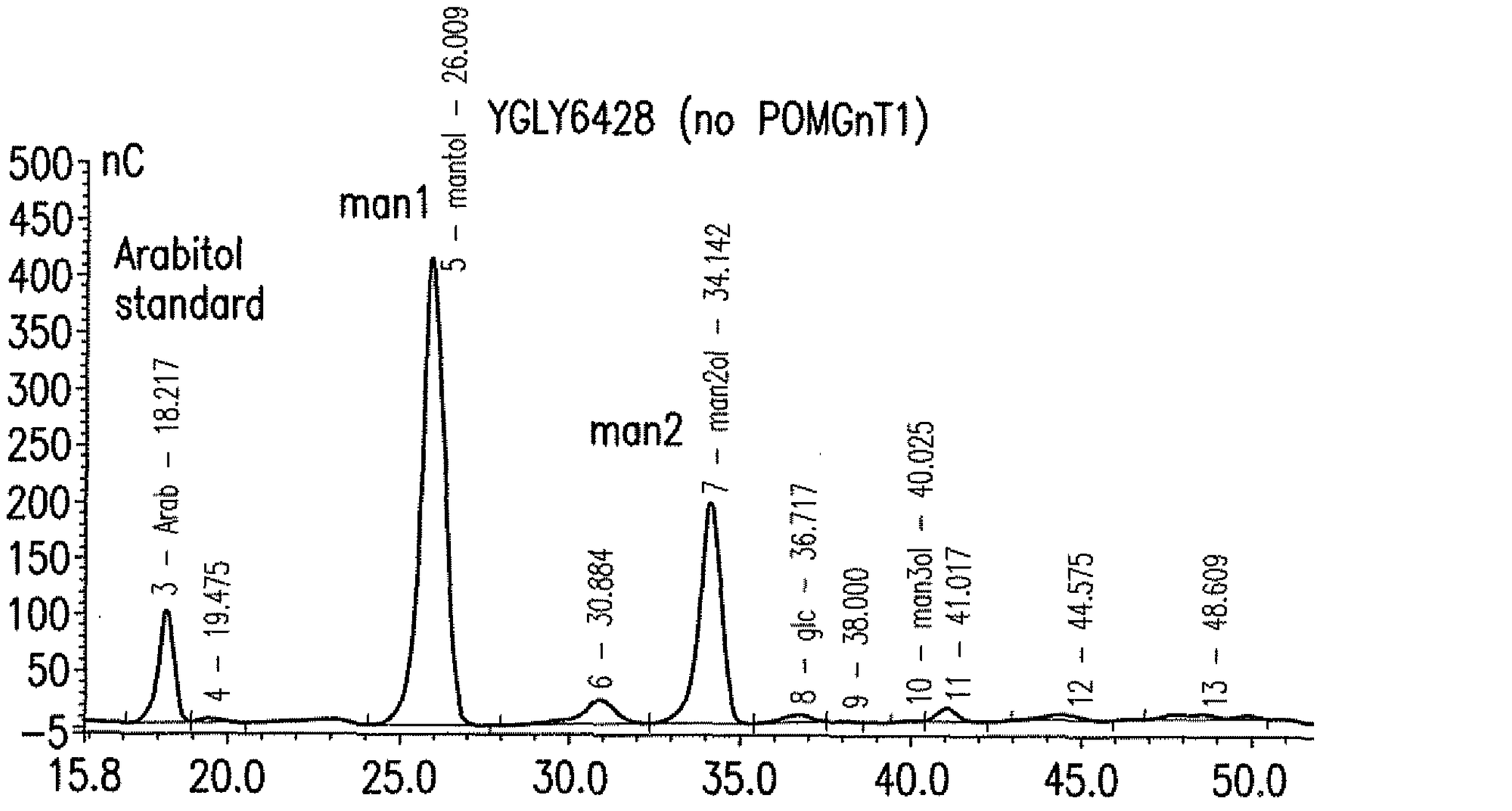

YGLY6428 (no POMGnT1)
nC
Arabitol standard
man1
man2
3 – Arab – 18.217
4 – 19.475
5 – mantol – 26.009
6 – 30.884
7 – man2ol – 34.142
8 – glc – 36.717
9 – 38.000
10 – man3ol – 40.025
11 – 41.017
12 – 44.575
13 – 48.609
500
450
400
350
300
250
200
150
100
50
–5
15.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.4A

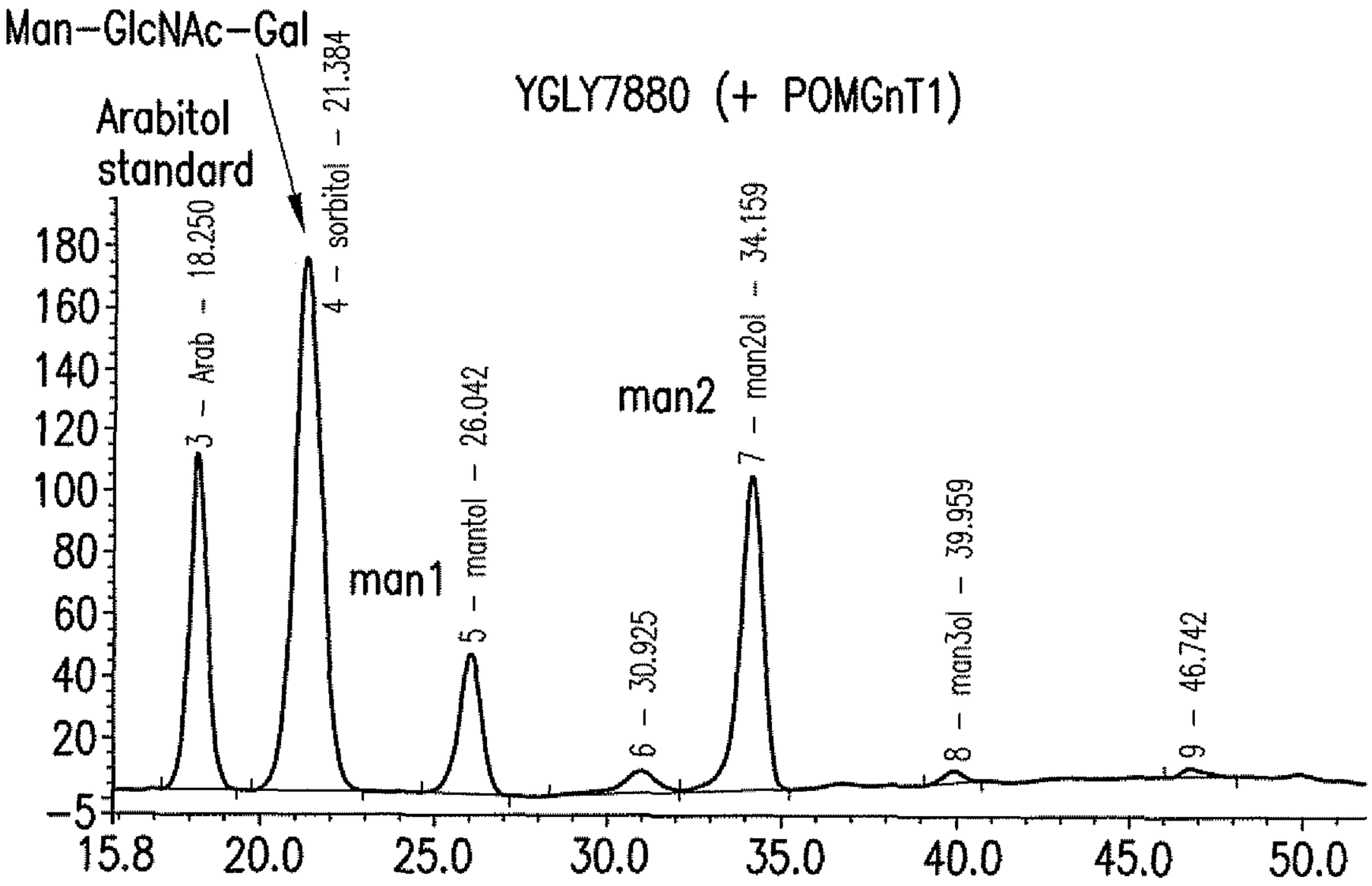

Man–GlcNAc–Gal
YGLY7880 (+ POMGnT1)
Arabitol standard
man1
man2
3 – Arab – 18.250
4 – sorbitol – 21.384
5 – mantol – 26.042
6 – 30.925
7 – man2ol – 34.159
8 – man3ol – 39.959
9 – 46.742
180
160
140
120
100
80
60
40
20
–5
15.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.4B

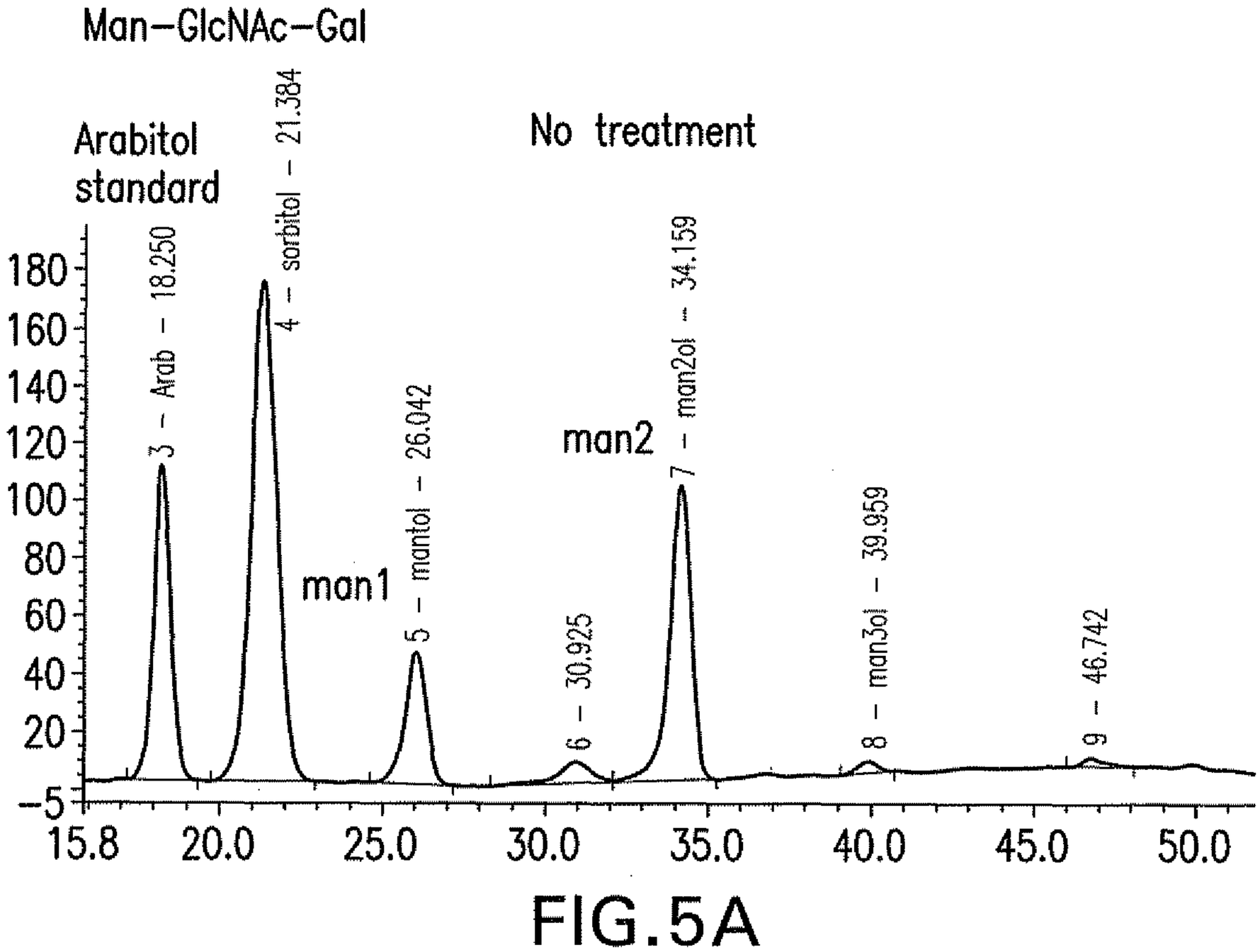

Man–GlcNAc–Gal
Arabitol standard
No treatment
3 – Arab – 18.250
4 – sorbitol – 21.384
5 – mantol – 26.042
6 – 30.925
7 – man2ol – 34.159
8 – man3ol – 39.959
9 – 46.742
man1
man2
180
160
140
120
100
80
60
40
20
−5
15.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.5A

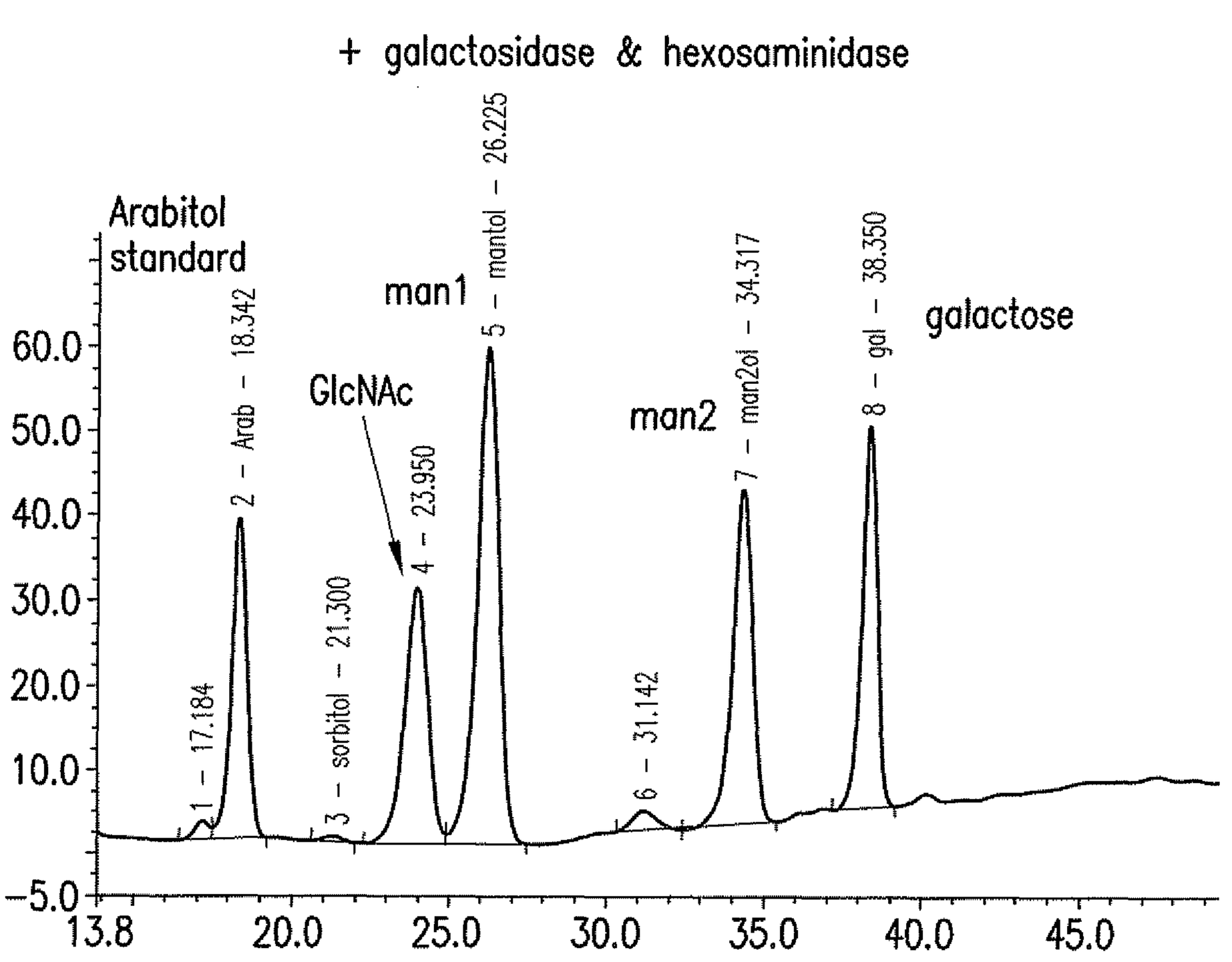

+ galactosidase & hexosaminidase
Arabitol standard
GlcNAc
man1
man2
galactose
1 – 17.184
2 – Arab – 18.342
3 – sorbitol – 21.300
4 – 23.950
5 – mantol – 26.225
6 – 31.142
7 – man2ol – 34.317
8 – gal – 38.350
60.0
50.0
40.0
30.0
20.0
10.0
−5.0
13.8
20.0
25.0
30.0
35.0
40.0
45.0

FIG.5B

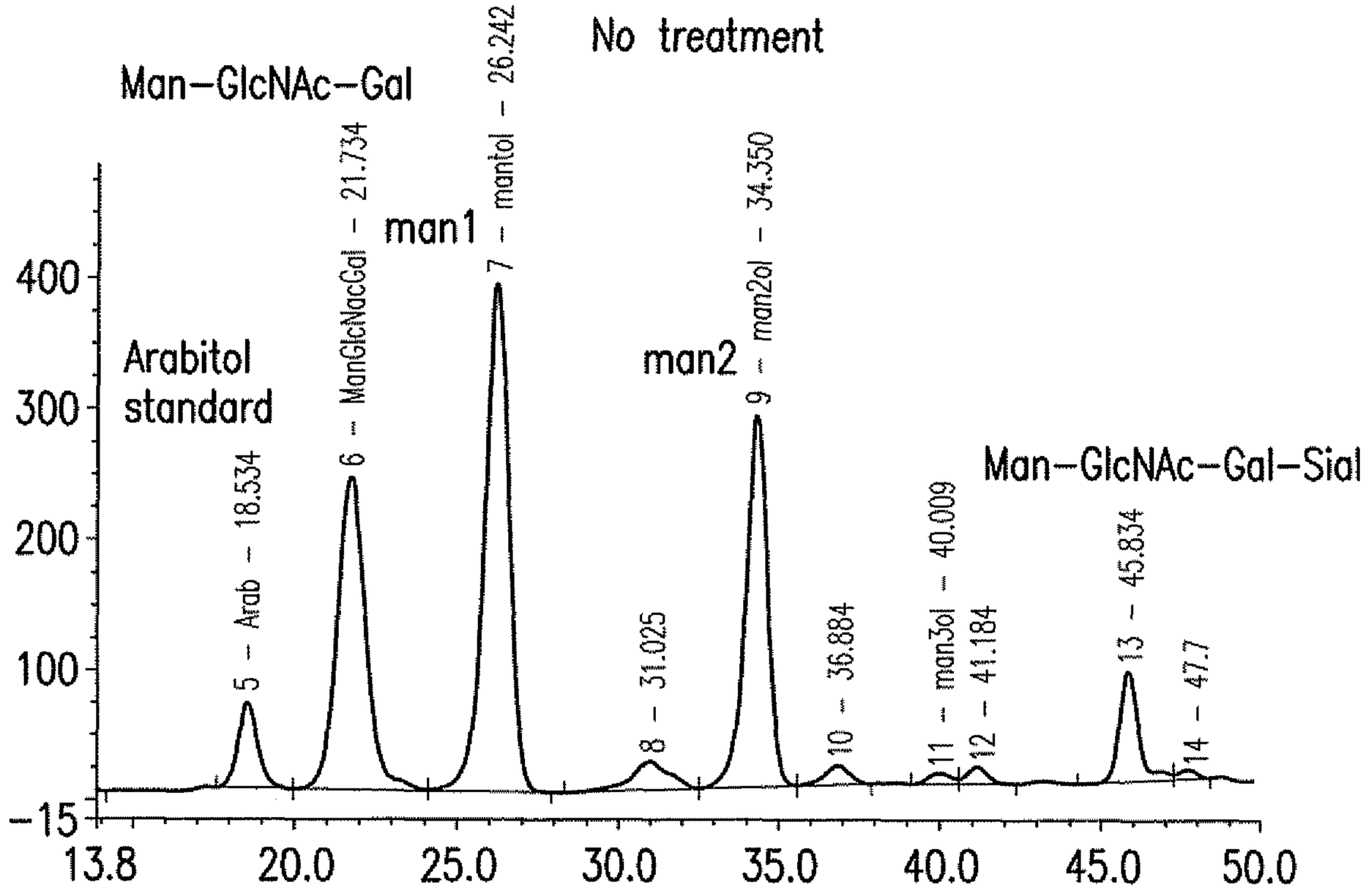

No treatment
Man–GlcNAc–Gal
Arabitol standard
man1
man2
Man–GlcNAc–Gal–Sial
5 – Arab – 18.534
6 – ManGlcNacGal – 21.734
7 – mantol – 26.242
8 – 31.025
9 – man2ol – 34.350
10 – 36.884
11 – man3ol – 40.009
12 – 41.184
13 – 45.834
14 – 47.7
400
300
200
100
–15
13.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.6A

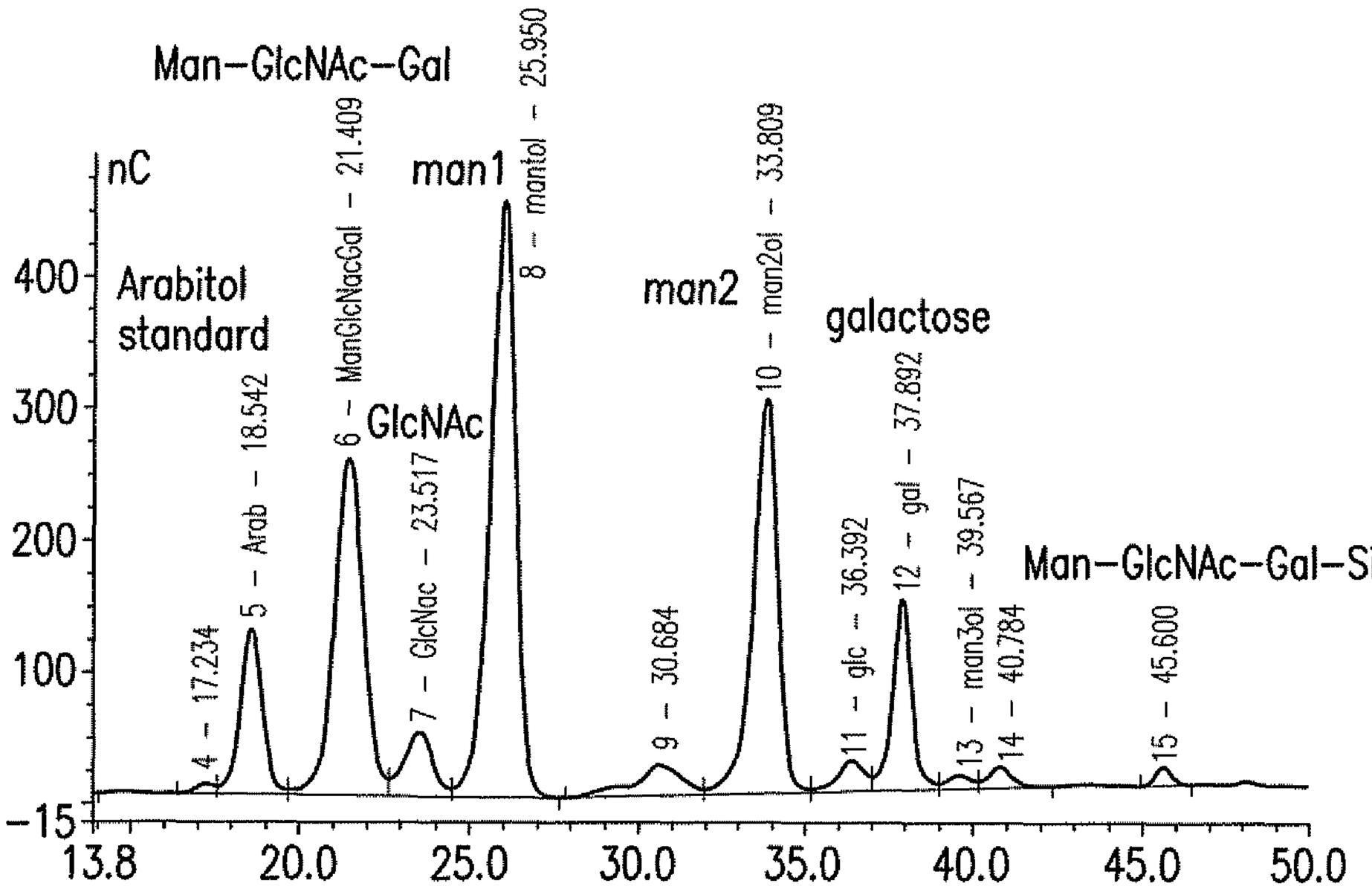

+ neuraminidase, galactosidase & hexosaminidase
Man–GlcNAc–Gal
nC
Arabitol standard
man1
man2
galactose
Man–GlcNAc–Gal–Sial
4 – 17.234
5 – Arab – 18.542
6 – ManGlcNacGal – 21.409
7 – GlcNac – 23.517
8 – mantol – 25.950
9 – 30.684
10 – man2ol – 33.809
11 – glc – 36.392
12 – gal – 37.892
13 – man3ol – 39.567
14 – 40.784
15 – 45.600
400
300
200
100
–15
13.8
20.0
25.0
30.0
35.0
40.0
45.0
50.0

FIG.6B

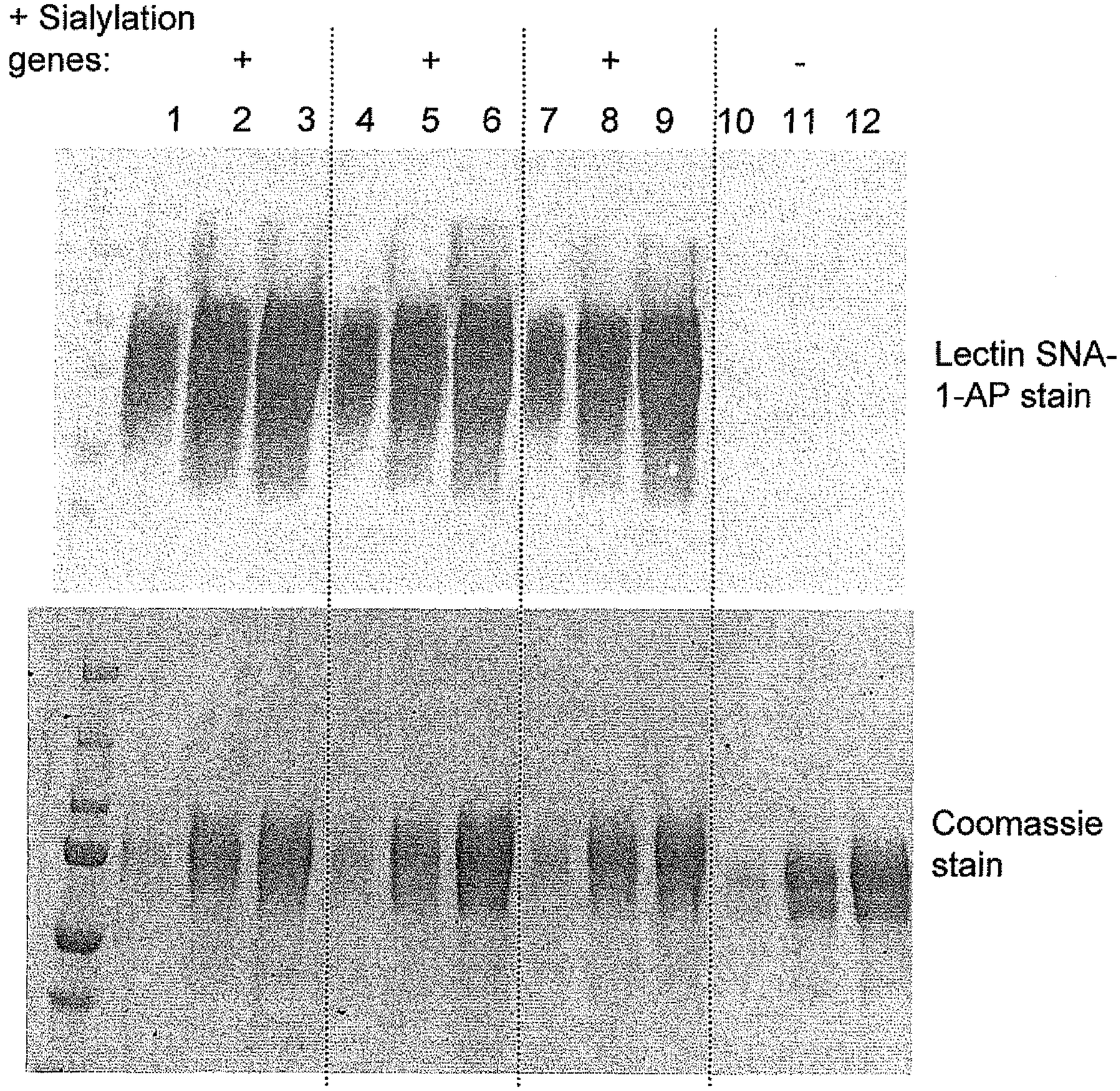

+ Sialylation genes:
+
+
+
-
1 2 3 4 5 6 7 8 9 10 11 12
Lectin SNA-1-AP stain
Coomassie stain

FIG.7

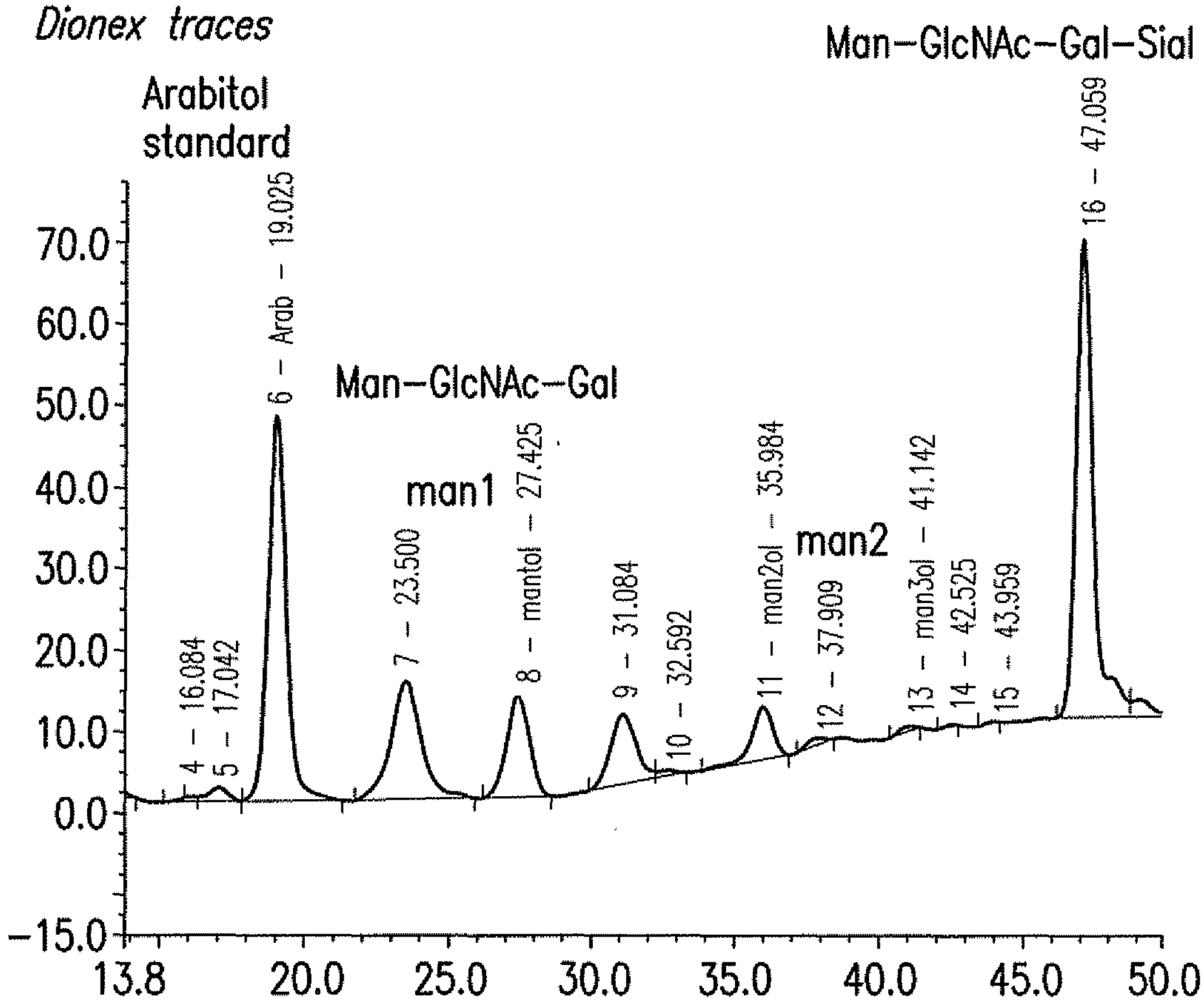


FIG.8A

Percentages of O–glycan species from PANEL A.

man1	Man2	Man–GlcNAc	Man–GlcNAc–Gal	Man–GlcNAc–Gal–Sia
11	6	0	27	56

FIG.8B

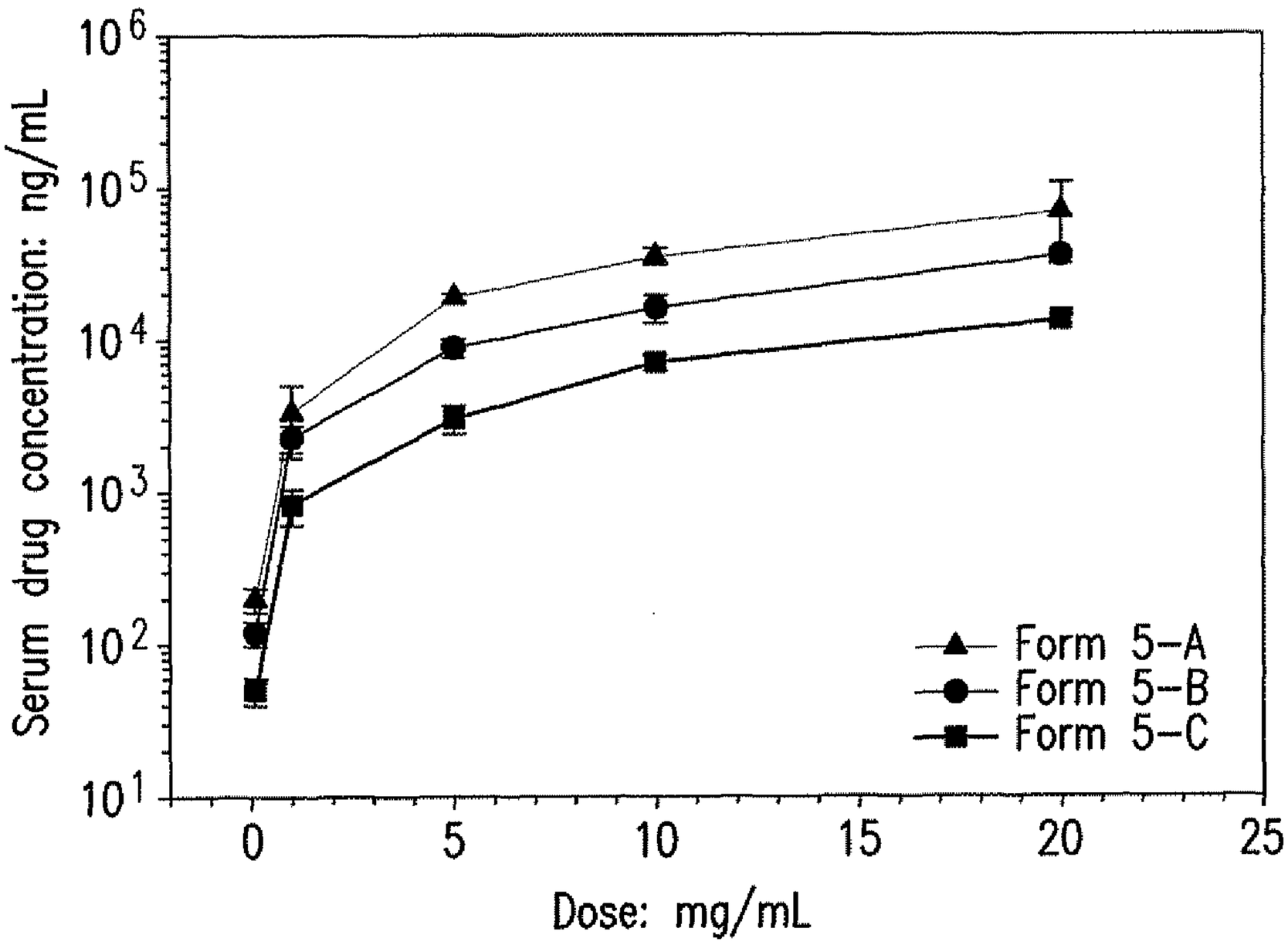

Serum drug concentration: ng/mL
10^6
10^5
10^4
10^3
10^2
10^1
0
5
10
15
20
25
Dose: mg/mL
Form 5-A
Form 5-B
Form 5-C

FIG.9

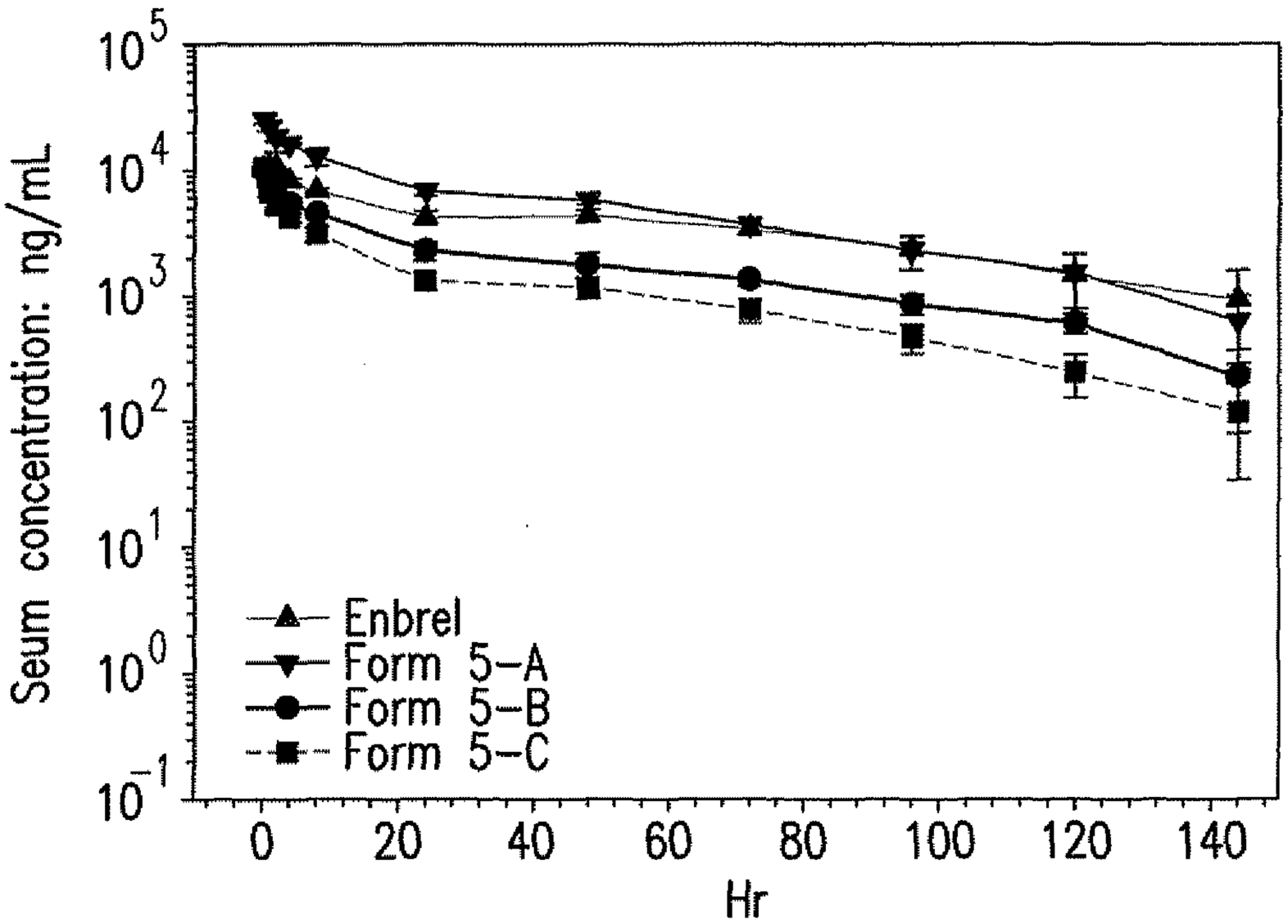


FIG.10A

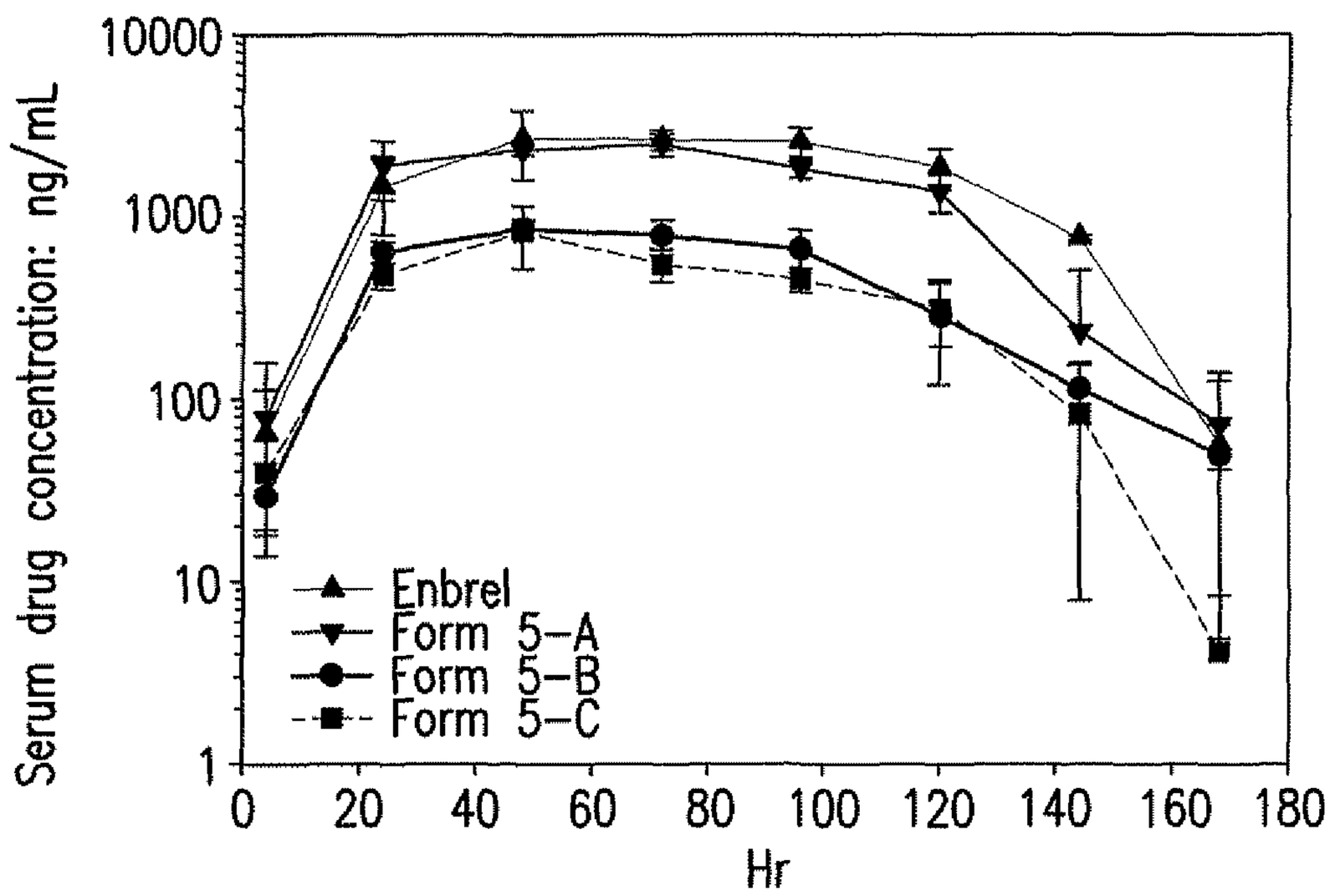


FIG.10B

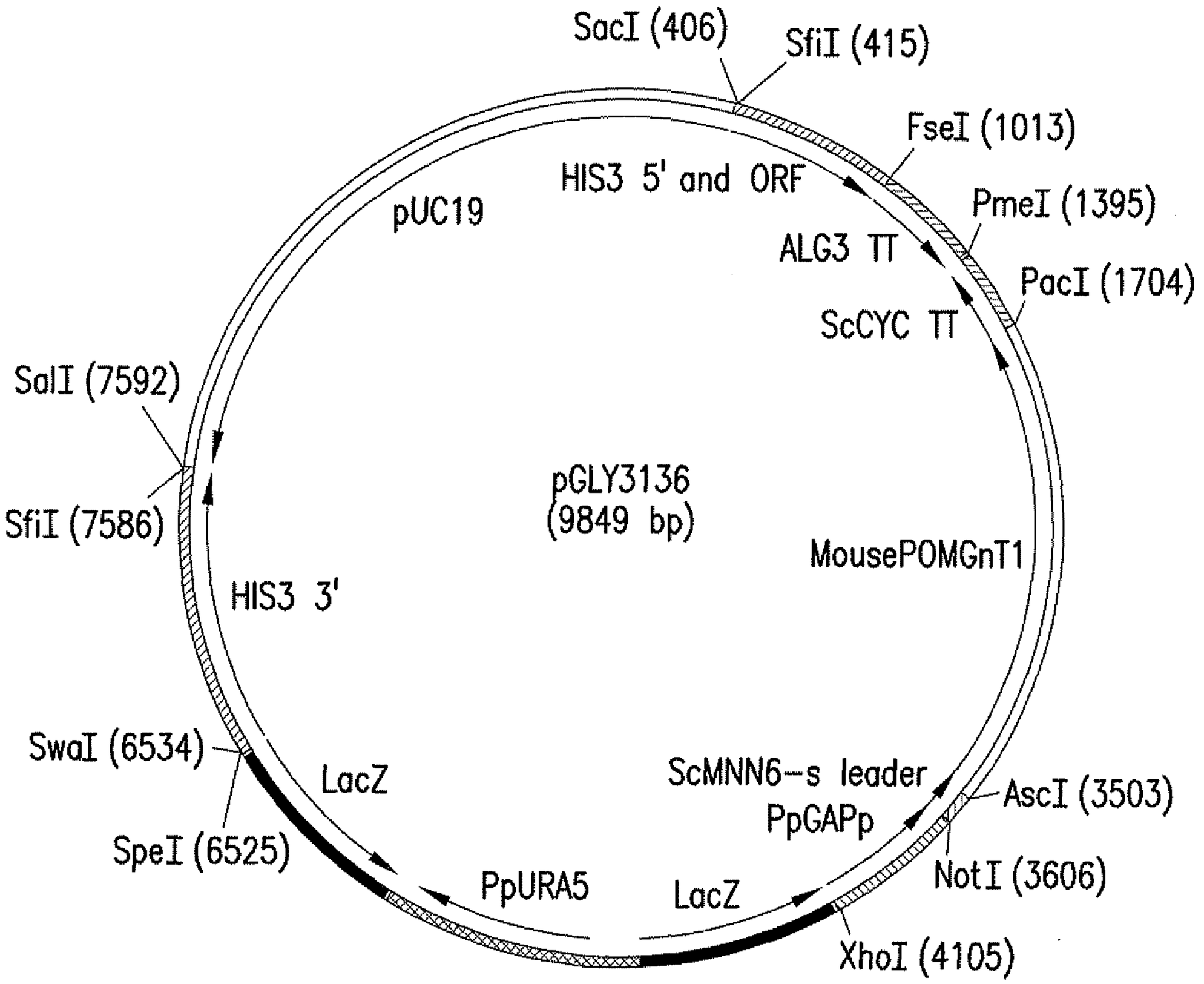

SacI (406)
SfiI (415)
FseI (1013)
PmeI (1395)
PacI (1704)
SalI (7592)
SfiI (7586)
SwaI (6534)
SpeI (6525)
HIS3 5' and ORF
pUC19
ALG3 TT
ScCYC TT
pGLY3136
(9849 bp)
MousePOMGnT1
HIS3 3'
LacZ
PpURA5
LacZ
ScMNN6-s leader
PpGAPp
AscI (3503)
NotI (3606)
XhoI (4105)

FIG.11

FIG. 1

ManP
Strain wild-type
Manα1,2-Manα1,2-Manα1,2-Manα1-Ser/Thr
Manβ1,2
Δpno1, Δmnn4l1, Δbmt1,2,3,4
Strain GFI 1.0
Manα1,2-Manα1,2-Manα1,2-Manα1-Ser/Thr
α1,2-mannosidase
Strain GFI 2.0
Manα1-Ser/Thr
PomGnT1, UDP-GlcNAc transporter
Strain GFI SO-1
GlcNAcβ1,2-Manα1-Ser/Thr
β1,4GalT, UDP-Gal transporter,
UDP-Gal epimerase
Strain GFI SO-2
Galβ1,4-GlcNAcβ1,2-Manα1-Ser/Thr
α2,6SialT, GNE, SPS, SPP, CSS, CST
Strain GFI SO-3
NANAα2,6-Galβ1,4-GlcNAcβ1,2-Manα1-Ser/Thr