



- (51) **International Patent Classification:**
C12N 1/00 (2006.01) *C12N 15/74* (2006.01)
- (21) **International Application Number:**
PCT/US2011/025878
- (22) **International Filing Date:**
23 February 2011 (23.02.2011)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/307,642 24 February 2010 (24.02.2010) 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):** **SETHURAMAN, Natarajan** [US/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **CHOI, Byung-Kwon** [KR/US]; 21 Lafayette Street, Suite 200, Lebanon, New Hampshire 03766 (US). **PRINZ, Bianka** [DE/US]; 10 Allen Street, Lebanon, New Hampshire 03766 (US). **MEEHL, Michael** [US/US]; 21 Lafayette Street, Lebanon, New Hampshire 03766 (US). **STADHEIM, Terrance A.** [US/US]; 22 Market Street, Lyme, NH 03768 (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, RO, RS, RU, 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.

- (84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, 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:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** METHOD FOR INCREASING N-GLYCOSYLATION SITE OCCUPANCY ON THERAPEUTIC GLYCOPROTEINS PRODUCED IN PICHIA PASTORIS

(57) **Abstract:** Described is a method for increasing the *N*-glycosylation site occupancy of a therapeutic glycoprotein produced in recombinant host cells modified as described herein and genetically engineered to express the glycoprotein compared to the *N*-glycosylation site occupancy of the therapeutic glycoprotein produced in a recombinant host cell not modified as described herein. In particular, the method provides recombinant host cells that overexpress a heterologous single-subunit oligosaccharyltransferase, which in particular embodiments is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of the yeast oligosaccharyltransferase (OTase) complex, for example, the *Leishmania major* STT3D protein, in the presence of expression of the host cell genes encoding the endogenous OTase complex. The method is useful for both producing therapeutic glycoproteins with increased *N*-glycosylation site occupancy in lower eukaryote cells such as yeast and filamentous fungi and in higher eukaryote cells such as plant and insect cells and mammalian cells.



TITLE OF THE INVENTION

METHOD FOR INCREASING N-GLYCOSYLATION SITE OCCUPANCY ON
THERAPEUTIC GLYCOPROTEINS PRODUCED IN *PICHIA PASTORIS*

5 CROSS REFERENCE TO RELATED APPLICATIONS

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

BACKGROUND OF THE INVENTION

10 (1) Field of the Invention

The present invention relates to methods for increasing the *N*-glycosylation site occupancy of a heterologous glycoprotein produced in a recombinant host cell modified according to the present invention and genetically engineered to express the glycoprotein compared to the *N*-glycosylation site occupancy of the therapeutic glycoprotein produced in a recombinant host cell not modified according to the present invention. In particular, the present invention provides recombinant host cells that overexpress a heterologous single-subunit oligosaccharyltransferase, which in particular embodiments is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of the yeast oligosaccharyltransferase (OTase) complex in the presence of the host cell's endogenous OTase complex and methods to using these host cells to produce heterologous glycoproteins.

(2) Description of Related Art

The ability to produce recombinant human proteins has led to major advances in human health care and remains an active area of drug discovery. Many therapeutic proteins require the posttranslational addition of glycans to specific asparagine residues (*N*-glycosylation) of the protein to ensure proper structure-function activity and subsequent stability in human serum. For therapeutic use in humans, glycoproteins require human-like *N*-glycosylation. Mammalian cell lines (e.g., Chinese hamster ovary (CHO) cells, human retinal cells) that can mimic human-like glycoprotein processing have several drawbacks including low protein titers, long fermentation times, heterogeneous products, and continued viral containment. It is therefore desirable to use an expression system that not only produces high protein titers with short fermentation times, but can also produce human-like glycoproteins.

Fungal hosts such as *Saccharomyces cerevisiae* or methylotrophic yeast such as *Pichia pastoris* have distinct advantages for therapeutic protein expression, for example, they do not secrete high amounts of endogenous proteins, strong inducible promoters for producing heterologous proteins are available, they can be grown in defined chemical media and without the use of animal sera, and they can produce high titers of recombinant proteins (Cregg *et al.*,

FEMS Microbiol. Rev. 24: 45-66 (2000)). However, glycosylated proteins expressed in yeast generally contain additional mannose sugars resulting in "high mannose" glycans. Because these high mannose *N*-glycans can result in adverse responses when administered to certain individuals, yeast have not generally been used to produce therapeutic glycoproteins intended for human use. However, methods for genetically engineering yeast to produce human-like *N*-glycans are described in U.S. Patent Nos. 7,029,872 and 7,449,308 along with methods described in U.S. Published Application Nos. 20040230042, 20050208617, 20040171826, 20050208617, and 20060286637. These methods have been used to construct recombinant yeast that can produce therapeutic glycoproteins that have predominantly human-like complex or hybrid *N*-glycans thereon instead of yeast type *N*-glycans.

It has been found that while the genetically engineered yeast can produce glycoproteins that have mammalian- or human-like *N*-glycans, the occupancy of *N*-glycan attachment sites on glycoproteins varies widely and is generally lower than the occupancy of these same sites in glycoproteins produced in mammalian cells. This has been observed for various recombinant antibodies produced in *Pichia pastoris*. However, variability of occupancy of *N*-glycan attachment sites has also been observed in mammalian cells as well. For example, Gawlitzek *et al.*, Identification of cell culture conditions to control *N*-glycosylation site-occupancy of recombinant glycoproteins expressed in CHO cells, Biotechnol. Bioengin. 103: 1164-1175 (2009), disclosed that *N*-glycosylation site occupancy can vary for particular sites for particular glycoproteins produced in CHO cells and that modifications in growth conditions can be made to control occupancy at these sites. International Published Application No. WO 2006107990 discloses a method for improving protein *N*-glycosylation of eukaryotic cells using the dolichol-linked oligosaccharide synthesis pathway. Control of *N*-glycosylation site occupancy has been reviewed by Jones *et al.*, Biochim. Biophys. Acta. 1726: 121-137 (2005). However, there still remains a need for methods for increasing *N*-glycosylation site occupancy of therapeutic proteins produced in recombinant host cells.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a method for producing therapeutic glycoproteins in recombinant host cells modified as disclosed herein wherein the *N*-glycosylation site occupancy of the glycoproteins produced in the host cells modified as disclosed herein is increased over the *N*-glycosylation site occupancy of the same glycoproteins produced in host cells not modified as disclosed herein. For example, in yeast host cells modified as disclosed herein, the *N*-glycosylation site occupancy of glycoproteins produced therein will be the same as or more similar to the *N*-glycosylation site occupancy of the same glycoproteins produced in recombinant mammalian or human cells.

To increase the *N*-glycosylation site occupancy on a glycoprotein produced in a recombinant host cell one or more heterologous single-subunit oligosaccharyltransferase (OTase) is/are overexpressed in the recombinant host cell either before or simultaneously with the expression of the glycoprotein in the host cell. In particular aspects, at least one of the heterologous single-subunit oligosaccharyltransferase is capable of functionally complementing a lethal mutation of one or more essential subunits comprising the endogenous host cell hetero-oligomeric oligosaccharyltransferase (OTase) complex. The *Leishmania major* STT3D protein is an example of a heterologous single-subunit oligosaccharyltransferase that has been shown to suppress a lethal mutation in the *STT3* locus and at least one locus selected from *WBPI*, *OST1*, *SWP1*, and *OST2* in *Saccharomyces cerevisiae* (Naseb *et al.*, Molec. Biol. Cell 19: 3758-3768 (2008)). In general, the one or more heterologous single-subunit oligosaccharyltransferases is/are overexpressed constitutively or inducibly in the presence of the proteins comprising the host cell's endogenous OTase complex, including the host cell's endogenous STT3 protein. Expression cassettes encoding the heterologous single-subunit oligosaccharyltransferase gene can either be integrated into any site within the host cell genome or located in the extrachromosomal space of the host cell, i.e., autonomously replicating genetic elements such as plasmids, viruses, 2 μ m plasmid, minichromosomes, and the like.

In particular embodiments, one or more of the single-subunit oligosaccharyltransferases is/are the *Leishmania sp.* STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In particular embodiments, the one or more single-subunit oligosaccharyltransferases is/are the *Leishmania major* STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. The nucleic acid molecules encoding the single-subunit oligosaccharyltransferases are not overexpressed in lieu of the expression of the endogenous genes encoding the proteins comprising the host cell's OTase complex, including the host cell STT3 protein. Instead the nucleic acid molecules encoding the single-subunit oligosaccharyltransferases are overexpressed constitutively or inducibly in the presence of the expression of the genes encoding the proteins comprising the host cell's endogenous oligosaccharyltransferase (OTase) complex, which includes expression of the endogenous gene encoding the host cell's STT3. Each expression cassette encoding a single-subunit OTase can either be integrated into any site within the host cell genome or located in the extrachromosomal space of the host cell, i.e., autonomously replicating genetic elements such as plasmids, viruses, 2 μ m plasmid, minichromosomes, and the like.

The present invention has been exemplified herein using *Pichia pastoris* host cells genetically engineered to produce mammalian- or human-like complex *N*-glycans; however, the present invention can be applied to other yeast or filamentous fungal host cells, in particular, yeast or filamentous fungi genetically engineered to produce mammalian- or human-like complex or hybrid *N*-glycans, to improve the overall *N*-glycosylation site occupancy of

glycoproteins produced in the yeast or filamentous fungus host cell. In further aspects, the host cells are yeast or filamentous fungi that produce recombinant heterologous proteins that have wild-type or endogenous host cell *N*-glycosylation patterns, e.g., hypermannosylated or high mannose *N*-glycans. In further aspects, the host cells are yeast or filamentous fungi that lack
 5 alpha-1,6-mannosyltransferase activity (e.g., och1p activity in the case of various yeast strains such as but not limited to *Saccharomyces cerevisiae* or *Pichia pastoris*) and thus produce recombinant heterologous proteins that have high mannose *N*-glycans. Furthermore, the present invention can also be applied to plant and mammalian expression system to improve the overall
 10 *N*-glycosylation site occupancy of glycoproteins produced in these plant or mammalian expression systems, particularly glycoproteins that have more than two *N*-linked glycosylation sites.

Therefore, in one aspect of the above, provided is a method for producing a heterologous glycoprotein in a recombinant host cell, comprising providing a recombinant host cell that includes one or more nucleic acid molecules encoding one or more heterologous single-
 15 subunit oligosaccharyltransferases and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans in a
 20 host cell, comprising providing a host cell that includes one or more nucleic acid molecules encoding one or more heterologous single-subunit oligosaccharyltransferases and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In general, in the above aspects, the endogenous host cell genes encoding the
 25 proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed.

In further aspects of the above method, the host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia
 30 lindneri*), *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*, and *Neurospora crassa*. In other aspects, the host cell is an insect, plant or mammalian host cell.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein in a lower eukaryote host cell, comprising providing a recombinant lower eukaryote host cell that includes at least one nucleic acid molecule encoding a
 35

heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In further aspects of the above method, the lower eukaryote host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *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*, and *Neurospora crassa*.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein in a recombinant yeast host cell, comprising providing a recombinant yeast host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In the above methods, the recombinant yeast host cell either produces the glycoprotein with a yeast *N*-glycan pattern or the yeast has been genetically engineered to produce glycoproteins with a yeast pattern but which lack hypermannosylation but which produce high mannose *N*-glycans. For example, the yeast can be genetically engineered to lack α 1,6-mannosyltransferase activity, e.g., Och1p activity. In further aspects, the yeast is genetically engineered to produce glycoproteins that have mammalian or human-like *N*-glycans.

In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania* sp. STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In further embodiments, the single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of an OTase complex, for example, a yeast OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *Saccharomyces cerevisiae* and/or *Pichia pastoris* STT3 locus, *WBP1* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. For example, in further aspects, the for example single-subunit oligosaccharyltransferase

is the *Leishmania major* STT3D protein, which is capable of functionally suppressing (or rescuing or complementing) the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* OTase complex.

In further aspects of the above method, the yeast host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, and *Candida albicans*.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein in a recombinant yeast host cell, comprising providing a recombinant yeast host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of a yeast oligosaccharyltransferase (OTase) complex, and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In the above methods, the recombinant yeast host cell either produces the glycoprotein with a yeast *N*-glycan pattern or the yeast has been genetically engineered to produce glycoproteins with a yeast pattern that includes high mannose *N*-glycans but which lack hypermannosylation. For example, the yeast can be genetically engineered to lack α 1,6-mannosyltransferase activity, e.g., Och1p activity. In further aspects, the yeast is genetically engineered to produce glycoproteins that have mammalian or human-like *N*-glycans.

In particular embodiments, the host cell further includes one or more nucleic acid molecules encoding the *Leishmania* sp. STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In particular embodiments, the host cell further includes one or more nucleic acids encoding the *Leishmania major* STT3A protein, STT3B protein, STT3C protein, or combinations thereof.

In further aspects of the above method, the yeast host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, and *Candida albicans*.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein in a filamentous fungus host cell, comprising providing a filamentous

fungus host cell that includes at least one nucleic acid molecule encoding a single-subunit heterologous oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein. The filamentous fungus host cell produces the glycoprotein in which the *N*-glycans have a filamentous fungus pattern or it is genetically engineered to produce glycoproteins that have mammalian or human-like *N*-glycans.

In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania* sp. STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3A protein, STT3B protein, STT3C protein, STT3D protein or combinations thereof. In further embodiments, the single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of an OTase complex, for example, a yeast OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *Saccharomyces cerevisiae* and/or *Pichia pastoris* STT3 locus, *WBP1* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. For example, in further aspects, the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein, which is capable of functionally suppressing (or rescuing or complementing) the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* OTase complex.

In further aspects of the above, the filamentous fungus host cell is selected from the group consisting of *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, and *Neurospora crassa*.

In further embodiments of any one of the above methods, the host cell is genetically engineered to produce glycoproteins comprising one or more mammalian- or human-like complex *N*-glycans selected from G0, G1, G2, A1, or A2. In further embodiments, the host cell is genetically engineered to produce glycoproteins comprising one or more mammalian- or human-like complex *N*-glycans that have bisected *N*-glycans or have multiantennary *N*-glycans. In other embodiments, the host cell is genetically engineered to produce glycoproteins comprising one or more mammalian- or human-like hybrid *N*-glycans selected from GlcNAcMan₃GlcNAc₂; GalGlcNAcMan₃GlcNAc₂; NANAGlcNAcMan₃GlcNAc₂; Man₅GlcNAc₂, GlcNAcMan₅GlcNAc₂, GalGlcNAcMan₅GlcNAc₂, and NANAGlcNAcMan₅GlcNAc₂. In further embodiments, the *N*-glycan structure consists of the G-2 structure Man₃GlcNAc₂.

In particular embodiments of any one of the above methods, the heterologous glycoprotein can be for example, erythropoietin (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and granulocyte-colony stimulating factor (G-CSF); granulocyte macrophage-colony stimulating factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin; soluble IgE receptor α -chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 - Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; or IL-2 receptor agonist.

In further embodiments of any one of the above methods, the heterologous protein is an antibody, examples of which, include but are not limited to, an anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, or anti-CD20 antibody.

In particular aspects of any one of the above methods, the host cell includes one or more nucleic acid molecules encoding one or more catalytic domains of a glycosidase, mannosidase, or glycosyltransferase activity derived from a member of the group consisting of UDP-GlcNAc transferase (GnT) I, GnT II, GnT III, GnT IV, GnT V, GnT VI, UDP-galactosyltransferase (GalT), fucosyltransferase, and sialyltransferase. In particular embodiments, the mannosidase is selected from the group consisting of *C. elegans* mannosidase IA, *C. elegans* mannosidase IB, *D. melanogaster* mannosidase IA, *H. sapiens* mannosidase IB, *P. citrinum* mannosidase I, mouse mannosidase IA, mouse mannosidase IB, *A. nidulans* mannosidase IA, *A. nidulans* mannosidase IB, *A. nidulans* mannosidase IC, mouse mannosidase II, *C. elegans* mannosidase II, *H. sapiens* mannosidase II, and mannosidase III.

In certain aspects of any one of the above methods, at least one catalytic domain is localized by forming a fusion protein comprising the catalytic domain and a cellular targeting signal peptide. The fusion protein can be encoded by at least one genetic construct formed by the in-frame ligation of a DNA fragment encoding a cellular targeting signal peptide with a DNA fragment encoding a catalytic domain having enzymatic activity. Examples of targeting signal peptides include, but are not limited to, membrane-bound proteins of the ER or Golgi, retrieval signals, Type II membrane proteins, Type I membrane proteins, membrane spanning nucleotide

sugar transporters, mannosidases, sialyltransferases, glucosidases, mannosyltransferases, and phospho-mannosyltransferases.

In particular aspects of any one of the above methods, the host cell further includes one or more nucleic acid molecules encoding one or more enzymes selected from the group consisting of UDP-GlcNAc transporter, UDP-galactose transporter, GDP-fucose transporter, CMP-sialic acid transporter, and nucleotide diphosphatases.

In further aspects of any one of the above methods, the host cell includes one or more nucleic acid molecules encoding an α 1,2-mannosidase activity, a UDP-GlcNAc transferase (GnT) I activity, a mannosidase II activity, and a GnT II activity.

In further still aspects of any one of the above methods, the host cell includes one or more nucleic acid molecules encoding an α 1,2-mannosidase activity, a UDP-GlcNAc transferase (GnT) I activity, a mannosidase II activity, a GnT II activity, and a UDP-galactosyltransferase (GalT) activity.

In further still aspects of any one of the above methods, the host cell is deficient in the activity of one or more enzymes selected from the group consisting of mannosyltransferases and phosphomannosyltransferases. In further still aspects, the host cell does not express an enzyme selected from the group consisting of 1,6 mannosyltransferase, 1,3 mannosyltransferase, and 1,2 mannosyltransferase.

In a particular aspect of any one of the above methods, the host cell is an *och1* mutant of *Pichia pastoris*.

Further provided is a host cell, comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed, which includes expression of the endogenous host cell STT3 gene.

Further provided is a lower eukaryotic host cell, comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed.

Further provided is a yeast host cell, comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed.

Further provided is a yeast host cell, comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of a yeast

oligosaccharyltransferase (OTase) complex; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed.

Further provided is a filamentous fungus host cell comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed.

Further provided is a filamentous fungus host cell, comprising (a) a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of a yeast or filamentous fungus oligosaccharyltransferase (OTase) complex; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed

In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania sp.* STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In particular embodiments, the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3A protein, STT3B protein, STT3C protein, STT3D protein, or combinations thereof. In further embodiments, the single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of an OTase complex, for example, a yeast OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *Saccharomyces cerevisiae* and/or *Pichia pastoris* STT3 locus, *WBP1* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. For example, in further aspects, the for example single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein, which is capable of functionally suppressing (or rescuing or complementing) the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* OTase complex.

In further aspects, the above host cells further include one or more nucleic acid molecules encoding a *Leishmania sp.* STT3A protein, STT3B protein, STT3C protein, or combinations thereof.

In further embodiments of any one of the above host cells, the host cell is genetically engineered to produce glycoproteins comprising one or more mammalian- or human-like complex *N*-glycans selected from G0, G1, G2, A1, or A2. In further embodiments, the host cell is genetically engineered to produce glycoproteins comprising one or more human-like complex *N*-glycans that bisected *N*-glycans or have multiantennary *N*-glycans. In other embodiments, the host cell is genetically engineered to produce glycoproteins comprising one or

more mammalian- or human-like hybrid *N*-glycans selected from GlcNAcMan₃GlcNAc₂; GalGlcNAcMan₃GlcNAc₂; NANAGlcNAcMan₃GlcNAc₂; Man₅GlcNAc₂; GlcNAcMan₅GlcNAc₂; GalGlcNAcMan₅GlcNAc₂; and NANAGlcNAcMan₅GlcNAc₂. In further embodiments, the *N*-glycan structure consists of the G-2 structure Man₃GlcNAc₂.

5 In particular embodiments of any one of the above host cells, the heterologous glycoprotein can be for example, selected from the group consisting of erythropoietin (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and granulocyte-colony stimulating factor (GCSF); granulocyte macrophage-colony stimulating factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; 10 thrombin; soluble IgE receptor α -chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor 15 inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; and IL-2 receptor agonist.

In further embodiments of any one of the above host cells, the heterologous 20 protein is an antibody, examples of which, include but are not limited to, an anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, or anti-CD20 antibody.

25 In particular aspects of the above host cells, the host cell includes one or more nucleic acid molecules encoding one or more catalytic domains of a glycosidase, mannosidase, or glycosyltransferase activity derived from a member of the group consisting of UDP-GlcNAc transferase (GnT) I, GnT II, GnT III, GnT IV, GnT V, GnT VI, UDP-galactosyltransferase (GalT), fucosyltransferase, and sialyltransferase. In particular embodiments, the mannosidase is 30 selected from the group consisting of *C. elegans* mannosidase IA, *C. elegans* mannosidase IB, *D. melanogaster* mannosidase IA, *H. sapiens* mannosidase IB, *P. citrinum* mannosidase I, mouse mannosidase IA, mouse mannosidase IB, *A. nidulans* mannosidase IA, *A. nidulans* mannosidase IB, *A. nidulans* mannosidase IC, mouse mannosidase II, *C. elegans* mannosidase II, *H. sapiens* mannosidase II, and mannosidase III.

35 In certain aspects of any one of the above host cells, at least one catalytic domain is localized by forming a fusion protein comprising the catalytic domain and a cellular targeting signal peptide. The fusion protein can be encoded by at least one genetic construct formed by the

in-frame ligation of a DNA fragment encoding a cellular targeting signal peptide with a DNA fragment encoding a catalytic domain having enzymatic activity. Examples of targeting signal peptides include, but are not limited to, those to membrane-bound proteins of the ER or Golgi, retrieval signals such as HDEL or KDEL, Type II membrane proteins, Type I membrane proteins, 5 membrane spanning nucleotide sugar transporters, mannosidases, sialyltransferases, glucosidases, mannosyltransferases, and phospho-mannosyltransferases.

In particular aspects of any one of the above host cells, the host cell further includes one or more nucleic acid molecules encoding one or more enzymes selected from the group consisting of UDP-GlcNAc transporter, UDP-galactose transporter, GDP-fucose 10 transporter, CMP-sialic acid transporter, and nucleotide diphosphatases.

In further aspects of any one of the above host cells, the host cell includes one or more nucleic acid molecules encoding an α 1,2-mannosidase activity, a UDP-GlcNAc transferase (GnT) I activity, a mannosidase II activity, and a GnT II activity.

In further still aspects of any one of the above host cells, the host cell includes one 15 or more nucleic acid molecules encoding an α 1,2-mannosidase activity, a UDP-GlcNAc transferase (GnT) I activity, a mannosidase II activity, a GnT II activity, and a UDP-galactosyltransferase (GalT) activity.

In further aspects of any one of the above host cells, the host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, 20 *Pichia membranaefaciens*, *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*, 25 *Fusarium venenatum*, *Neurospora crassa*, plant cells, insect cells, and mammalian cells.

In further still aspects of any one of the above host cells, the host cell is deficient in the activity of one or more enzymes selected from the group consisting of mannosyltransferases and phosphomannosyltransferases. In further still aspects, the host cell does not express an enzyme selected from the group consisting of 1,6 mannosyltransferase, 1,3 30 mannosyltransferase, and 1,2 mannosyltransferase.

In a particular aspect of any one of the above host cells, the host cell is *Pichia pastoris*. In a further aspect, the host cell is an *och1* mutant of *Pichia pastoris*.

The methods and host cells herein can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*- 35 glycosylation sites of the glycoproteins in the composition are occupied.

Further, the methods and host cells herein can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-

glycosylation sites of the glycoproteins in the composition are occupied and which in further aspects have mammalian- or human-like *N*-glycans that lack fucose.

Further, the methods and yeast or filamentous fungus host cells are genetically engineered to produce mammalian-like or human-like *N*-glycans can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied and which in further aspects have mammalian- or human-like *N*-glycans that lack fucose.

In some aspects, the yeast or filamentous host cells genetically engineered to produce fucosylated mammalian- or human-like *N*-glycans can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied and which in further aspects have mammalian- or human-like *N*-glycans that have fucose.

The methods and host cells herein can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% antibody molecules in the compositions have both *N*-glycosylation sites occupied.

Further, the methods and host cells herein can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% antibody molecules in the compositions have both *N*-glycosylation sites occupied and the *N*-glycans lack fucose.

Further, the methods and yeast or filamentous fungus host cells herein can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% antibody molecules in the compositions have both *N*-glycosylation sites occupied and the *N*-glycans lack fucose.

Further, the methods and yeast or filamentous fungus host cells genetically engineered to produce mammalian-like or human-like *N*-glycans can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% antibody molecules in the compositions have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans that lack fucose. In some aspects, the yeast or filamentous host cells genetically engineered to produce fucosylated mammalian- or human-like *N*-glycans can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% antibody molecules in the compositions have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans with fucose.

Further provided is a glycoprotein composition comprising a plurality of antibodies wherein about 70% to about 99% of the intact antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have

a G2 structure, 5-17 mole% of the *N*-glycans have a Man5 structure, and 3-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier.

Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein about 70% to 99% of the intact antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier.

In particular embodiments, the antibodies comprise an antibody selected from the group consisting of anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, and anti-CD20 antibody.

Further provided are compositions comprising one or more glycoproteins produced by the host cells and methods described herein.

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₎Man₃GlcNAc₂; Gal₍₁₋₄₎GlcNAc₍₁₋₄₎Man₃GlcNAc₂; NANA₍₁₋₄₎Gal₍₁₋₄₎GlcNAc₍₁₋₄₎Man₃GlcNAc₂.

In particular embodiments, the glycoprotein compositions provided herein comprise glycoproteins having at least one hybrid *N*-glycan selected from the group consisting of GlcNAcMan₃GlcNAc₂; GalGlcNAcMan₃GlcNAc₂; NANAGalGlcNAcMan₃GlcNAc₂; GlcNAcMan₅GlcNAc₂; GalGlcNAcMan₅GlcNAc₂; and NANAGalGlcNAcMan₅GlcNAc₂. 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.

In particular embodiments, the glycoprotein compositions provided herein comprise glycoproteins having at least one complex *N*-glycan selected from the group consisting of GlcNAc₂Man₃GlcNAc₂; GalGlcNAc₂Man₃GlcNAc₂; Gal₂GlcNAc₂Man₃GlcNAc₂; NANAGal₂GlcNAc₂Man₃GlcNAc₂; and NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂. 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.

In particular embodiments, the *N*-glycan is fucosylated. In general, the fucose is in an α 1,3-linkage with the GlcNAc at the reducing end of the *N*-glycan, an α 1,6-linkage with the

GlcNAc at the reducing end of the *N*-glycan, an α 1,2-linkage with the Gal at the non-reducing end of the *N*-glycan, an α 1,3-linkage with the GlcNAc at the non-reducing end of the *N*-glycan, or an α 1,4-linkage with a GlcNAc at the non-reducing end of the *N*-glycan.

Therefore, in particular aspects of the above the glycoprotein compositions, the glycoform is in an α 1,3-linkage or α 1,6-linkage fucose to produce a glycoform selected from the group consisting of GlcNAcMan₅GlcNAc₂(Fuc), GlcNAcMan₃GlcNAc₂(Fuc), GlcNAc₂Man₃GlcNAc₂(Fuc), GalGlcNAc₂Man₃GlcNAc₂(Fuc), Gal₂GlcNAc₂Man₃GlcNAc₂(Fuc), NANA Gal₂GlcNAc₂Man₃GlcNAc₂(Fuc), and NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂(Fuc); in an α 1,3-linkage or α 1,4-linkage fucose to produce a glycoform selected from the group consisting of GlcNAc(Fuc)Man₅GlcNAc₂, GlcNAc(Fuc)Man₃GlcNAc₂, GlcNAc₂(Fuc₁₋₂)Man₃GlcNAc₂, GalGlcNAc₂(Fuc₁₋₂)Man₃GlcNAc₂, Gal₂GlcNAc₂(Fuc₁₋₂)Man₃GlcNAc₂, NANA Gal₂GlcNAc₂(Fuc₁₋₂)Man₃GlcNAc₂, and NANA₂Gal₂GlcNAc₂(Fuc₁₋₂)Man₃GlcNAc₂; or in an α 1,2-linkage fucose to produce a glycoform selected from the group consisting of Gal(Fuc)GlcNAc₂Man₃GlcNAc₂, Gal₂(Fuc₁₋₂)GlcNAc₂Man₃GlcNAc₂, NANA Gal₂(Fuc₁₋₂)GlcNAc₂Man₃GlcNAc₂, and NANA₂Gal₂(Fuc₁₋₂)GlcNAc₂Man₃GlcNAc₂.

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

In further aspects, the glycoproteins comprise high mannose *N*-glycans, including but not limited to, Man₅GlcNAc₂, or *N*-glycans that consist of the Man₃GlcNAc₂ *N*-glycan structure.

Definitions

As used herein, the terms “*N*-glycan” and “glycoform” are used interchangeably and refer to an *N*-linked oligosaccharide, for example, 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 co-translationally in the lumen of the ER and continues post-translationally in the Golgi apparatus for *N*-linked glycoproteins.

N-glycans have a common pentasaccharide core of Man₃GlcNAc₂ (“Man” refers to mannose; “Glc” refers to glucose; and “NAc” refers to *N*-acetyl; GlcNAc refers to *N*-acetylglucosamine). Usually, *N*-glycan structures are presented with the non-reducing end to the left and the reducing end to the right. The reducing end of the *N*-glycan is the end that is attached to the Asn residue comprising the glycosylation site on the protein. *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 $\text{Man}_3\text{GlcNAc}_2$ ("Man3") core structure which is also referred to as the "triannose 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."

With respect to complex *N*-glycans, the terms "G-2", "G-1", "G0", "G1", "G2", "A1", and "A2" mean the following. "G-2" refers to an *N*-glycan structure that can be characterized as $\text{Man}_3\text{GlcNAc}_2$; the term "G-1" refers to an *N*-glycan structure that can be characterized as $\text{GlcNAcMan}_3\text{GlcNAc}_2$; the term "G0" refers to an *N*-glycan structure that can be characterized as $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "G1" refers to an *N*-glycan structure that can be characterized as $\text{GalGlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "G2" refers to an *N*-glycan structure that can be characterized as $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; the term "A1" refers to an *N*-glycan structure that can be characterized as $\text{NANAGal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$; and, the term "A2" refers to an *N*-glycan structure that can be characterized as $\text{NANA}_2\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$. Unless otherwise indicated, the terms "G-2", "G-1", "G0", "G1", "G2", "A1", and "A2" refer to *N*-glycan species that lack fucose attached to the GlcNAc residue at the reducing end of the *N*-glycan. When the term includes an "F", the "F" indicates that the *N*-glycan species contains a fucose residue on the GlcNAc residue at the reducing end of the *N*-glycan. For example, G0F, G1F, G2F, A1F, and A2F all indicate that the *N*-glycan further includes a fucose residue attached to the GlcNAc residue at the reducing end of the *N*-glycan. Lower eukaryotes such as yeast and filamentous fungi do not normally produce *N*-glycans that produce fucose.

With respect to multiantennary *N*-glycans, the term "multiantennary *N*-glycan" refers to *N*-glycans that further comprise a GlcNAc residue on the mannose residue comprising the non-reducing end of the 1,6 arm or the 1,3 arm of the *N*-glycan or a GlcNAc residue on each of the mannose residues comprising the non-reducing end of the 1,6 arm and the 1,3 arm of the *N*-glycan. Thus, multiantennary *N*-glycans can be characterized by the formulas GlcNAc_2 -

4)Man₃GlcNAc₂, Gal(1-4)GlcNAc(2-4)Man₃GlcNAc₂, or NANA(1-4)Gal(1-4)GlcNAc(2-4)Man₃GlcNAc₂. The term "1-4" refers to 1, 2, 3, or 4 residues.

With respect to bisected *N*-glycans, the term "bisected *N*-glycan" refers to *N*-glycans in which a GlcNAc residue is linked to the mannose residue at the reducing end of the *N*-glycan. A bisected *N*-glycan can be characterized by the formula GlcNAc₃Man₃GlcNAc₂ wherein each mannose residue is linked at its non-reducing end to a GlcNAc residue. In contrast, when a multiantennary *N*-glycan is characterized as GlcNAc₃Man₃GlcNAc₂, the formula indicates that two GlcNAc residues are linked to the mannose residue at the non-reducing end of one of the two arms of the *N*-glycans and one GlcNAc residue is linked to the mannose residue at the non-reducing end of the other arm of the *N*-glycan.

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).

As used herein, the term "glycoprotein" refers to any protein having one or more *N*-glycans attached thereto. Thus, the term refers both to proteins that are generally recognized in the art as a glycoprotein and to proteins which have been genetically engineered to contain one or more *N*-linked glycosylation sites.

As used herein, a "humanized glycoprotein" or a "human-like glycoprotein" refers alternatively to a protein having attached thereto *N*-glycans having fewer than four mannose residues, and synthetic glycoprotein intermediates (which are also useful and can be manipulated further *in vitro* or *in vivo*) having at least five mannose residues. Preferably, glycoproteins produced according to the invention contain at least 30 mole %, preferably at least 40 mole % and more preferably 50, 60, 70, 80, 90, or even 100 mole % of the Man₅GlcNAc₂ intermediate, at least transiently. This may be achieved, *e.g.*, by engineering a host cell of the invention to express a "better", *i.e.*, a more efficient glycosylation enzyme. For example, a mannosidase is selected such that it will have optimal activity under the conditions present at the site in the host cell where proteins are glycosylated and is introduced into the host cell preferably by targeting the enzyme to a host cell organelle where activity is desired.

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. Preferred host cells are yeasts and fungi.

When referring to "mole percent" of a glycan present in a preparation of a glycoprotein, the term means the molar percent of a particular glycan present in the pool of *N*-linked oligosaccharides released when the protein preparation is treated with PNGase and then quantified by a method that is not affected by glycoform composition, (for instance, labeling a PNGase released glycan pool with a fluorescent tag such as 2-aminobenzamide and then separating by high performance liquid chromatography or capillary electrophoresis and then quantifying glycans by fluorescence intensity). For example, 50 mole percent GlcNAc₂Man₃GlcNAc₂Gal₂NANA₂ means that 50 percent of the released glycans are GlcNAc₂Man₃GlcNAc₂Gal₂NANA₂ and the remaining 50 percent are comprised of other *N*-linked oligosaccharides. In embodiments, the mole percent of a particular glycan in a preparation of glycoprotein will be between 20% and 100%, preferably above 25%, 30%, 35%, 40% or 45%, more preferably above 50%, 55%, 60%, 65% or 70% and most preferably above 75%, 80% 85%, 90% or 95%.

The term "operably 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.

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 operably 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.

The term "transfect", "transfection", "transfecting" and the like refer to the introduction of a heterologous nucleic acid into eukaryote cells, both higher and lower eukaryote cells. Historically, the term "transformation" has been used to describe the introduction of a nucleic acid into a yeast or fungal cell; however, herein the term "transfection" is used to refer to the introduction of a nucleic acid into any eukaryote cell, including yeast and fungal cells.

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

The term "lower eukaryotic cells" includes yeast and filamentous fungi. 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*. *Pichia* sp., any *Saccharomyces* sp., *Hansenula polymorpha*, any *Kluyveromyces* sp., *Candida albicans*, any *Aspergillus* sp., *Trichoderma reesei*, *Chrysosporium lucknowense*, any *Fusarium* sp. and *Neurospora crassa*.

As used herein, the terms "antibody," "immunoglobulin," "immunoglobulins" and "immunoglobulin molecule" are used interchangeably. Each immunoglobulin molecule has a unique structure that allows it to bind its specific antigen, but all immunoglobulins have the same overall structure as described herein. The basic immunoglobulin structural unit is known to comprise a tetramer of subunits. Each tetramer has two identical pairs of polypeptide chains, each pair having one "light" chain (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The amino-terminal portion of each chain includes a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The carboxy-terminal portion of each chain defines a constant region primarily responsible for effector function. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, and define the antibody's isotype as IgG, IgM, IgA, IgD, and IgE, respectively.

The light and heavy chains are subdivided into variable regions and constant regions (See generally, Fundamental Immunology (Paul, W., ed., 2nd ed. Raven Press, N.Y., 1989), Ch. 7. The variable regions of each light/heavy chain pair form the antibody binding site. Thus, an intact antibody has two binding sites. Except in bifunctional or bispecific antibodies, the two binding sites are the same. The chains all exhibit the same general structure of relatively conserved framework regions (FR) joined by three hypervariable regions, also called complementarity determining regions or CDRs. The CDRs from the two chains of each pair are aligned by the framework regions, enabling binding to a specific epitope. The terms include naturally occurring forms, as well as fragments and derivatives. Included within the scope of the term are classes of immunoglobulins (Igs), namely, IgG, IgA, IgE, IgM, and IgD. Also included within the scope of the terms are the subtypes of IgGs, namely, IgG1, IgG2, IgG3, and IgG4. The term is used in the broadest sense and includes single monoclonal antibodies (including agonist

and antagonist antibodies) as well as antibody compositions which will bind to multiple epitopes or antigens. The terms specifically cover monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (for example, bispecific antibodies), and antibody fragments so long as they contain or are modified to contain at least the portion of the CH2 domain of the heavy chain immunoglobulin constant region which comprises an *N*-linked glycosylation site of the CH2 domain, or a variant thereof. Included within the terms are molecules comprising only the Fc region, such as immunoadhesions (U.S. Published Patent Application No. 2004/0136986; the disclosure of which is incorporated herein by reference), Fc fusions, and antibody-like molecules.

The term "Fc fragment" refers to the 'fragment crystallized' C-terminal region of the antibody containing the CH2 and CH3 domains. The term "Fab fragment" refers to the 'fragment antigen binding' region of the antibody containing the VH, CH1, VL and CL domains.

The term "monoclonal antibody" (mAb) as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each mAb is directed against a single determinant on the antigen. In addition to their specificity, monoclonal antibodies are advantageous in that they can be produced, for example, by hybridoma culture, uncontaminated by other immunoglobulins. The term "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.*, (1975) *Nature*, 256:495, or may be made by recombinant DNA methods (*See*, for example, U.S. Pat. No. 4,816,567; the disclosure of which is incorporated herein by reference).

The term "fragments" within the scope of the terms "antibody" or "immunoglobulin" include those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation and those produced recombinantly, so long as the fragment remains capable of specific binding to a target molecule. Among such fragments are Fc, Fab, Fab', Fv, F(ab')₂, and single chain Fv (scFv) fragments. Hereinafter, the term "immunoglobulin" also includes the term "fragments" as well.

Immunoglobulins further include immunoglobulins or fragments that have been modified in sequence but remain capable of specific binding to a target molecule, including: interspecies chimeric and humanized antibodies; antibody fusions; heteromeric antibody

complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (*See*, for example, *Intracellular Antibodies: Research and Disease Applications*, (Marasco, ed., Springer-Verlag New York, Inc., 1998).

The term "catalytic antibody" refers to immunoglobulin molecules that are capable of catalyzing a biochemical reaction. Catalytic antibodies are well known in the art and have been described in U.S. Patent Application Nos. 7,205,136; 4,888,281; 5,037,750 to Schochetman *et al.*, U.S. Patent Application Nos. 5,733,757; 5,985,626; and 6,368,839 to Barbas, III *et al.* (the disclosures of which are all incorporated herein by reference).

The interaction of antibodies and antibody-antigen complexes with cells of the immune system and the variety of responses, including antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC), clearance of immunocomplexes (phagocytosis), antibody production by B cells and IgG serum half-life are defined respectively in the following: Daeron *et al.*, *Annu. Rev. Immunol.* 15: 203-234 (1997); Ward and Ghetie, *Therapeutic Immunol.* 2:77-94 (1995); Cox and Greenberg, *Semin. Immunol.* 13: 339-345 (2001); Heyman, *Immunol. Lett.* 88:157-161 (2003); and Ravetch, *Curr. Opin. Immunol.* 9: 121-125 (1997).

As used herein, the term "consisting essentially of" will be understood to imply the inclusion of a stated integer or group of integers; while excluding modifications or other integers which would materially affect or alter the stated integer. With respect to species of *N*-glycans, the term "consisting essentially of" a stated *N*-glycan will be understood to include the *N*-glycan whether or not that *N*-glycan is fucosylated at the *N*-acetylglucosamine (GlcNAc) which is directly linked to the asparagine residue of the glycoprotein.

As used herein, the term "predominantly" or variations such as "the predominant" or "which is predominant" will be understood to mean the glycan species that has the highest mole percent (%) of total neutral *N*-glycans after the glycoprotein has been treated with PNGase and released glycans analyzed by mass spectroscopy, for example, MALDI-TOF MS or HPLC. In other words, the phrase "predominantly" is defined as an individual entity, such as a specific glycoform, is present in greater mole percent than any other individual entity. For example, if a composition consists of species A at 40 mole percent, species B at 35 mole percent and species C at 25 mole percent, the composition comprises predominantly species A, and species B would be the next most predominant species. Some host cells may produce compositions comprising neutral *N*-glycans and charged *N*-glycans such as mannosylphosphate. Therefore, a composition of glycoproteins can include a plurality of charged and uncharged or neutral *N*-glycans. In the present invention, it is within the context of the total plurality of neutral *N*-glycans in the composition in which the predominant *N*-glycan determined. Thus, as used herein, "predominant *N*-glycan" means that of the total plurality of neutral *N*-glycans in the composition, the predominant *N*-glycan is of a particular structure.

As used herein, the term “essentially free of” a particular sugar residue, such as fucose, or galactose and the like, is used to indicate that the glycoprotein composition is substantially devoid of *N*-glycans which contain such residues. Expressed in terms of purity, essentially free means that the amount of *N*-glycan structures containing such sugar residues does not exceed 10%, and preferably is below 5%, more preferably below 1%, most preferably below 0.5%, wherein the percentages are by weight or by mole percent. Thus, substantially all of the *N*-glycan structures in a glycoprotein composition according to the present invention are free of, for example, fucose, or galactose, or both.

As used herein, a glycoprotein composition “lacks” or “is lacking” a particular sugar residue, such as fucose or galactose, when no detectable amount of such sugar residue is present on the *N*-glycan structures at any time. For example, in preferred embodiments of the present invention, the glycoprotein compositions are produced by lower eukaryotic organisms, as defined above, including yeast (for example, *Pichia* sp.; *Saccharomyces* sp.; *Kluyveromyces* sp.; *Aspergillus* sp.), and will “lack fucose,” because the cells of these organisms do not have the enzymes needed to produce fucosylated *N*-glycan structures. Thus, the term “essentially free of fucose” encompasses the term “lacking fucose.” However, a composition may be “essentially free of fucose” even if the composition at one time contained fucosylated *N*-glycan structures or contains limited, but detectable amounts of fucosylated *N*-glycan structures as described above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 A-H shows the genealogy of *P. pastoris* strain **YGLY13992** (**Figure 1F**) and strain **YGLY14401** (**Figure 1G**) beginning from wild-type strain **NRRL-Y11430** (**Figure 1A**).

Figure 2 shows a map of plasmid pGLY6301 encoding the *LmSTT3D* ORF under the control of the *Pichia pastoris* alcohol oxidase I (AOX1) promoter and *S. cerevisiae* *CYC* transcription termination sequence. The plasmid is a roll-in vector that targets the *URA6* locus. The selection of transformants uses arsenic resistance encoded by the *S. cerevisiae* *ARR3* ORF under the control of the *P. pastoris* *RPL10* promoter and *S. cerevisiae* *CYC* transcription termination sequence.

Figure 3 shows a map of plasmid pGLY6294 encoding the *LmSTT3D* ORF under the control of the *P. pastoris* *GAPDH* promoter and *S. cerevisiae* *CYC* transcription termination sequence. The plasmid is a KINKO vector that targets the *TRP1* locus: the 3' end of the *TRP1* ORF is adjacent to the *P. pastoris* *ALG3* transcription termination sequence. The selection of transformants uses nourseothricin resistance encoded by the *Streptomyces noursei* nourseothricin acetyltransferase (*NAT*) ORF under the control of the *Ashbya gossypii* *TEF1* promoter (PTEF) and *Ashbya gossypii* *TEF1* termination sequence (TTEF).

Figure 4 shows a map of plasmid pGLY6. Plasmid pGLY6 is an integration vector that targets the *URA5* locus and contains a nucleic acid molecule comprising the *S. cerevisiae* invertase gene or transcription unit (ScSUC2) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris URA5* gene (PpURA5-5') and on the other side by a nucleic acid molecule comprising the a nucleotide sequence from the 3' region of the *P. pastoris URA5* gene (PpURA5-3').

Figure 5 shows a map of plasmid pGLY40. Plasmid pGLY40 is an integration vector that targets the *OCH1* locus and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (*lacZ* repeat) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *OCH1* gene (PpOCH1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *OCH1* gene (PpOCH1-3').

Figure 6 shows a map of plasmid pGLY43a. Plasmid pGLY43a is an integration vector that targets the *BMT2* locus and contains a nucleic acid molecule comprising the *K. lactis* UDP-*N*-acetylglucosamine (UDP-GlcNAc) transporter gene or transcription unit (KlGlcNAc Transp.) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (*lacZ* repeat). The adjacent genes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *BMT2* gene (PpPBS2-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *BMT2* gene (PpPBS2-3').

Figure 7 shows a map of plasmid pGLY48. Plasmid pGLY48 is an integration vector that targets the *MNN4L1* locus and contains an expression cassette comprising a nucleic acid molecule encoding the mouse homologue of the UDP-GlcNAc transporter (MmGlcNAc Transp.) open reading frame (ORF) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter (PpGAPDH Prom) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* termination sequence (ScCYC TT) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by *lacZ* repeats (*lacZ* repeat) and in which the expression cassettes together are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris MNN4L1* gene (PpMNN4L1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4L1* gene (PpMNN4L1-3').

Figure 8 shows as map of plasmid pGLY45. Plasmid pGLY45 is an integration vector that targets the *PNO1/MNN4* loci contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit (PpURA5) flanked by nucleic acid molecules comprising *lacZ* repeats (*lacZ* repeat) which in turn is flanked on one side by a nucleic acid

molecule comprising a nucleotide sequence from the 5' region of the *PNO1* gene (PpPNO1-5') and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4* gene (PpMNN4-3').

Figure 9 shows a map of plasmid pGLY1430. Plasmid pGLY1430 is a KINKO integration vector that targets the *ADE1* locus without disrupting expression of the locus and contains in tandem four expression cassettes encoding (1) the human GlcNAc transferase I catalytic domain (codon optimized) fused at the *N*-terminus to *P. pastoris* *SEC12* leader peptide (CO-NA10), (2) mouse homologue of the UDP-GlcNAc transporter (MmTr), (3) the mouse mannosidase IA catalytic domain (FB) fused at the *N*-terminus to *S. cerevisiae* *SEC12* leader peptide (FB8), and (4) the *P. pastoris* *URA5* gene or transcription unit (PpURA5) flanked by lacZ repeats (lacZ). All flanked by the 5' region of the *ADE1* gene and ORF (*ADE1* 5' and ORF) and the 3' region of the *ADE1* gene (PpADE1-3'). PpPMA1 prom is the *P. pastoris* *PMA1* promoter; PpPMA1 TT is the *P. pastoris* *PMA1* termination sequence; SEC4 is the *P. pastoris* *SEC4* promoter; OCH1 TT is the *P. pastoris* *OCH1* termination sequence; ScCYC TT is the *S. cerevisiae* *CYC* termination sequence; PpOCH1 Prom is the *P. pastoris* *OCH1* promoter; PpALG3 TT is the *P. pastoris* *ALG3* termination sequence; and PpGAPDH is the *P. pastoris* *GADPH* promoter.

Figure 10 shows a map of plasmid pGLY582. Plasmid pGLY582 is an integration vector that targets the *HIS1* locus and contains in tandem four expression cassettes encoding (1) the *S. cerevisiae* UDP-glucose epimerase (ScGAL10), (2) the human galactosyltransferase I (hGalT) catalytic domain fused at the *N*-terminus to the *S. cerevisiae* *KRE2-s* leader peptide (33), (3) the *P. pastoris* *URA5* gene or transcription unit (PpURA5) flanked by lacZ repeats (lacZ repeat), and (4) the *D. melanogaster* UDP-galactose transporter (DmUGT). All flanked by the 5' region of the *HIS1* gene (PpHIS1-5') and the 3' region of the *HIS1* gene (PpHIS1-3'). PMA1 is the *P. pastoris* *PMA1* promoter; PpPMA1 TT is the *P. pastoris* *PMA1* termination sequence; GAPDH is the *P. pastoris* *GADPH* promoter and ScCYC TT is the *S. cerevisiae* *CYC* termination sequence; PpOCH1 Prom is the *P. pastoris* *OCH1* promoter and PpALG12 TT is the *P. pastoris* *ALG12* termination sequence.

Figure 11 shows a map of plasmid pGLY167b. Plasmid pGLY167b is an integration vector that targets the *ARG1* locus and contains in tandem three expression cassettes encoding (1) the *D. melanogaster* mannosidase II catalytic domain (codon optimized) fused at the *N*-terminus to *S. cerevisiae* *MNN2* leader peptide (CO-KD53), (2) the *P. pastoris* *HIS1* gene or transcription unit, and (3) the rat *N*-acetylglucosamine (GlcNAc) transferase II catalytic domain (codon optimized) fused at the *N*-terminus to *S. cerevisiae* *MNN2* leader peptide (CO-TC54). All flanked by the 5' region of the *ARG1* gene (PpARG1-5') and the 3' region of the *ARG1* gene (PpARG1-3'). PpPMA1 prom is the *P. pastoris* *PMA1* promoter; PpPMA1 TT is the *P. pastoris* *PMA1* termination sequence; PpGAPDH is the *P. pastoris* *GADPH* promoter;

ScCYC TT is the *S. cerevisiae* CYC termination sequence; PpOCH1 Prom is the *P. pastoris* OCH1 promoter; and PpALG12 TT is the *P. pastoris* ALG12 termination sequence.

Figure 12 shows a map of plasmid pGLY3411 (pSH1092). Plasmid pGLY3411 (pSH1092) is an integration vector that contains the expression cassette comprising the *P.*

pastoris URA5 gene or transcription unit (PpURA5) flanked by *lacZ* repeats (*lacZ* repeat) flanked on one side with the 5' nucleotide sequence of the *P. pastoris* BMT4 gene (PpPBS4 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris* BMT4 gene (PpPBS4 3').

Figure 13 shows a map of plasmid pGLY3419 (pSH1110). Plasmid pGLY3430 (pSH1115) is an integration vector that contains an expression cassette comprising the *P.*

pastoris URA5 gene or transcription unit (PpURA5) flanked by *lacZ* repeats (*lacZ* repeat) flanked on one side with the 5' nucleotide sequence of the *P. pastoris* BMT1 gene (PBS1 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris* BMT1 gene (PBS1 3').

Figure 14 shows a map of plasmid pGLY3421 (pSH1106). Plasmid pGLY4472 (pSH1186) contains an expression cassette comprising the *P. pastoris* URA5 gene or

transcription unit (PpURA5) flanked by *lacZ* repeats (*lacZ* repeat) flanked on one side with the 5' nucleotide sequence of the *P. pastoris* BMT3 gene (PpPBS3 5') and on the other side with the 3' nucleotide sequence of the *P. pastoris* BMT3 gene (PpPBS3 3').

Figure 15 shows a map of plasmid pGLY3673. Plasmid pGLY3673 is a KINKO integration vector that targets the *PRO1* locus without disrupting expression of the locus and contains expression cassettes encoding the *T. reesei* α -1,2-mannosidase catalytic domain fused at the N-terminus to *S. cerevisiae* α MATpre signal peptide (α MATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell.

Figure 16 shows a map of pGLY6833 encoding the light and heavy chains of an anti-Her2 antibody. The plasmid is a roll-in vector that targets the *TRP2* locus. The ORFs encoding the light and heavy chains are under the control of a *P. pastoris* AOX1 promoter and the *P. pastoris* CIT1 3UTR transcription termination sequence. Selection of transformants uses zeocin resistance encoded by the zeocin resistance protein (Zeocin^R) ORF under the control of the *P. pastoris* TEF1 promoter and *S. cerevisiae* CYC termination sequence.

Figure 17 shows a map of pGLY6564 encoding the light and heavy chains of an anti-RSV antibody. The plasmid is a roll-in vector that targets the *TRP2* locus. The ORF encoding the heavy chain is under the control of a *P. pastoris* AOX1 promoter and the *S. cerevisiae* CYC transcription termination sequence. The ORF encoding the light chain is under the control of a *P. pastoris* AOX1 promoter and the *P. pastoris* AOX1 transcription termination sequence. Selection of transformants uses zeocin resistance encoded by the zeocin resistance protein (Zeocin^R) ORF under the control of the *P. pastoris* TEF1 promoter and *S. cerevisiae* CYC termination sequence.

Figure 18 shows the percent *N*-glycosylation site occupancy of anti-Her2 and anti-RSV antibodies produced in control strains versus strains in which the *LmSTT3D* is constitutively expressed (GAPDH promoter) or inducibly expressed (AOX1 promoter).

Figure 19 shows a comparison of *N*-glycosylation site occupancy of the anti-Her2 antibody produced in strain **YGLY13992** and strain **YGLY17351** compared to *N*-glycosylation site occupancy of a commercially available anti-Her2 antibody produced in CHO cells (HERCEPTIN). Strain **YGLY13992** does not include an expression cassette encoding the *LmSTT3D* whereas strain **YGLY17351** includes an expression cassette encoding the *LmSTT3* under the control of the inducible *PpAOX1* promoter.

Figure 20 shows the percent *N*-glycosylation site occupancy of anti-Her2 antibodies produced in strain **YGLY17351** grown in various bioreactors was consistent regardless of bioreactor scale.

Figure 21 shows the results of a CE and Q-TOF analysis of a commercial lot of anti-Her2 antibody (HERCEPTIN).

Figure 22 shows the results of a CE and Q-TOF analysis for the same commercial lot as used for **Figure 21** but after treatment with PNGase F for a period of time.

Figure 23 A-D shows the genealogy of *P. pastoris* strain **YGLY12900** beginning from **YGLY7961**.

Figure 24 shows a map of plasmid pGLY2456. Plasmid pGLY2456 is a KINKO integration vector that targets the *TRP2* locus without disrupting expression of the locus and contains six expression cassettes encoding (1) the mouse CMP-sialic acid transporter codon optimized (CO mCMP-Sia Transp), (2) the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase codon optimized (CO hGNE), (3) the *Pichia pastoris* *ARG1* gene or transcription unit, (4) the human CMP-sialic acid synthase codon optimized (CO hCMP-NANA S), (5) the human *N*-acetylneuraminase-9-phosphate synthase codon optimized (CO hSIAP S), and, (6) the mouse α -2,6-sialyltransferase catalytic domain codon optimized fused at the *N*-terminus to *S. cerevisiae* *KRE2* leader peptide (comST6-33). All flanked by the 5' region of the *TRP2* gene and ORF (PpTRP2 5') and the 3' region of the *TRP2* gene (PpTRP2-3'). PpPMA1 prom is the *P. pastoris* *PMA1* promoter; PpPMA1 TT is the *P. pastoris* *PMA1* termination sequence; CYC TT is the *S. cerevisiae* *CYC* termination sequence; PpTEF Prom is the *P. pastoris* *TEF1* promoter; PpTEF TT is the *P. pastoris* *TEF1* termination sequence; PpALG3 TT is the *P. pastoris* *ALG3* termination sequence; and pGAP is the *P. pastoris* *GAPDH* promoter.

Figure 25 shows a map of plasmid pGLY5048. Plasmid pGLY5048 is an integration vector that targets the *STE13* locus and contains expression cassettes encoding (1) the *T. reesei* α -1,2-mannosidase catalytic domain fused at the *N*-terminus to *S. cerevisiae* α MATpre signal peptide (α MATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris* *URA5* gene or transcription unit.

Figure 26 shows a map of plasmid pGLY5019. Plasmid pGLY5019 is an integration vector that targets the *DAP2* locus and contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NAT^R) ORF operably linked to the *Ashbya gossypii TEF1* promoter and *A. gossypii TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris DAP2* gene and on the other side with the 3' nucleotide sequence of the *P. pastoris DAP2* gene.

Figure 27 shows a plasmid map of pGLY5085. Plasmid pGLY5085 is a KINKO plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY2456 except that the *P. pastoris ARG1* gene has been replaced with an expression cassette encoding hygromycin resistance (HygR) and the plasmid targets the *P. pastoris TRP5* locus. The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP5* gene ending at the stop codon followed by a *P. pastoris ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP5* gene.

Figure 28 shows a plasmid map of pGLY7240. The plasmid is an integration vector that targets the *TRP2* locus and contains an ORF encoding the zeocin resistance protein (Zeocin^R) under the control of the *P. pastoris TEF1* promoter and *S. cerevisiae CYC* termination sequence. The plasmid encodes the GM-CSF/CWP1 fusion protein operably linked at the 5' end to the *Pichia pastoris AOX1* promoter and at the 3' end to the *S. cerevisiae CYC* transcription termination sequence.

Figure 29 shows a Western blot of GM-CSF produced in strain YGLY16349, which co-expresses LmSTT3D, that the majority of GM-CSF (Lanes 2-8) is glycosylated with 2N-linked sites in contrast to the control strain (YGLY15560, lane 9) where GM-CSF is predominantly N-glycosylated with 1 site along with the minor portions of 2 N sites and non-glycosylated.

Figure 30 shows a Q-TOP analysis of GM-CSF expressed from YGLY15560 (A) and YGLY16349 (B), respectively. Non-glycosylated GM-CSF was not detected.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for producing a therapeutic glycoprotein in a host cell in which the *N*-glycosylation site occupancy of the glycoprotein is increased over the *N*-glycosylation site occupancy of the same glycoprotein produced in a host cell not modified as disclosed herein. When the present invention is practiced in a lower eukaryote host cell, e.g., yeast host cells or filamentous fungal host cells, the *N*-glycosylation site occupancy of recombinant glycoproteins produced in the host cell is the same as or more similar to the *N*-glycosylation site occupancy of the same recombinant glycoproteins produced in mammalian or human host cells.

To increase the *N*-glycosylation site occupancy on a glycoprotein produced in a recombinant host cell, at least one nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase, which in particular embodiments at least one is capable of functionally suppressing a lethal mutation of one or more essential subunits comprising the endogenous host cell hetero-oligomeric oligosaccharyltransferase (OTase) complex, is overexpressed in the recombinant host cell either before or simultaneously with the expression of the glycoprotein in the host cell.

The *Leishmania major* STT3A protein, *Leishmania major* STT3B protein, and *Leishmania major* STT3D protein, are single-subunit oligosaccharyltransferases that have been shown to suppress the lethal phenotype of a deletion of the *STT3* locus in *Saccharomyces cerevisiae* (Naseb *et al.*, Molec. Biol. Cell 19: 3758-3768 (2008)). Naseb *et al.* (ibid.) further showed that the *Leishmania major* STT3D protein could suppress the lethal phenotype of a deletion of the *WBP1*, *OST1*, *SWP1*, or *OST2* loci. Hese *et al.* (Glycobiology 19: 160-171 (2009)) teaches that the *Leishmania major* STT3A (STT3-1), STT3B (STT3-2), and STT3D (STT3-4) proteins can functionally complement deletions of the *OST2*, *SWP1*, and *WBP1* loci. The *Leishmania major* STT3D (*LmSTT3D*) protein is a heterologous single-subunit oligosaccharyltransferases that is capable of suppressing a lethal phenotype of a Δ *stt3* mutation and at least one lethal phenotype of a Δ *wbp1*, Δ *ost1*, Δ *swp1*, and Δ *ost2* mutation that is shown in the examples herein to be capable of enhancing the *N*-glycosylation site occupancy of heterologous glycoproteins, for example antibodies, produced by the host cell.

The one or more heterologous single-subunit oligosaccharyltransferases is/are overexpressed constitutively or inducibly in the presence of the proteins comprising the host cell's endogenous OTase complex, including the host cell's STT3 protein. An expression cassette encoding each heterologous single-subunit oligosaccharyltransferase gene can either be integrated into any site within the host cell genome or located in the extrachromosomal space of the host cell, i.e., autonomously replicating genetic elements such as plasmids, viruses, 2 μ m plasmid, minichromosomes, and the like. In general, the heterologous single-subunit oligosaccharyltransferases are provided to the host cell in expression cassettes, each comprising a nucleic acid molecule encoding a single-subunit oligosaccharyltransferase open reading frame (ORF) operably linked to a heterologous constitutive or inducible promoter and other heterologous transcriptional or translational regulatory elements suitable for expressing heterologous proteins in a particular host cell. One or more copies of each expression cassette is/are integrated into one or more locations in the host cell's genome either by site-specific targeting of a particular locus for integration or randomly integrating the expression cassette into the genome. The locus for targeted integration can be selected based upon the suitability of the locus for ectopic constitutive or inducible expression of the single-subunit oligosaccharyltransferase in the expression cassette. Methods for integrating heterologous

nucleic acid molecules into a host cell genome by techniques such as single- and double-crossover homologous recombination and the like are well known in the art (*See* for example, U.S. Published Application No. 20090124000 and International Published Application No. WO2009085135, the disclosures of which are incorporated herein by reference). Alternatively, or in addition to integrating one or more copies of the expression cassette into the host cell genome, one or more copies of the expression cassette are located in the extrachromosomal space of the host cell using a 2 μ plasmid, viral vector, mini-chromosome, or other genetic vector that replicates autonomously.

While the present invention has been exemplified herein with *Pichia pastoris* host cells genetically engineered to produce mammalian or human-like glycosylation patterns comprising complex *N*-glycans, the present invention to increase the overall amount of *N*-glycosylation site occupancy of the glycoproteins produced in the host cell compared to that of glycoproteins produced in the host not modified as disclosed herein to express the single-subunit oligosaccharyltransferase gene can also be applied to *Pichia pastoris* host cells that are not genetically engineered to produce glycoproteins that have mammalian or human glycosylation patterns but instead express glycoproteins that have endogenous or wild-type glycosylation patterns, for example hypermannosylated *N*-glycosylation or when the host cell lacks alpha-1,6-mannosyltransferase (och1p) activity, high mannose *N*-glycosylation. The present invention can also be applied to other yeast or filamentous fungi or to plant or algal host cells, which express glycoproteins that have endogenous or wild-type glycosylation patterns, for example hypermannosylated *N*-glycosylation or when the host cell lacks alpha-1,6-mannosyltransferase (och1p) activity, high mannose *N*-glycosylation, or which have been genetically engineered to produce mammalian or human-like complex or hybrid *N*-glycans to increase the overall amount of *N*-glycosylation site occupancy of the glycoproteins produced in the host cell compared to that of glycoproteins produced in the host not modified as disclosed herein to express the single-subunit oligosaccharyltransferase gene. The present invention can also be applied to mammalian expression systems to increase the overall *N*-glycosylation site occupancy of glycoproteins that have more than two *N*-linked sites compared to that of glycoproteins produced in the host cell not modified as disclosed herein to express the single-subunit oligosaccharyltransferase gene.

The OTase complex of animals, plants, and fungi is a hetero-oligomeric protein complex. In the well-studied model organism *Saccharomyces cerevisiae*, the OTase complex currently appears to consist of at least eight different subunits: Ost1p, Ost2p, Wbp1, Stt3p, Swp1p, Ost4p, Ost5p, and Ost3p/Ost6p (Silberstein & Gilmore, FASEB J. 10: 849–858 (1996); Knauer & Lehle, Biochim. Biophys. Acta. 1426: 259–273 (1999); Dempski & Imperiali, Curr. Opin. Chem. Biol. 6: 844–850 (2002); Yan & Lennarz, J. Biol. Chem. 277: 47692–47700 (2005); Kelleher & Gilmore, Glycobiol. 16:47R–62R (2006); Weerapana & Imperiali, Glycobiol. 16: 91R–101R (2006)). In *Pichia pastoris*, the OTase complex appears to include at least Ost1p,

Ost2p, Ost3p, Ost4p, Ost6p, Wbp1, Swp1p, and Stt3p (See Shutter *et al.*, Nat. Biotechnol. 27: 561-566 (2009)).

It has been hypothesized that the *STT3* protein is the catalytic subunit in the OTase complex (Yan & Lennarz, J. Biol. Chem. 277: 47692-47700 (2002); Kelleher *et al.*, Mol. Cell. 12: 101-111 (2003); Nilsson *et al.*, J. Cell Biol. 161: 715-725 (2003)). Support for this hypothesis is from experiments showing that the prokaryotic homologue of yeast Stt3p is an active oligosaccharyltransferase in the absence of any other accessory proteins (Wacker *et al.*, Science. 298: 1790-1793 (2002); Kowarik *et al.*, Science 314: 1148-1150 (2006)). Proteins homologous to yeast Stt3p are encoded in almost all eukaryotic genomes (Kelleher & Gilmore, Glycobiol. 16:47R-62R (2006)). However, comparative genome analysis suggests that the composition of the OTase became increasingly complex during the evolutionary divergence of eukaryotes.

Single-subunit oligosaccharyltransferases are present in *Giardia* and kinetoplastids, whereas four subunit oligosaccharyltransferases consisting of the STT3, OST1, OST2, and WBP1 homologues are found in diplomonads, entamoebas, and apicomplexan species. Additionally, multiple forms of the putative STT3 proteins can be encoded in trypanosomatid genomes: three STT3 homologues are found in *Trypanosoma brucei* and four in *Leishmania major* (McConville *et al.*, Microbiol. Mol. Biol. Rev. 66: 122-154 (2002); Berriman *et al.*, Science. 309: 416-422 (2005); Ivens *et al.*, Science. 309: 436-442 (2005); Samuelson *et al.*, Proc. Natl. Acad. Sci. USA 102: 1548-1553 (2005); Kelleher & Gilmore, Glycobiol. 16:47R-62R (2006)).

In trypanosomatid parasites, *N*-linked glycosylation principally follows the pathway described for fungal or animal cells, but with different oligosaccharide structures transferred to protein (Parodi, Glycobiology 3: 193-199 (1993); McConville *et al.*, Microbiol. Mol. Biol. Rev. 66: 122-154 (2002)). It has been shown that, depending on the species, either Man₆GlcNAc₂ or Man₇GlcNAc₂ is the largest glycan transferred to protein in the genus *Leishmania* (Parodi, Glycobiology 3: 193-199 (1993)). Unlike the yeast and mammalian oligosaccharyltransferase that preferably use Glc₃Man₉GlcNAc₂, the trypanosome oligosaccharyltransferase is not selective and transfers different lipid-linked oligosaccharides at the same rate (Bosch *et al.*, J. Biol. Chem. 263:17360-17365 (1988)). Therefore, the simplest eukaryotic oligosaccharyltransferase is a single subunit STT3 protein, similar to the oligosaccharyltransferase found in bacterial *N*-glycosylation systems. Nasab *et al.*, Molecular Biology of the Cell 19: 3758-3768 (2008) expressed each of the four *Leishmania major* STT3 proteins individually in *Saccharomyces cerevisiae* and found that three of them, *LmSTT3A* protein, *LmSTT3B* protein, and *LmSTT3D* protein, were able to complement a deletion of the yeast *STT3* locus. In addition, *LmSTT3D* expression suppressed the lethal phenotype of single and double deletions in genes encoding various essential OTase subunits. The *LmSTT3* proteins

did not incorporate into the yeast OTase complex but instead formed a homodimeric enzyme, capable of replacing the endogenous, multimeric enzyme of the yeast cell. The results indicate that while these single-subunit oligosaccharyltransferases may resemble the prokaryotic enzymes, they use substrates typical for eukaryote glycosylation: that is, the N-X-S/T *N*-glycosylation recognition site and dolicholpyrophosphate-linked high mannose oligosaccharides.

N-glycosylation site occupancy in yeast has also been discussed in reports by, for example, Schultz and Aeby, *Molec. Cell. Proteomics* 8: 357-364 (2009); Hese *et al.*, *op. cit.*) and Nasab *et al.*, (*op. cit.*). Expression of the *Toxoplasma gondii* or *Trypanosoma cruzi* STT3 protein in *Saccharomyces cerevisiae* has been shown to complement the lethal phenotype of an *stt3* deletion (Shams-Eldin *et al.*, *Mol. Biochem. Parasitol.* 143: 6–11 (2005); Castro *et al.*, *Proc. Natl. Acad. Sci. USA* 103: 14756–14760 (2006) and while the *Trypanosoma cruzi* STT3 protein integrates into the yeast OTase complex the *Leishmania major* STT3 proteins appear to form homodimers instead (Nasab *et al.*, *op. cit.*). However, in these reports, the *LmSTT3D* protein had been tested for its functional suppression of a lethal mutation of the endogenous yeast *STT3* locus and other essential components of the yeast OTase complex in studies that measured *N*-glycosylation site occupancy of endogenous proteins. In addition, the yeast strains that were used in the studies produced glycoproteins that had a yeast glycosylation pattern, not a mammalian or human-like glycosylation pattern comprising hybrid or complex *N*-glycans.

In contrast to the above reports, in the present invention the open reading frame encoding a heterologous single-subunit oligosaccharyltransferase (as exemplified herein with the open reading frame encoding the *LmSTT3D*) protein is overexpressed constitutively or inducibly in the recombinant host cell in which the host cell further expresses the endogenous genes encoding the proteins comprising the host cell oligosaccharyltransferase (OTase) complex, which includes the expression of the endogenous host cell STT3 gene. Thus, the host cell expresses both the heterologous single-subunit oligosaccharyltransferase and the endogenous host cell OTase complex, including the endogenous host cell SST3 protein. Furthermore, with respect to recombinant yeast, filamentous fungus, algal, or plant host cells, the host cells can further be genetically engineered to produce glycoproteins that comprise a mammalian or human-like glycosylation pattern comprising complex and/or hybrid *N*-glycans and not glycoproteins that have the host cells' endogenous glycosylation pattern.

The present invention has been exemplified herein using *Pichia pastoris* host cells genetically engineered to produce mammalian- or human-like complex *N*-glycans; however, the present invention can be applied to other yeast ost cells (including but not limited to *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Ogataea minuta*, and *Pichia pastoris*) or filamentous fungi (including but not limited to *Trichoderma reesei*) that produce glycoproteins that have yeast or fungal *N*-glycans (either hypermannosylated *N*-glycans or high mannose *N*-glycans) or genetically engineered to produce glycoproteins that have mammalian- or human-like

high mannose, complex, or hybrid *N*-glycans to improve the overall *N*-glycosylation site occupancy of glycoproteins produced in the host cell. Furthermore, the present invention can also be applied to plant and mammalian expression system to improve the overall *N*-glycosylation site occupancy of glycoproteins produced in these plant or mammalian expression systems, particularly glycoproteins that have more than two *N*-linked glycosylation sites.

Therefore, in one aspect of the above, provided is a method for producing a heterologous glycoprotein in a host cell, comprising providing a host cell that includes a nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans in a host cell, comprising providing a host cell that is genetically engineered to produce glycoproteins that have human-like *N*-glycans and includes a nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein and wherein the endogenous host cell genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex are expressed; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

Expression of the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex includes expression of the endogenous host cell gene encoding the endogenous STT3 protein or homologue. In the case of yeast host cells, the endogenous host cell genes encoding the proteins comprising the OTase complex are expressed, which includes the expression of the endogenous *STT3* gene. Currently, the genes encoding proteins comprising the *Saccharomyces cerevisiae* OTase complex are known to include *OST1*, *OST2*, *OST3*, *OST4*, *OST5*, *OST6*, *WBP1*, *SWP1*, and *STT3* (See for example, Spirig *et al.*, Molec. Gen. Genet. 256: 628-637 (1997) and in *Pichia pastoris*, the OTase complex appears to include at least Ost1p, Ost2p, Ost3p, Ost4p, Ost6p, Wbp1, Swp1p, and Stt3p (See Shutter *et al.*, op. cit.).

In general, the heterologous single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of an OTase complex, for example, a yeast OTase complex. Thus, the heterologous single-subunit oligosaccharyltransferase is capable of functionally complementing or rescuing a lethal mutation of at least one essential protein of an OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *Saccharomyces cerevisiae* and/or *Pichia pastoris* *STT3*

locus, *WBPI* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. In general, heterologous single-subunit oligosaccharyltransferases that can be used in the methods herein for increasing *N*-glycosylation site occupancy is a heterologous single-subunit oligosaccharyltransferase that in particular embodiments is capable of functionally suppressing (or rescuing or complementing) the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* and/or *Pichia pastoris* OTase complex. For example, in further aspects, the heterologous single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein, which is capable of functionally suppressing (or rescuing or complementing) the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* or *Pichia pastoris* OTase complex. Therefore, for a particular host cell, a particular heterologous single-subunit oligosaccharyltransferase is suitable for expression in the particular host cell provided the single-subunit heterologous oligosaccharyltransferase is capable of suppressing the lethal phenotype of at least one essential protein of the yeast OTase complex. In further aspect, a heterologous single-subunit heterologous oligosaccharyltransferase is selected for expression in a particular host cell provided the single-subunit heterologous oligosaccharyltransferase is capable of suppressing the lethal phenotype of at least one essential protein of the *Saccharomyces cerevisiae* and/or *Pichia pastoris* OTase complex. The essential proteins include *OST1*, *OST2*, *WBPI*, *SWP1*, and *STT3*.

As used herein, a lethal mutation includes a deletion or disruption of the gene encoding the essential protein of the OTase complex or a mutation in the coding sequence that renders the essential protein non-functional. The term can further include knock-down mutations wherein production of a functional essential protein is abrogated using shRNA or RNAi.

Further provided is a host cell, comprising a first nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase; and a second nucleic acid molecule encoding a heterologous glycoprotein; and the host cell expresses its endogenous genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex, which includes expressing the endogenous host cell gene encoding the host cell STT3 protein, which in yeast is the *STT3* gene. In further aspects of a yeast host cell, the host cell expresses the endogenous genes encoding the proteins comprising the OTase complex.

In particular aspects of any of the above, the host cell further comprises one or more a nucleic acid molecules encoding additional heterologous oligosaccharyltransferases, which can include single-subunit or multimeric oligosaccharyltransferases. For example, the host cell can comprise one or more nucleic acid molecules encoding one or more single-subunit oligosaccharyltransferases selected from the group consisting of the *LmSTT3A* protein, *LmSTT3B* protein, and *LmSTT3D* protein. In further aspects, the host cell can further include a nucleic acid molecule encoding *LmSTT3C* protein. In further aspects of any one of the above, the host cell can include one or more nucleic acid molecules encoding one or more

oligosaccharyltransferases selected from the group consisting of the *Toxoplasma gondii* STT3 protein, *Trypanosoma cruzi* STT3 protein, *Trypanosoma brucei* STT3 protein, and *C. elegans* STT3 protein. In further still aspects of any one of the above, the host cell can further include a nucleic acid molecule encoding the *Pichia pastoris* STT3 protein.

Lower eukaryotes such as yeast or filamentous fungi are often used for expression of recombinant glycoproteins because they can be economically cultured, give high yields, and when appropriately modified are capable of suitable glycosylation. Yeast in particular offers established genetics allowing for rapid transfections, tested protein localization strategies and facile gene knock-out techniques. Suitable vectors have expression control sequences, such as promoters, including 3-phosphoglycerate kinase or other glycolytic enzymes, and an origin of replication, termination sequences, and the like as desired.

Useful lower eukaryote host cells 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* and *Neurospora crassa*. Various yeasts, such as *K. lactis*, *Pichia pastoris*, *Pichia methanolica*, and *Hansenula polymorpha* are particularly suitable for cell culture because they are able to grow to high cell densities and secrete large quantities of recombinant protein. 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. In the case of lower eukaryotes, cells are routinely grown from between about one and a half to three days.

Therefore, provided is a method for producing a heterologous glycoprotein in a lower eukaryote host cell, comprising providing a lower eukaryote host cell that includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

Further provided is a lower eukaryote host cell, comprising a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and a second nucleic acid molecule encoding a heterologous glycoprotein; and wherein the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are expressed.

Further provided is a yeast or filamentous fungus host cell, comprising a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and a second nucleic acid molecule encoding a heterologous glycoprotein; and wherein the endogenous

host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are expressed. This includes expression of the endogenous STT3 gene, which in yeast is the *STT3* gene.

In particular aspects, the above yeast or filamentous fungus host cell can be a host cell that produces glycoproteins that have a yeast-like or filamentous fungus-like glycosylation pattern. The yeast glycosylation pattern can include hypermannosylated *N*-glycans or the yeast can be genetically engineered to lack α 1,6-mannosyltransferase activity, that is, the yeast host is genetically engineered to lack och1p activity, in which case, the yeast produces glycoproteins that have high mannose *N*-glycans that are not further hypermannosylated.

In particular embodiments of the above methods and host cells, the heterologous single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of the OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *STT3* locus, *WBP1* locus, *OST1* locus, *SWP1* locus, or *OST2* locus, or homologue thereof. In further aspects, the for example single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein.

The methods and host cells herein provide a means for producing heterologous glycoproteins in a host cell wherein the *N*-glycosylation site occupancy of a composition of the heterologous glycoproteins is greater than the *N*-glycosylation site occupancy for the heterologous produced in the host cell not modified as described herein to express a heterologous single-subunit oligosaccharyltransferase and the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex. For a lower eukaryote host cell such as yeast, when the *N*-glycosylation site occupancy of a heterologous glycoprotein is lower than that obtained for the heterologous glycoprotein when produced in mammalian or human cells, the *N*-glycosylation site occupancy of the glycoprotein produced in the host cell can be made the same as or more similar to the *N*-glycosylation site occupancy of the glycoprotein in the mammalian or human cell by producing the glycoprotein in a host cell that express a heterologous single-subunit oligosaccharyltransferase and the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex. As shown in the examples, *Pichia pastoris* host cells that express a heterologous single-subunit oligosaccharyltransferase and the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are capable of producing antibodies wherein the *N*-glycosylation site occupancy of the antibodies is similar to that of the antibodies produced in Chinese hamster ovary (CHO) cells (See also **Figure 19**).

A method for measuring *N*-glycosylation site occupancy is to separate and measure the amount of glycosylated protein and non-glycosylated protein and determine the *N*-glycosylation site occupancy using the formula

(Moles glycosylated protein)/(moles glycosylated protein + moles non-glycosylated protein) x 100 = percent *N*-glycosylation site occupancy

When measuring the *N*-glycosylation site occupancy of antibodies in an antibody composition, the antibodies in the composition are reduced and the moles of glycosylated and non-glycosylated heavy chains determined. Each heavy chain has one *N*-glycosylation site at Asn-297. The percent *N*-glycosylation site occupancy is determined based upon total moles of *N*-glycans released and the total moles of antibody heavy chains. For example, an *N*-glycosylation site occupancy of 94% would indicate that 94% of the heavy chains in the composition have an *N*-glycan at Asn-297 and 6% of the heavy chains would lack an *N*-glycan. Antibodies consist of two heavy chains and two light chains. In the above example, antibodies in the composition can have both heavy chains linked to an *N*-glycan, one of the two heavy chains with an *N*-glycan, or neither chain with an *N*-glycan. Therefore, a 94% *N*-glycosylation site occupancy of heavy chains would suggest that about 88% of the antibodies in the composition would have both heavy chains *N*-glycosylated and 11.4% of the antibodies would have only one of the two heavy chains *N*-glycosylated. To get a qualitative indication that the above is correct, whole antibodies are analyzed by a method such as Q-TOF (hybrid quadrupole time of flight mass spectrometer with MS/MS capability).

A general method for measuring *N*-glycosylation site occupancy of antibodies can use the following method, which is exemplified in Example 3. The antibodies are reduced to heavy chains (HC) and light chains (LC) and the amount of glycosylated heavy chain (GHC) and non-glycosylated heavy chains (NGHC) are determined by a method such as capillary electrophoresis. The *N*-glycosylation site occupancy using the formula

Moles GHC)/(moles GHC + moles NGHC) x 100 = percent *N*-glycosylation site HC occupancy

For any *N*-glycosylation site, the site is either occupied or not. Therefore, *N*-glycan occupancy of 100% would be equivalent to a ratio of 1:1 (1 mole of *N*-glycan per 1 mole of *N*-glycosylation site, e.g., heavy chain from reduced antibody) or 2:1 (2 moles of *N*-glycan per 1 mole of protein with two *N*-glycosylation sites, e.g., non-reduced antibody). *N*-glycan occupancy of 80% would be equivalent to a ratio of 0.8:1 (0.8 mole of *N*-glycan per 1 mole of *N*-glycosylation site, e.g., heavy chain from reduced antibody) or 1.6:1 (1.6 moles of *N*-glycan per mole of protein with two *N*-glycosylation sites, e.g., non-reduced antibody).

An estimate of the proportion of whole antibodies in which both heavy chains are glycosylated can be approximated by the formula (fraction GHC)² x 100 = fully occupied antibodies (whole, non-reduced antibodies in which both *N*-glycosylation sites are occupied). Example 3 shows that the methods herein enable the production of antibody compositions

wherein about 70% to about 98% of the non-reduced whole antibody molecules in the composition have both *N*-glycosylation sites occupied. Since measurement of *N*-glycosylation site occupancy was determined using reduced antibody molecules, the results herein show that for compositions comprising glycoprotein molecules containing a single glycosylation site, more than 84% to at least 99% of the glycoprotein molecules were *N*-glycosylated. Therefore, the methods and host cells herein enable production of glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied.

Another method for measuring *N*-glycosylation site occupancy of glycoproteins in a glycoprotein composition can be accomplished by releasing the *N*-glycans from the glycoproteins in the composition and measuring the molar amount of the *N*-glycans released and the molar amount of glycoprotein times the number of glycosylation sites on the glycoprotein. The following formula can be used

$$\frac{(\text{Total moles of } N\text{-glycans})}{(\text{Total moles of glycoprotein} \times \text{No. of sites})} \times 100 = \text{percent } N\text{-glycosylation site occupancy.}$$

The above formula will give the percent of total *N*-glycosylation sites that are occupied.

Lower eukaryotes, particularly yeast, can be genetically modified so that they express glycoproteins in which the glycosylation pattern is mammalian or human-like or humanized. In this manner, glycoprotein compositions can be produced in which a specific desired glycoform is predominant in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or genetically engineering the host cells and/or supplying exogenous enzymes to mimic all or part of the mammalian glycosylation pathway as described in U.S. Published Application No. 2004/0018590, the disclosure of which is 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.

Lower eukaryotes such as yeast can be genetically modified so that they express glycoproteins in which the glycosylation pattern is mammalian-like or 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.*, U.S. Patent No. 7,449,308, the disclosure of which is incorporated herein by reference. Thus, in particular aspects of the invention, the host cell is yeast, for example, a methylotrophic yeast such as *Pichia pastoris* or *Ogataea minuta* and mutants thereof and genetically engineered variants thereof. In this manner, glycoprotein compositions can be produced in which a specific desired glycoform is predominant in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or genetically engineering the host cells and/or supplying exogenous enzymes to

mimic all or part of the mammalian glycosylation pathway as described in U.S. Patent No. 7,449,308, the disclosure of which is 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.

In one embodiment, the host cell further includes an α 1,2-mannosidase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target the α 1,2-mannosidase activity to the ER or Golgi apparatus of the host cell. Passage of a recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a Man₅GlcNAc₂ glycoform, for example, a recombinant glycoprotein composition comprising predominantly a Man₅GlcNAc₂ glycoform. For example, 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 Man₅GlcNAc₂ glycoform.

In a further embodiment, the immediately preceding host cell further 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 GlcNAcMan₅GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAcMan₅GlcNAc₂ 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 GlcNAcMan₅GlcNAc₂ glycoform. The glycoprotein produced in the above cells can be treated

in vitro with a hexaminidase to produce a recombinant glycoprotein comprising a Man₅GlcNAc₂ glycoform.

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 GlcNAcMan₃GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAcMan₃GlcNAc₂ 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 GlcNAc₂Man₃GlcNAc₂ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a Man₃GlcNAc₂ glycoform.

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 GlcNAc₂Man₃GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a GlcNAc₂Man₃GlcNAc₂ 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 GlcNAc₂Man₃GlcNAc₂ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexaminidase to produce a recombinant glycoprotein comprising a Man₃GlcNAc₂ glycoform.

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 GalGlcNAc₂Man₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform, or mixture thereof for example a recombinant glycoprotein composition comprising predominantly a GalGlcNAc₂Man₃GlcNAc₂ glycoform or Gal₂GlcNAc₂Man₃GlcNAc₂ 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 Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a galactosidase to produce a recombinant glycoprotein comprising a GlcNAc₂Man₃GlcNAc₂ glycoform, for example a recombinant glycoprotein composition comprising predominantly a
5 GlcNAc₂Man₃GlcNAc₂ glycoform.

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
10 Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly a NANA₂Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform or NANAGal₂GlcNAc₂Man₃GlcNAc₂ 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
15 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
20 a recombinant glycoprotein comprising predominantly a Gal₂GlcNAc₂Man₃GlcNAc₂ glycoform or GalGlcNAc₂Man₃GlcNAc₂ glycoform or mixture thereof.

Any one of the preceding host cells can further 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
25 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.

In further embodiments, the host cell that produces glycoproteins that have predominantly GlcNAcMan₅GlcNAc₂ *N*-glycans further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the
30 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 GalGlcNAcMan₅GlcNAc₂ glycoform.

In a further embodiment, the immediately preceding host cell that produced
35 glycoproteins that have predominantly the GalGlcNAcMan₅GlcNAc₂ *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 NANAGalGlcNAcMan5GlcNAc₂ glycoform.

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 α 1,2-fucosyltransferase, α 1,3-fucosyltransferase, α 1,4-fucosyltransferase, and α 1,6-fucosyltransferase.

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.

Host cells further include *Pichia pastoris* that are genetically engineered to eliminate glycoproteins having phosphomannose residues by deleting or disrupting one or both of the phosphomannosyltransferase genes *PNO1* and *MNN4B* (See for example, U.S. Patent Nos. 7,198,921 and 7,259,007; the disclosures of which are all incorporated herein by reference), which in further aspects can also include deleting or disrupting the *MNN4A* gene. Disruption includes disrupting the open reading frame encoding the particular enzymes or disrupting expression of the open reading frame or abrogating translation of RNAs encoding one or more of the β -mannosyltransferases and/or phosphomannosyltransferases using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures.

Host cells further include lower eukaryote cells (e.g., yeast such as *Pichia pastoris*) that are genetically modified to control *O*-glycosylation of the glycoprotein by deleting or disrupting one or more of the protein *O*-mannosyltransferase (Dol-P-Man:Protein (Ser/Thr) Mannosyl Transferase genes) (*PMTs*) (See U.S. Patent No. 5,714,377; the disclosure of which is incorporated herein by reference) or grown in the presence of Pmtp inhibitors and/or an α 1,2 mannosidase as disclosed in Published International Application No. WO 2007061631 the

disclosure of which is incorporated herein by reference), or both. Disruption includes disrupting the open reading frame encoding the *Pmtp* or disrupting expression of the open reading frame or abrogating translation of RNAs encoding one or more of the *Pmtps* using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures.

Pmtp inhibitors include but are not limited to a benzylidene thiazolidinediones. Examples of benzylidene thiazolidinediones that can be used are 5-[[3,4-bis(phenylmethoxy)phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid; 5-[[3-(1-Phenylethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid; and 5-[[3-(1-Phenyl-2-hydroxy)ethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid.

In particular embodiments, the function or expression of at least one endogenous *PMT* gene is reduced, disrupted, or deleted. For example, in particular embodiments the function or expression of at least one endogenous *PMT* gene selected from the group consisting of the *PMT1*, *PMT2*, *PMT3*, and *PMT4* genes is reduced, disrupted, or deleted; or the host cells are cultivated in the presence of one or more *PMT* inhibitors. In further embodiments, the host cells include one or more *PMT* gene deletions or disruptions and the host cells are cultivated in the presence of one or more *Pmtp* inhibitors. In particular aspects of these embodiments, the host cells also express a secreted α -1,2-mannosidase.

PMT deletions or disruptions and/or *Pmtp* inhibitors control *O*-glycosylation by reducing *O*-glycosylation occupancy; that is by reducing the total number of *O*-glycosylation sites on the glycoprotein that are glycosylated. The further addition of an α -1,2-mannosidase that is secreted by the cell controls *O*-glycosylation by reducing the mannose chain length of the *O*-glycans that are on the glycoprotein. Thus, combining *PMT* deletions or disruptions and/or *Pmtp* inhibitors with expression of a secreted α -1,2-mannosidase controls *O*-glycosylation by reducing occupancy and chain length. In particular circumstances, the particular combination of *PMT* deletions or disruptions, *Pmtp* inhibitors, and α -1,2-mannosidase is determined empirically as particular heterologous glycoproteins (antibodies, for example) may be expressed and transported through the Golgi apparatus with different degrees of efficiency and thus may require a particular combination of *PMT* deletions or disruptions, *Pmtp* inhibitors, and α -1,2-mannosidase. In another aspect, genes encoding one or more endogenous mannosyltransferase enzymes are deleted. The deletion(s) can be in combination with providing the secreted α -1,2-mannosidase and/or *PMT* inhibitors or can be in lieu of providing the secreted α -1,2-mannosidase and/or *PMT* inhibitors.

Thus, the control of *O*-glycosylation can be useful for producing particular glycoproteins in the host cells disclosed herein in better total yield or in yield of properly assembled glycoprotein. The reduction or elimination of *O*-glycosylation appears to have a

beneficial effect on the assembly and transport of glycoproteins such as whole antibodies as they traverse the secretory pathway and are transported to the cell surface. Thus, in cells in which *O*-glycosylation is controlled, the yield of properly assembled glycoproteins such as antibody fragments is increased over the yield obtained in host cells in which *O*-glycosylation is not controlled.

To reduce or eliminate the likelihood of *N*-glycans and *O*-glycans with β -linked mannose residues, which are resistant to α -mannosidases, the recombinant glycoengineered *Pichia pastoris* host cells are genetically engineered to eliminate glycoproteins having α -mannosidase-resistant *N*-glycans by deleting or disrupting one or more of the β -mannosyltransferase genes (e.g., *BMT1*, *BMT2*, *BMT3*, and *BMT4*) (See, U.S. Patent No. 7,465,577 and U.S. Patent No. 7,713,719). The deletion or disruption of *BMT2* and one or more of *BMT1*, *BMT3*, and *BMT4* also reduces or eliminates detectable cross reactivity to antibodies against host cell protein.

Yield of glycoprotein can in some situations be improved by overexpressing nucleic acid molecules encoding mammalian or human chaperone proteins or replacing the genes encoding one or more endogenous chaperone proteins with nucleic acid molecules encoding one or more mammalian or human chaperone proteins. In addition, the expression of mammalian or human chaperone proteins in the host cell also appears to control *O*-glycosylation in the cell. Thus, further included are the host cells herein wherein the function of at least one endogenous gene encoding a chaperone protein has been reduced or eliminated, and a vector encoding at least one mammalian or human homolog of the chaperone protein is expressed in the host cell. Also included are host cells in which the endogenous host cell chaperones and the mammalian or human chaperone proteins are expressed. In further aspects, the lower eukaryotic host cell is a yeast or filamentous fungi host cell. Examples of the use of chaperones of host cells in which human chaperone proteins are introduced to improve the yield and reduce or control *O*-glycosylation of recombinant proteins has been disclosed in Published International Application No. WO 2009105357 and WO2010019487 (the disclosures of which are incorporated herein by reference). Like above, further included are lower eukaryotic host cells wherein, in addition to replacing the genes encoding one or more of the endogenous chaperone proteins with nucleic acid molecules encoding one or more mammalian or human chaperone proteins or overexpressing one or more mammalian or human chaperone proteins as described above, the function or expression of at least one endogenous gene encoding a protein *O*-mannosyltransferase (PMT) protein is reduced, disrupted, or deleted. In particular embodiments, the function of at least one endogenous *PMT* gene selected from the group consisting of the *PMT1*, *PMT2*, *PMT3*, and *PMT4* genes is reduced, disrupted, or deleted.

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)Man₃GlcNAc₂, Gal(1-4)GlcNAc(2-4)Man₃GlcNAc₂, and NANA(1-4)Gal(1-4)GlcNAc(2-4)Man₃GlcNAc₂; hybrid *N*-glycans maybe selected from the group consisting of GlcNAcMan₃GlcNAc₂; GalGlcNAcMan₃GlcNAc₂; NANAGalGlcNAcMan₃GlcNAc₂ GlcNAcMan₅GlcNAc₂, GalGlcNAcMan₅GlcNAc₂, and NANAGalGlcNAcMan₅GlcNAc₂; and high Mannose *N*-glycans maybe selected from the group consisting of Man₅GlcNAc₂, Man₆GlcNAc₂, Man₇GlcNAc₂, Man₈GlcNAc₂, and Man₉GlcNAc₂. Further included are glycoproteins having *N*-glycans consisting of the *N*-glycan structure Man₃GlcNAc₂, for example, as shown in U.S. Published Application No. 20050170452.

Therefore, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans in a lower eukaryote host cell, comprising providing a lower eukaryote host cell that is genetically engineered to produce glycoproteins that have human-like *N*-glycans and includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans in a yeast or filamentous fungus host cell, comprising providing a yeast or filamentous fungus host cell that is genetically engineered to produce glycoproteins that have human-like *N*-glycans and includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

Further provided is a yeast or filamentous fungus host cell genetically engineered to produce glycoproteins having mammalian- or human-like *N*-glycans, comprising a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase; and a second nucleic acid molecule encoding a heterologous glycoprotein; and wherein the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are expressed. This includes expression of the endogenous STT3 gene, which in yeast is the *STT3* gene.

In general, in the above methods and host cells, the single-subunit oligosaccharyltransferase is capable of functionally suppressing the lethal phenotype of a mutation of at least one essential protein of the OTase complex. In further aspects, the essential protein of the OTase complex is encoded by the *STT3* locus, *WBP1* locus, *OST1* locus, *SWP1*

locus, or *OST2* locus, or homologue thereof. In further aspects, the for example single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein.

Promoters are DNA sequence elements for controlling gene expression. In particular, promoters specify transcription initiation sites and can include a TATA box and upstream promoter elements. The promoters selected are those which would be expected to be operable in the particular host system selected. For example, yeast promoters are used when a yeast such as *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Ogataea minuta*, or *Pichia pastoris* is the host cell whereas fungal promoters would be used in host cells such as *Aspergillus niger*, *Neurospora crassa*, or *Tricoderma reesei*. Examples of yeast promoters include but are not limited to the *GAPDH*, *AOX1*, *SEC4*, *HHI*, *PMA1*, *OCH1*, *GAL1*, *PGK*, *GAP*, *TPI*, *CYC1*, *ADH2*, *PHO5*, *CUP1*, *MFa1*, *FLD1*, *PMA1*, *PDI*, *TEF*, *RPL10*, and *GUT1* promoters. Romanos *et al.*, Yeast 8: 423-488 (1992) provide a review of yeast promoters and expression vectors. Hartner *et al.*, Nucl. Acid Res. 36: e76 (pub on-line 6 June 2008) describes a library of promoters for fine-tuned expression of heterologous proteins in *Pichia pastoris*.

The promoters that are operably linked to the nucleic acid molecules disclosed herein can be constitutive promoters or inducible promoters. An inducible promoter, for example the *AOX1* promoter, is a promoter that directs transcription at an increased or decreased rate upon binding of a transcription factor in response to an inducer. Transcription factors as used herein include any factor that can bind to a regulatory or control region of a promoter and thereby affect transcription. The RNA synthesis or the promoter binding ability of a transcription factor within the host cell can be controlled by exposing the host to an inducer or removing an inducer from the host cell medium. Accordingly, to regulate expression of an inducible promoter, an inducer is added or removed from the growth medium of the host cell. Such inducers can include sugars, phosphate, alcohol, metal ions, hormones, heat, cold and the like. For example, commonly used inducers in yeast are glucose, galactose, alcohol, and the like.

Transcription termination sequences that are selected are those that are operable in the particular host cell selected. For example, yeast transcription termination sequences are used in expression vectors when a yeast host cell such as *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, or *Pichia pastoris* is the host cell whereas fungal transcription termination sequences would be used in host cells such as *Aspergillus niger*, *Neurospora crassa*, or *Tricoderma reesei*. Transcription termination sequences include but are not limited to the *Saccharomyces cerevisiae* *CYC* transcription termination sequence (*ScCYC* TT), the *Pichia pastoris* *ALG3* transcription termination sequence (*ALG3* TT), the *Pichia pastoris* *ALG6* transcription termination sequence (*ALG6* TT), the *Pichia pastoris* *ALG12* transcription termination sequence (*ALG12* TT), the *Pichia pastoris* *AOX1* transcription termination sequence (*AOX1* TT), the *Pichia pastoris* *OCH1* transcription termination sequence (*OCH1* TT) and *Pichia pastoris* *PMA1* transcription

termination sequence (*PMA1* TT). Other transcription termination sequences can be found in the examples and in the art.

For genetically engineering yeast, selectable markers can be used to construct the recombinant host cells include drug resistance markers and genetic functions which allow the yeast host cell to synthesize essential cellular nutrients, e.g. amino acids. Drug resistance markers which are commonly used in yeast include chloramphenicol, kanamycin, methotrexate, G418 (geneticin), Zeocin, and the like. Genetic functions which allow the yeast host cell to synthesize essential cellular nutrients are used with available yeast strains having auxotrophic mutations in the corresponding genomic function. Common yeast selectable markers provide genetic functions for synthesizing leucine (*LEU2*), tryptophan (*TRP1* and *TRP2*), proline (*PRO1*), uracil (*URA3*, *URA5*, *URA6*), histidine (*HIS3*), lysine (*LYS2*), adenine (*ADE1* or *ADE2*), and the like. Other yeast selectable markers include the *ARR3* gene from *S. cerevisiae*, which confers arsenite resistance to yeast cells that are grown in the presence of arsenite (Bobrowicz *et al.*, Yeast, 13:819-828 (1997); Wysocki *et al.*, J. Biol. Chem. 272:30061-30066 (1997)). A number of suitable integration sites include those enumerated in U.S. Patent No. 7,479,389 (the disclosure of which is incorporated herein by reference) and include homologs to loci known for *Saccharomyces cerevisiae* and other yeast or fungi. Methods for integrating vectors into yeast are well known (See for example, U.S. Patent No. 7,479,389, U.S. Patent No. 7,514,253, U.S. Published Application No. 2009012400, and WO2009/085135; the disclosures of which are all incorporated herein by reference). Examples of insertion sites include, but are not limited to, *Pichia ADE* genes; *Pichia TRP* (including *TRP1* through *TRP2*) genes; *Pichia MCA* genes; *Pichia CYM* genes; *Pichia PEP* genes; *Pichia PRB* genes; and *Pichia LEU* genes. The *Pichia ADE1* and *ARG4* genes have been described in Lin Cereghino *et al.*, Gene 263:159-169 (2001) and U.S. Patent No. 4,818,700 (the disclosure of which is incorporated herein by reference), the *HIS3* and *TRP1* genes have been described in Cosano *et al.*, Yeast 14:861-867 (1998), *HIS4* has been described in GenBank Accession No. X56180.

The methods disclosed herein can be adapted for use in mammalian, plant, and insect cells. Examples of animal cells include, but are not limited to, SC-I cells, LLC-MK cells, CV-I cells, CHO cells, COS cells, murine cells, human cells, HeLa cells, 293 cells, VERO cells, MDBK cells, MDCK cells, MDOK cells, CRFK cells, RAF cells, TCMK cells, LLC-PK cells, PK15 cells, WI-38 cells, MRC-5 cells, T-FLY cells, BHK cells, SP2/0, NSO cells, and derivatives thereof. Insect cells include cells of *Drosophila melanogaster* origin. These cells can be genetically engineered to render the cells capable of making immunoglobulins that have particular or predominantly particular *N*-glycans. For example, U.S. Patent No. 6,949,372 discloses methods for making glycoproteins in insect cells that are sialylated. Yamane-Ohnuki *et al.* Biotechnol. Bioeng. 87: 614-622 (2004), Kanda *et al.*, Biotechnol. Bioeng. 94: 680-688 (2006), Kanda *et al.*, Glycobiol. 17: 104-118 (2006), and U.S. Pub. Application Nos.

2005/0216958 and 2007/0020260 (the disclosures of which are incorporated herein by reference) disclose mammalian cells that are capable of producing immunoglobulins in which the *N*-glycans thereon lack fucose or have reduced fucose. U.S. Published Patent Application No.

2005/0074843 (the disclosure of which is incorporated herein by reference) discloses making
5 antibodies in mammalian cells that have bisected *N*-glycans.

The regulatable promoters selected for regulating expression of the expression cassettes in mammalian, insect, or plant cells should be selected for functionality in the cell-type chosen. Examples of suitable regulatable promoters include but are not limited to the tetracycline-regulatable promoters (*See* for example, Berens & Hillen, *Eur. J. Biochem.* 270: 3109-3121 (2003)), RU 486-inducible promoters, ecdysone-inducible promoters, and kanamycin-regulatable systems. These promoters can replace the promoters exemplified in the expression cassettes described in the examples. The capture moiety can be fused to a cell surface anchoring protein suitable for use in the cell-type chosen. Cell surface anchoring proteins including GPI proteins are well known for mammalian, insect, and plant cells. GPI-anchored fusion proteins
10 has been described by Kennard *et al.*, *Methods Biotechnol.* Vo. 8: *Animal Cell Biotechnology* (Ed. Jenkins. Human Press, Inc., Totowa, NJ) pp. 187-200 (1999). The genome targeting sequences for integrating the expression cassettes into the host cell genome for making stable recombinants can replace the genome targeting and integration sequences exemplified in the examples. Transfection methods for making stable and transiently transfected mammalian,
15 insect, and plant host cells are well known in the art. Once the transfected host cells have been constructed as disclosed herein, the cells can be screened for expression of the immunoglobulin of interest and selected as disclosed herein.

Therefore, in a further aspect of the above, provided is a method for producing a heterologous glycoprotein in a mammalian or insect host cell, comprising providing a
25 mammalian or insect host cell that includes includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (e.g., *Leishmania major* STT3 protein) and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein. In further aspects, the host cell is genetically engineered to produce glycoproteins with human-like *N*-glycans or *N*-glycans not normally endogenous to the host cell.
30

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein wherein the *N*-glycosylation site occupancy of the heterologous glycoprotein is greater than 83% in a mammalian or insect host cell, comprising providing a mammalian or insect host cell that includes a nucleic acid molecule encoding a heterologous
35 single-subunit oligosaccharyltransferase (e.g., *Leishmania major* STT3 protein) and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein

wherein the *N*-glycosylation site occupancy of the heterologous glycoprotein is greater than 83%. In further aspects, the host cell is genetically engineered to produce glycoproteins with human-like *N*-glycans or *N*-glycans not normally endogenous to the host cell.

In a further embodiment of the above methods, the endogenous host cell genes encoding the proteins comprising the oligosaccharyltransferase (OTase) complex are expressed.

In particular embodiments of the above methods, the *N*-glycosylation site occupancy is at least 94%. In further still embodiments, the *N*-glycosylation site occupancy is at least 99%.

Further provided is a mammalian or insect host cell, comprising a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (e.g., the *Leishmania major* STT3D protein); and a second nucleic acid molecule encoding a heterologous glycoprotein; and wherein the endogenous host cell genes encoding the proteins comprising the endogenous host cell oligosaccharyltransferase (OTase) complex are expressed.

In particular embodiments, the higher eukaryote cell, tissue, or organism can also be from the plant kingdom, for example, wheat, rice, corn, tobacco, and the like. Alternatively, bryophyte cells can be selected, for example from species of the genera *Physcomitrella*, *Funaria*, *Sphagnum*, *Ceratodon*, *Marchantia*, and *Sphaerocarpos*. Exemplary of plant cells is the bryophyte cell of *Physcomitrella patens*, which has been disclosed in WO 2004/057002 and WO2008/006554 (the disclosures of which are all incorporated herein by reference). Expression systems using plant cells can further manipulated to have altered glycosylation pathways to enable the cells to produce immunoglobulins that have predominantly particular *N*-glycans. For example, the cells can be genetically engineered to have a dysfunctional or no core fucosyltransferase and/or a dysfunctional or no xylosyltransferase, and/or a dysfunctional or no β 1,4-galactosyltransferase. Alternatively, the galactose, fucose and/or xylose can be removed from the immunoglobulin by treatment with enzymes removing the residues. Any enzyme resulting in the release of galactose, fucose and/or xylose residues from *N*-glycans which are known in the art can be used, for example α -galactosidase, β -xylosidase, and α -fucosidase. Alternatively, an expression system can be used which synthesizes modified *N*-glycans which can not be used as substrates by 1,3-fucosyltransferase and/or 1,2-xylosyltransferase, and/or 1,4-galactosyltransferase. Methods for modifying glycosylation pathways in plant cells are disclosed in U.S. Patent Nos. 7,449,308, 6,998,267 and 7,388,081 (the disclosures of which are incorporated herein by reference) which disclose methods for genetically engineering plants to make recombinant glycoproteins that have human-like *N*-glycans. WO 2008006554 (the disclosure of which is incorporated herein by reference) discloses methods for making glycoproteins such as antibodies in plants genetically engineered to make glycoproteins without xylose or fucose. WO 2007006570 (the disclosure of which is incorporated herein by reference)

discloses methods for genetically engineering bryophytes, ciliates, algae, and yeast to make glycoproteins that have animal or human-like glycosylation patterns.

Therefore, in a further aspect of the above, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans in a plant host cell, comprising providing a plant host cell that is genetically engineered to produce glycoproteins that have mammalian- or human-like *N*-glycans and includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (e.g., the *Leishmania major* STT3D protein) and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein.

In a further aspect of the above, provided is a method for producing a heterologous glycoprotein with mammalian- or human-like complex or hybrid *N*-glycans wherein the *N*-glycosylation site occupancy of the heterologous glycoprotein is greater than 83% in a plant host cell, comprising providing a plant host cell that is genetically engineered to produce glycoproteins that have mammalian- or human-like *N*-glycans and includes a nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (e.g., the *Leishmania major* STT3D protein) and a nucleic acid molecule encoding the heterologous glycoprotein; and culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein with mammalian- or human-like *N*-glycans wherein the *N*-glycosylation site occupancy of the heterologous glycoprotein is greater than 83%.

In a further embodiment of the above methods, the endogenous host cell genes encoding the proteins comprising the endogenous host cell oligosaccharyltransferase (OTase) complex are expressed.

In particular embodiments of the above methods, the *N*-glycosylation site occupancy is at least 94%. In further still embodiments, the *N*-glycosylation site occupancy is at least 99%.

Further provided is a plant host cell, comprising a first nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (e.g., the *Leishmania major* STT3D protein); and a second nucleic acid molecule encoding a heterologous glycoprotein; and wherein the endogenous host cell genes encoding the proteins comprising the endogenous host cell oligosaccharyltransferase (OTase) complex are expressed.

The host cells and methods herein are useful for producing a wide range of recombinant proteins and glycoproteins. Examples of recombinant proteins and glycoproteins that can be produced in the host cells disclosed herein include but are not limited to erythropoietin (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and granulocyte-colony stimulating factor (GCSF); granulocyte macrophage-colony stimulating factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C;

antithrombin III; thrombin; soluble IgE receptor α -chain; immunoglobulins or antibodies such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; and IL-2 receptor agonist.

The recombinant host cells and methods disclosed herein are particularly useful for producing antibodies, Fc fusion proteins, and the like where it is desirable to provide antibody or Fc fusion protein compositions wherein the percent galactose-containing *N*-glycans is increased compared to the percent galactose obtainable in the host cells prior to modification as taught herein. Examples of antibodies that can be made in the host cells herein include but are not limited to human antibodies, humanized antibodies, chimeric antibodies, heavy chain antibodies (e.g., camel or llama). Specific antibodies include but are not limited to the following antibodies recited under their generic name (target): Muromonab-CD3 (anti-CD3 receptor antibody), Abciximab (anti-CD41 7E3 antibody), Rituximab (anti-CD20 antibody), Daclizumab (anti-CD25 antibody), Basiliximab (anti-CD25 antibody), Palivizumab (anti-RSV (respiratory syncytial virus) antibody), Infliximab (anti-TNF α antibody), Trastuzumab (anti-Her2 antibody), Gemtuzumab ozogamicin (anti-CD33 antibody), Alemtuzumab (anti-CD52 antibody), Ibritumomab tiuxeten (anti-CD20 antibody), Adalimumab (anti-TNF α antibody), Omalizumab (anti-IgE antibody), Tositumomab-¹³¹I (iodinated derivative of an anti-CD20 antibody), Efalizumab (anti-CD11a antibody), Cetuximab (anti-EGF receptor antibody), Golimumab (anti-TNF α antibody), Bevacizumab (anti VEGF-A antibody), and variants thereof. Examples of Fc-fusion proteins that can be made in the host cells disclosed herein include but are not limited to etanercept (TNFR-Fc fusion protein), FGF-21-Fc fusion proteins, GLP-1-Fc fusion proteins, RAGE-Fc fusion proteins, EPO-Fc fusion proteins, ActRIIA-Fc fusion proteins, ActRIIB-Fc fusion proteins, glucagon-Fc fusions, oxyntomodulin-Fc-fusions, and analogs and variants thereof.

Thus, the methods and host cells herein can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied and the glycoproteins have mammalian- or human-like *N*-glycans.

Further, the methods and host cells herein can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-

glycosylation sites of the glycoproteins in the composition are occupied and the glycoproteins have mammalian- or human-like *N*-glycans that lack fucose.

Further, the methods and yeast or filamentous fungus host cells genetically engineered to produce mammalian-like or human-like *N*-glycans can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied and the glycoproteins have mammalian- or human-like *N*-glycans that lack fucose.

In some aspects, the yeast or filamentous host cells genetically engineered to produce fucosylated mammalian- or human-like *N*-glycans can be used to produce glycoprotein compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the *N*-glycosylation sites of the glycoproteins in the composition are occupied and the glycoproteins have mammalian- or human-like *N*-glycans that have fucose.

The recombinant cells disclosed herein can be used to produce antibodies and Fc fragments suitable for chemically conjugating to a heterologous peptide or drug molecule. For example, WO2005047334, WO2005047336, WO2005047337, and WO2006107124 (the disclosures of which are incorporated herein by reference) disclose chemically conjugating peptides or drug molecules to Fc fragments. EP1180121, EP1105409, and US 6,593,295 (the disclosures of which are incorporated herein by reference) disclose chemically conjugating peptides and the like to blood components, which includes whole antibodies.

Thus, the methods and host cells herein can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans.

Further, the methods and host cells herein can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans that lack fucose.

Further, the methods and yeast or filamentous fungus host cells genetically engineered to produce mammalian-like or human-like *N*-glycans can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans that lack fucose.

In some aspects, the yeast or filamentous host cells genetically engineered to produce fucosylated mammalian- or human-like *N*-glycans can be used to produce antibody compositions in which at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and the antibodies have mammalian- or human-like *N*-glycans that have fucose.

As shown in Example 3, the *N*-glycosylation composition of antibodies produced in *Pichia pastoris* strains, which have been genetically engineered to make galactose-terminated *N*-glycans, appear to range from about 50-60 mole%G0, 18-24 mole%G1, 3-8% mole% G2, 12-17 mole% Man5, and 3-6 mole% hybrids.

Therefore, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 70% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man5 structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 70% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Therefore, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 75% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man5 structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 75% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Further still, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 80% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man5 structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 80% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1

structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Therefore, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 85% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man₅ structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 85% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Further still, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 90% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man₅ structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 90% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Therefore, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 95% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man₅ structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 95% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure,

and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Further still, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 98% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man₅ structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 98% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

Therefore, provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50-70 mole% of the *N*-glycans have a G0 structure, 15-25 mole% of the *N*-glycans have a G1 structure, 4-12 mole% of the *N*-glycans have a G2 structure, 5-17 mole% of the *N*-glycans have a Man₅ structure, and 5-15 mole% of the *N*-glycans have a hybrid structure, and a pharmaceutically acceptable carrier. Further still is provided is a glycoprotein composition comprising a plurality of antibodies wherein at least 99% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 53 to 58 mole% of the *N*-glycans have a G0 structure, 20-22 mole% of the *N*-glycans have a G1 structure, and about 16 to 18 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure, and a pharmaceutically acceptable carrier. In further aspects of the above, the *N*-glycans further include fucose.

In particular embodiments, the antibodies comprise an antibody selected from the group consisting of anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, and anti-CD20 antibody.

All patents and publications referenced or mentioned herein are indicative of the levels of skill of those skilled in the art to which the invention pertains, and each such referenced patent or publication is hereby incorporated by reference to the same extent as if it had been incorporated by reference in its entirety individually or set forth herein in its entirety.

The following examples are intended to promote a further understanding of the present invention.

EXAMPLE 1

Plasmids comprising expression cassettes encoding the *Leishmania major* STT3D (LmSTT3D) open reading frame (ORF) operably linked to an inducible or constitutive promoter were constructed as follows.

The open reading frame encoding the LmSTT3D (SEQ ID NO:12) was codon-optimized for optimal expression in *P. pastoris* and synthesized by GeneArt AG, Brandenburg, Germany. The codon-optimized nucleic acid molecule encoding the LmSTT3D was designated pGLY6287 and has the nucleotide sequence shown in SEQ ID NO:11.

Plasmid pGLY6301 (**Figure 2**) is a roll-in integration plasmid that targets the URA6 locus in *P. pastoris*. The expression cassette encoding the LmSTT3D comprises a nucleic acid molecule encoding the LmSTT3D ORF codon-optimized for effective expression in *P. pastoris* operably linked at the 5' end to a nucleic acid molecule that has the inducible *P. pastoris* AOX1 promoter sequence (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae* CYC transcription termination sequence (SEQ ID NO:24). For selecting transformants, the plasmid comprises an expression cassette encoding the *S. cerevisiae* ARR3 ORF in which the nucleic acid molecule encoding the ORF (SEQ ID NO:32) is operably linked at the 5' end to a nucleic acid molecule having the *P. pastoris* RPL10 promoter sequence (SEQ ID NO:25) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae* CYC transcription termination sequence (SEQ ID NO:24). The plasmid further includes a nucleic acid molecule for targeting the URA6 locus (SEQ ID NO:33). Plasmid pGLY6301 was constructed by cloning the DNA fragment encoding the codon-optimized LmSTT3D ORF (pGLY6287) flanked by an *EcoRI* site at the 5' end and an *FseI* site at the 3' end into plasmid pGFI30t, which had been digested with *EcoRI* and *FseI*.

Plasmid pGLY6294 (**Figure 3**) is a KINKO integration vector that targets the TRP1 locus in *P. pastoris* without disrupting expression of the locus. KINKO (Knock-In with little or No Knock-Out) integration vectors enable insertion of heterologous DNA into a targeted locus without disrupting expression of the gene at the targeted locus and have been described in U.S. Published Application No. 20090124000. The expression cassette encoding the LmSTT3D comprises a nucleic acid molecule encoding the LmSTT3D ORF operably linked at the 5' end to a nucleic acid molecule that has the constitutive *P. pastoris* GAPDH promoter sequence (SEQ ID NO:26) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae* CYC transcription termination sequence (SEQ ID NO:24). For selecting transformants, the plasmid comprises an expression cassette encoding the Nourseothricin resistance (NAT^R) ORF (originally from pAG25 from EROSCARF, Scientific Research and Development GmbH, Daimlerstrasse 13a, D-61352 Bad Homburg, Germany, See Goldstein *et al.*, Yeast 15: 1541 (1999)); wherein the nucleic acid molecule encoding the ORF (SEQ ID NO:34) is operably linked to at the 5' end to a

nucleic acid molecule having the *Ashbya gossypii* *TEF1* promoter sequence (SEQ ID NO:86) and at the 3' end to a nucleic acid molecule that has the *Ashbya gossypii* *TEF1* termination sequence (SEQ ID NO:87). The two expression cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the ORF encoding Trp1p ending at the stop codon (SEQ ID NO:30) linked to a nucleic acid molecule having the *P. pastoris* *ALG3* termination sequence (SEQ ID NO:29) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP1* gene (SEQ ID NO:31). Plasmid pGLY6294 was constructed by cloning the DNA fragment encoding the codon-optimized *LmSTT3D* ORF (pGLY6287) flanked by a *NotI* site at the 5' end and a *PacI* site at the 3' end into plasmid pGLY597, which had been digested with *NotI* and *FseI*. An expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance ORF (NAT) operably linked to the *Ashbya gossypii* *TEF1* promoter (PTEF) and *Ashbya gossypii* *TEF1* termination sequence (TTEF).

The above plasmids can be used to introduce the *LmSTT3D* expression cassettes into *P. pastoris* to increase the *N*-glycosylation site occupancy on glycoproteins produced therein as shown in the following examples.

EXAMPLE 2

Genetically engineered *Pichia pastoris* strain **YGLY13992** is a strain that produces recombinant human anti-Her2 antibodies and *Pichia pastoris* strain **YGLY14401** is a strain that produces recombinant human anti-RSV antibodies. Construction of the strains is illustrated schematically in **Figure 1A-1H**. Briefly, the strains were constructed as follows.

The strain **YGLY8316** was constructed from wild-type *Pichia pastoris* strain **NRRL-Y 11430** using methods described earlier (See for example, U.S. Patent No. 7,449,308; U.S. Patent No. 7,479,389; U.S. Published Application No. 20090124000; Published PCT Application No. WO2009085135; Nett and Gerngross, Yeast 20:1279 (2003); Choi *et al.*, Proc. Natl. Acad. Sci. USA 100:5022 (2003); Hamilton *et al.*, Science 301:1244 (2003)). All plasmids were made in a pUC19 plasmid using standard molecular biology procedures. For nucleotide sequences that were optimized for expression in *P. pastoris*, the native nucleotide sequences were analyzed by the GENEOPTIMIZER software (GeneArt, Regensburg, Germany) and the results used to generate nucleotide sequences in which the codons were optimized for *P. pastoris* expression. Yeast strains were transformed by electroporation (using standard techniques as recommended by the manufacturer of the electroporator BioRad).

Plasmid pGLY6 (**Figure 4**) is an integration vector that targets the *URA5* locus. It contains a nucleic acid molecule comprising the *S. cerevisiae* invertase gene or transcription unit (*ScSUC2*; SEQ ID NO:38) flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris* *URA5* gene (SEQ ID NO:39) and on the

other side by a nucleic acid molecule comprising the nucleotide sequence from the 3' region of the *P. pastoris* *URA5* gene (SEQ ID NO:40). Plasmid pGLY6 was linearized and the linearized plasmid transformed into wild-type strain **NRRL-Y 11430** to produce a number of strains in which the *ScSUC2* gene was inserted into the *URA5* locus by double-crossover homologous recombination. Strain **YGLY1-3** was selected from the strains produced and is auxotrophic for uracil.

Plasmid pGLY40 (**Figure 5**) is an integration vector that targets the *OCH1* locus and contains a nucleic acid molecule comprising the *P. pastoris* *URA5* gene or transcription unit (SEQ ID NO:41) flanked by nucleic acid molecules comprising *lacZ* repeats (SEQ ID NO:42) which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *OCH1* gene (SEQ ID NO:43) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *OCH1* gene (SEQ ID NO:44). Plasmid pGLY40 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY1-3** to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *OCH1* locus by double-crossover homologous recombination. Strain **YGLY2-3** was selected from the strains produced and is prototrophic for *URA5*. Strain **YGLY2-3** was counterselected in the presence of 5-fluoroorotic acid (5-FOA) to produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain in the *OCH1* locus. This renders the strain auxotrophic for uracil. Strain **YGLY4-3** was selected.

Plasmid pGLY43a (**Figure 6**) is an integration vector that targets the *BMT2* locus and contains a nucleic acid molecule comprising the *K. lactis* UDP-*N*-acetylglucosamine (UDP-GlcNAc) transporter gene or transcription unit (*KIMNN2-2*, SEQ ID NO:45) adjacent to a nucleic acid molecule comprising the *P. pastoris* *URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The adjacent genes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *BMT2* gene (SEQ ID NO:46) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *BMT2* gene (SEQ ID NO:47). Plasmid pGLY43a was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY4-3** to produce a number of strains in which the *KIMNN2-2* gene and *URA5* gene flanked by the *lacZ* repeats has been inserted into the *BMT2* locus by double-crossover homologous recombination. The *BMT2* gene has been disclosed in Mille *et al.*, J. Biol. Chem. 283: 9724-9736 (2008) and U.S. Patent No.7,465,557. Strain **YGLY6-3** was selected from the strains produced and is prototrophic for uracil. Strain **YGLY6-3** was counterselected in the presence of 5-FOA to produce strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. This renders the strain auxotrophic for uracil. Strain **YGLY8-3** was selected.

Plasmid pGLY48 (**Figure 7**) is an integration vector that targets the *MNN4LI* locus and contains an expression cassette comprising a nucleic acid molecule encoding the

mouse homologue of the UDP-GlcNAc transporter (SEQ ID NO:48) open reading frame (ORF) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris GAPDH* promoter (SEQ ID NO:26) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae CYC* termination sequences (SEQ ID NO:24) adjacent to a nucleic acid molecule comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats and in which the expression cassettes together are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *P. pastoris MNN4L1* gene (SEQ ID NO:49) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4L1* gene (SEQ ID NO:50). Plasmid pGLY48 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY8-3** to produce a number of strains in which the expression cassette encoding the mouse UDP-GlcNAc transporter and the *URA5* gene have been inserted into the *MNN4L1* locus by double-crossover homologous recombination. The *MNN4L1* gene (also referred to as *MNN4B*) has been disclosed in U.S. Patent No. 7,259,007. Strain **YGLY10-3** was selected from the strains produced and then counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. Strain **YGLY12-3** was selected.

Plasmid pGLY45 (**Figure 8**) is an integration vector that targets the *PNO1/MNN4* loci and contains a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats which in turn is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *PNO1* gene (SEQ ID NO:51) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *MNN4* gene (SEQ ID NO:52). Plasmid pGLY45 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY12-3** to produce a number of strains in which the *URA5* gene flanked by the *lacZ* repeats has been inserted into the *PNO1/MNN4* loci by double-crossover homologous recombination. The *PNO1* gene has been disclosed in U.S. Patent No. 7,198,921 and the *MNN4* gene (also referred to as *MNN4B*) has been disclosed in U.S. Patent No. 7,259,007. Strain **YGLY14-3** was selected from the strains produced and then counterselected in the presence of 5-FOA to produce a number of strains in which the *URA5* gene has been lost and only the *lacZ* repeats remain. Strain **YGLY16-3** was selected.

Plasmid pGLY1430 (**Figure 9**) is a KINKO integration vector that targets the *ADE1* locus without disrupting expression of the locus and contains in tandem four expression cassettes encoding (1) the human GlcNAc transferase I catalytic domain (NA) fused at the *N*-terminus to *P. pastoris SEC12* leader peptide (10) to target the chimeric enzyme to the ER or Golgi, (2) mouse homologue of the UDP-GlcNAc transporter (MmTr), (3) the mouse mannosidase IA catalytic domain (FB) fused at the *N*-terminus to *S. cerevisiae SEC12* leader peptide (8) to target the chimeric enzyme to the ER or Golgi, and (4) the *P. pastoris URA5* gene

or transcription unit. KINKO (Knock-In with little or No Knock-Out) integration vectors enable insertion of heterologous DNA into a targeted locus without disrupting expression of the gene at the targeted locus and have been described in U.S. Published Application No. 20090124000.

The expression cassette encoding the NA10 comprises a nucleic acid molecule encoding the

human GlcNAc transferase I catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:53) fused at the 5' end to a nucleic acid molecule encoding the *SEC12* leader 10 (SEQ ID NO:54), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence. The expression cassette encoding MmTr

comprises a nucleic acid molecule encoding the mouse homologue of the UDP-GlcNAc transporter ORF operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *SEC4* promoter (SEQ ID NO:55) and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *OCH1* termination sequences (SEQ ID NO:56). The expression cassette encoding the FB8 comprises a nucleic acid molecule encoding the mouse mannosidase IA catalytic domain

(SEQ ID NO:57) fused at the 5' end to a nucleic acid molecule encoding the *SEC12-m* leader 8 (SEQ ID NO:58), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GADPH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris* *URA5* gene or transcription unit flanked by

nucleic acid molecules comprising *lacZ* repeats. The four tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and complete ORF of the *ADE1* gene (SEQ ID NO:59) followed by a *P. pastoris* *ALG3* termination sequence (SEQ ID NO:29) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ADE1* gene (SEQ ID NO:60). Plasmid pGLY1430

was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY16-3** to produce a number of strains in which the four tandem expression cassette have been inserted into the *ADE1* locus immediately following the *ADE1* ORF by double-crossover homologous recombination. The strain **YGLY2798** was selected from the strains produced and is auxotrophic for arginine and now prototrophic for uridine, histidine, and adenine. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY3794** was selected and is capable of making glycoproteins that have predominantly galactose terminated *N*-glycans.

Plasmid pGLY582 (**Figure 10**) is an integration vector that targets the *HIS1* locus and contains in tandem four expression cassettes encoding (1) the *S. cerevisiae* UDP-glucose epimerase (*ScGAL10*), (2) the human galactosyltransferase I (*hGalT*) catalytic domain fused at the N-terminus to the *S. cerevisiae* *KRE2-s* leader peptide (33) to target the chimeric enzyme to the ER or Golgi, (3) the *P. pastoris* *URA5* gene or transcription unit flanked by *lacZ* repeats, and

(4) the *D. melanogaster* UDP-galactose transporter (*DmUGT*). The expression cassette encoding the *ScGAL10* comprises a nucleic acid molecule encoding the *ScGAL10* ORF (SEQ ID NO:61) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter (SEQ ID NO:88) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence (SEQ ID NO:62). The expression cassette encoding the chimeric galactosyltransferase I comprises a nucleic acid molecule encoding the *hGalT* catalytic domain codon optimized for expression in *P. pastoris* (SEQ ID NO:63) fused at the 5' end to a nucleic acid molecule encoding the *KRE2-s* leader 33 (SEQ ID NO:64), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris* *URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The expression cassette encoding the *DmUGT* comprises a nucleic acid molecule encoding the *DmUGT* ORF (SEQ ID NO:65) operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *OCH1* promoter (SEQ ID NO:66) and operably linked at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *ALG12* transcription termination sequence (SEQ ID NO:67). The four tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *HIS1* gene (SEQ ID NO:68) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *HIS1* gene (SEQ ID NO:69). Plasmid pGLY582 was linearized and the linearized plasmid transformed into strain YGLY3794 to produce a number of strains in which the four tandem expression cassette have been inserted into the *HIS1* locus by homologous recombination. Strain YGLY3853 was selected and is auxotrophic for histidine and prototrophic for uridine.

Plasmid pGLY167b (**Figure 11**) is an integration vector that targets the *ARG1* locus and contains in tandem three expression cassettes encoding (1) the *D. melanogaster* mannosidase II catalytic domain (KD) fused at the *N*-terminus to *S. cerevisiae* *MNN2* leader peptide (53) to target the chimeric enzyme to the ER or Golgi, (2) the *P. pastoris* *HIS1* gene or transcription unit, and (3) the rat *N*-acetylglucosamine (GlcNAc) transferase II catalytic domain (TC) fused at the *N*-terminus to *S. cerevisiae* *MNN2* leader peptide (54) to target the chimeric enzyme to the ER or Golgi. The expression cassette encoding the KD53 comprises a nucleic acid molecule encoding the *D. melanogaster* mannosidase II catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:70) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 53 (SEQ ID NO:71), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The *HIS1* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris* *HIS1* gene or

transcription unit (SEQ ID NO:72). The expression cassette encoding the TC54 comprises a nucleic acid molecule encoding the rat GlcNAc transferase II catalytic domain codon-optimized for expression in *P. pastoris* (SEQ ID NO:73) fused at the 5' end to a nucleic acid molecule encoding the *MNN2* leader 54 (SEQ ID NO:74), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris PMA1* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris PMA1* transcription termination sequence. The three tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *ARG1* gene (SEQ ID NO:75) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *ARG1* gene (SEQ ID NO:76). Plasmid pGLY167b was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY3853** to produce a number of strains (in which the three tandem expression cassette have been inserted into the *ARG1* locus by double-crossover homologous recombination. The strain **YGLY4754** was selected from the strains produced and is auxotrophic for arginine and prototrophic for uridine and histidine. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY4799** was selected.

Plasmid pGLY3411 (**Figure 12**) is an integration vector that contains the expression cassette comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats flanked on one side with the 5' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:77) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT4* gene (SEQ ID NO:78). Plasmid pGLY3411 was linearized and the linearized plasmid transformed into **YGLY4799** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT4* locus by double-crossover homologous recombination. Strain **YGLY6903** was selected from the strains produced and is prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strains **YGLY7432** and **YGLY7433** were selected.

Plasmid pGLY3419 (**Figure 13**) is an integration vector that contains an expression cassette comprising the *P. pastoris URA5* gene flanked by *lacZ* repeats flanked on one side with the 5' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:79) and on the other side with the 3' nucleotide sequence of the *P. pastoris BMT1* gene (SEQ ID NO:80). Plasmid pGLY3419 was linearized and the linearized plasmid transformed into strain **YGLY7432** and **YGLY7433** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strains **YGLY7656** and **YGLY7651** were selected from the strains produced and are prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan. The strains were then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strains **YGLY7930** and **YGLY7940** were selected.

Plasmid pGLY3421 (**Figure 14**) is an integration vector that contains an expression cassette comprising the *P. pastoris* *URA5* gene flanked by *lacZ* repeats flanked on one side with the 5' nucleotide sequence of the *P. pastoris* *BMT3* gene (SEQ ID NO:81) and on the other side with the 3' nucleotide sequence of the *P. pastoris* *BMT3* gene (SEQ ID NO:82).

Plasmid pGLY3419 was linearized and the linearized plasmid transformed into strain **YGLY7930** and **YGLY7940** to produce a number of strains in which the *URA5* expression cassette has been inserted into the *BMT1* locus by double-crossover homologous recombination. The strains **YGLY7965** and **YGLY7961** were selected from the strains produced and are prototrophic for uracil, adenine, histidine, proline, arginine, and tryptophan.

Plasmid pGLY3673 (**Figure 15**) is a KINKO integration vector that targets the *PRO1* locus without disrupting expression of the locus and contains expression cassettes encoding the *T. reesei* α -1,2-mannosidase catalytic domain fused at the N-terminus to *S. cerevisiae* α MATpre signal peptide (α MATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell. The expression cassette encoding the α MATTrMan comprises a nucleic acid molecule encoding the *T. reesei* catalytic domain (SEQ ID NO:83) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* α MATpre signal peptide (SEQ ID NO:13), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* promoter (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence (SEQ ID NO:24). The cassette is flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and complete ORF of the *PRO1* gene (SEQ ID NO:89) followed by a *P. pastoris* *ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *PRO1* gene (SEQ ID NO:90). The plasmid contains the *PpARG1* gene. Plasmid pGLY3673 was transformed into strains **YGLY7965** and **YGLY7961** to produce a number of strains of which strains **YGLY78316** and **YGLY8323** were selected from the strains produced.

Plasmid pGLY6833 (**Figure 16**) is a roll-in integration plasmid encoding the light and heavy chains of an anti-Her2 antibody that targets the *TRP2* locus in *P. pastoris*. The expression cassette encoding the anti-Her2 heavy chain comprises a nucleic acid molecule encoding the heavy chain ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:15) operably linked at the 5' end to a nucleic acid molecule encoding the *Saccharomyces cerevisiae* mating factor pre-signal sequence (SEQ ID NO:14) which in turn is fused at its N-terminus to a nucleic acid molecule that has the inducible *P. pastoris* *AOX1* promoter sequence (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule that has the *P. pastoris* *CITI* transcription termination sequence (SEQ ID NO:85). The expression cassette encoding the anti-Her2 light chain comprises a nucleic acid molecule encoding the light chain ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:17) operably linked at the 5' end to

a nucleic acid molecule encoding the *Saccharomyces cerevisiae* mating factor pre-signal sequence (SEQ ID NO:14) which in turn is fused at its *N*-terminus to a nucleic acid molecule that has the inducible *P. pastoris AOX1* promoter sequence (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule that has the *P. pastoris CIT1* transcription termination sequence (SEQ ID NO:85). For selecting transformants, the plasmid comprises an expression cassette encoding the Zeocin ORF in which the nucleic acid molecule encoding the ORF (SEQ ID NO:35) is operably linked at the 5' end to a nucleic acid molecule having the *S. cerevisiae TEF* promoter sequence (SEQ ID NO:37) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:24). The plasmid further includes a nucleic acid molecule for targeting the *TRP2* locus (SEQ ID NO:91).

Plasmid p GLY6564 (**Figure 17**) is a roll-in integration plasmid encoding the light and heavy chains of an anti-RSV antibody that targets the *TRP2* locus in *P. pastoris*. The expression cassette encoding the anti-RSV heavy chain comprises a nucleic acid molecule encoding the heavy chain ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:19) operably linked at the 5' end to a nucleic acid molecule encoding the *Saccharomyces cerevisiae* mating factor pre-signal sequence (SEQ ID NO:14) which in turn is fused at its *N*-terminus to a nucleic acid molecule that has the inducible *P. pastoris AOX1* promoter sequence (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule that has the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:24). The expression cassette encoding the anti-RSV light chain comprises a nucleic acid molecule encoding the light chain ORF codon-optimized for effective expression in *P. pastoris* (SEQ ID NO:21) operably linked at the 5' end to a nucleic acid molecule encoding the *Saccharomyces cerevisiae* mating factor pre-signal sequence (SEQ ID NO:14) which in turn is fused at its *N*-terminus to a nucleic acid molecule that has the inducible *P. pastoris AOX1* promoter sequence (SEQ ID NO:23) and at the 3' end to a nucleic acid molecule that has the *P. pastoris AOX1* transcription termination sequence (SEQ ID NO:36). For selecting transformants, the plasmid comprises an expression cassette encoding the Zeocin ORF in which the nucleic acid molecule encoding the ORF (SEQ ID NO:35) is operably linked at the 5' end to a nucleic acid molecule having the *S. cerevisiae TEF* promoter sequence (SEQ ID NO:37) and at the 3' end to a nucleic acid molecule having the *S. cerevisiae CYC* transcription termination sequence (SEQ ID NO:24). The plasmid further includes a nucleic acid molecule for targeting the *TRP2* locus (SEQ ID NO:91).

Strain **YGLY13992** was generated by transforming pGLY6833, which encodes the anti-Her2 antibody, into **YGLY8316**. The strain **YGLY13992** was selected from the strains produced. In this strain, the expression cassettes encoding the anti-Her2 heavy and light chains are targeted to the *Pichia pastoris TRP2* locus (PpTRP2). This strain does not include the *LmSTT3D* expression cassette. Strain **YGLY14401** was generated by transforming pGLY6564, which encodes the anti-RSV antibody, into **YGLY8323**. The strain **YGLY14401** was selected

from the strains produced. In this strain, the expression cassettes encoding the anti-RSV heavy and light chains are targeted to the *Pichia pastoris* *TRP2* locus (PpTRP2). This strain does not include the *LmSTT3D* expression cassette.

Transformation of the appropriate strains disclosed herein with the above
 5 *LmSTT3D* expression/integration plasmid vectors was performed essentially as follows. Appropriate *Pichia pastoris* strains were grown in 50 mL YPD media (yeast extract (1%), peptone (2%), and dextrose (2%)) overnight to an OD of about 0.2 to 6. After incubation on ice for 30 minutes, cells were pelleted by centrifugation at 2500-3000 rpm for five minutes. Media was removed and the cells washed three times with ice cold sterile 1 M sorbitol before
 10 resuspension in 0.5 mL ice cold sterile 1 M sorbitol. Ten μ L linearized DNA (5-20 μ g) and 100 μ L 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 μ F, 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
 15 hours to overnight at room temperature (24° C) before plating the cells on selective media.

Strains **YGLY13992** and **YGLY14401** were each then transformed with pGLY6301, which encodes the *LmSTT3D* under the control of the inducible *AOX1* promoter, or pGLY6294, which encodes the *LmSTT3D* under the control of the constitutive *GAPDH* promoter, as described above to produce the strains described in Example 3.

20

EXAMPLE 3

Integration/expression plasmid pGLY6301, which comprises the expression cassette in which the ORF encoding the *LmSTT3D* is operably-linked to the inducible *PpAOX1* promoter, or pGLY6294, which comprises the expression cassette in which the ORF encoding
 25 the *LmSTT3D* is operably-linked to the constitutive *PpGAPDH* promoter, was linearized with *SpeI* or *SfiI*, respectively, and the linearized plasmids transformed into *Pichia pastoris* strain **YGLY13992** or **YGLY14401** to produce strains **YGLY17351**, **YGLY17368**, **YGLY17319**, and **YGLY17354** shown in Table 1. Transformations were performed essentially as described in Example 2.

30

Table 1		
Strain	Antibody	<i>LmSTT3D</i> expression
YGLY13992	Anti-Her2	none
YGLY17351	Anti-Her2	inducible
YGLY17368	Anti-Her2	constitutive
YGLY14401	Anti-RSV	none
YGLY17319	Anti-RSV	inducible

YGLY17354	Anti-RSV	constitutive
-----------	----------	--------------

The genomic integration of pGLY6301 at the *URA6* locus was confirmed by colony PCR (cPCR) using the primers, PpURA6out/UP (5'-

CTGAGGAGTCAGATATCAGCTCAATCTCCAT-3'; SEQ ID NO: 1) and Puc19/LP (5'-

TCCGGCTCGTATGTTGTGTGGAATTGT-3'; SEQ ID NO: 2) or ScARR3/UP (5'-

GGCAATAGTCGCGAGAATCCTTAAACCAT-3'; SEQ ID NO: 3) and PpURA6out/LP (5'-

CTGGATGTTTGATGGGTTTCAGTTTCAGCTGGA-3'; SEQ ID NO: 4).

The genomic integration of pGLY6294 at the *TRP1* locus was confirmed by cPCR using the primers, PpTRP-5'out/UP (5'- CCTCGTAAAGATCTGCGGTTTGCAAAGT-3'; SEQ

ID NO: 5) and PpALG3TT/LP (5'-CCTCCCACTGGAACCGATGATATGGAA-3'; SEQ ID

NO: 6) or PpTEFTT/UP (5'-GATGCGAAGTTAAGTGCGCAGAAAGTAATATCA-3'; SEQ ID

NO: 7) and PpTRP1-3'out/LP (5'-CGTGTGTACCTTGAAACGTCAATGATACTTTGA-3';

SEQ ID NO: 8). Integration of the expression cassette encoding the *LmSTT3D* into the genome was confirmed using cPCR primers, *LmSTT3D*/iUP (5'-

GCGACTGGTTCCAATTGACAAGCTT-3' (SEQ ID NO: 9) and *LmSTT3D*/iLP (5'-

CAACAGTAGAACCAGAAGCCTCGTAAGTACAG-3' (SEQ ID NO: 10). The PCR

conditions were one cycle of 95 °C for two minutes, 35 cycles of 95° C for 20 seconds, 55° C for 20 seconds, and 72° C for one minute; followed by one cycle of 72° C for 10 minutes.

The strains were cultivated in a Sixfor fermentor to produce the antibodies for *N*-glycosylation site occupancy analysis. Cell growth conditions of the transformed strains for antibody production were generally as follows.

Protein expression for the transformed yeast strains was carried out at in shake flasks at 24° C with buffered glycerol-complex medium (BMGY) consisting of 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. The induction medium for protein expression was buffered methanol-complex medium (BMMY) consisting of 1% methanol instead of glycerol in BMGY. Pmt inhibitor Pmti-3 in methanol was added to the growth medium to a final concentration of 18.3 μM at the time the induction medium was added. Cells were harvested and centrifuged at 2,000 rpm for five minutes.

SixFors Fermentor Screening Protocol followed the parameters shown in Table 2.

Table 2		
SixFors Fermentor Parameters		
Parameter	Set-point	Actuated Element
pH	6.5 ± 0.1	30% NH ₄ OH
Temperature	24 ± 0.1	Cooling Water & Heating Blanket

Dissolved O ₂	n/a	Initial impeller speed of 550 rpm is ramped to 1200 rpm over first 10 hr, then fixed at 1200 rpm for remainder of run
--------------------------	-----	---

At time of about 18 hours post-inoculation, SixFors vessels containing 350 mL media A (*See* Table 3 below) plus 4% glycerol were inoculated with strain of interest. A small dose (0.3 mL of 0.2 mg/mL in 100% methanol) of Pmti-3 (5-[[3-(1-Phenyl-2-hydroxy)ethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic Acid) (*See* Published International Application No. WO 2007061631) was added with inoculum. At time about 20 hour, a bolus of 17 mL 50% glycerol solution (Glycerol Fed-Batch Feed, *See* Table 4 below) plus a larger dose (0.3 mL of 4 mg/mL) of Pmti-3 was added per vessel. At about 26 hours, when the glycerol was consumed, as indicated by a positive spike in the dissolved oxygen (DO) concentration, a methanol feed (*See* Table 5 below) was initiated at 0.7 mL/hr continuously. At the same time, another dose of Pmti-3 (0.3 mL of 4 mg/mL stock) was added per vessel. At time about 48 hours, another dose (0.3 mL of 4 mg/mL) of Pmti-3 was added per vessel. Cultures were harvested and processed at time about 60 hours post-inoculation.

Table 3 Composition of Media A		
Soytone L-1	20	g/L
Yeast Extract	10	g/L
KH ₂ PO ₄	11.9	g/L
K ₂ HPO ₄	2.3	g/L
Sorbitol	18.2	g/L
Glycerol	40	g/L
Antifoam Sigma 204	8	drops/L
10X YNB w/Ammonium Sulfate w/o Amino Acids (134 g/L)	100	mL/L
250X Biotin (0.4 g/L)	10	mL/L
500X Chloramphenicol (50 g/L)	2	mL/L
500X Kanamycin (50 g/L)	2	mL/L

Table 4 Glycerol Fed-Batch Feed		
Glycerol	50	% m/m
PTM1 Salts (see Table IV-E below)	12.5	mL/L
250X Biotin (0.4 g/L)	12.5	mL/L

Table 5 Methanol Feed		
Methanol	100	% m/m
PTM1 Salts (<i>See</i> Table 6)	12.5	mL/L
250X Biotin (0.4 g/L)	12.5	mL/L

Table 6 PTM1 Salts		
CuSO ₄ -5H ₂ O	6	g/L
NaI	80	mg/L
MnSO ₄ -7H ₂ O	3	g/L
NaMoO ₄ -2H ₂ O	200	mg/L
H ₃ BO ₃	20	mg/L
CoCl ₂ -6H ₂ O	500	mg/L
ZnCl ₂	20	g/L
FeSO ₄ -7H ₂ O	65	g/L
Biotin	200	mg/L
H ₂ SO ₄ (98%)	5	mL/L

The occupancy of *N*-glycan on anti-Her2 or anti-RSV antibodies was determined
 5 using capillary electrophoresis (CE) as follows. The antibodies were recovered from the cell
 culture medium and purified by protein A column chromatography. The protein A purified
 sample (100-200 µg) was concentrated to about 100 µL and then buffer was exchanged with 100
 mM Tris-HCl pH 9.0 with 1% SDS. Then, the sample along with 2 µL of 10 kDa internal
 standard provided by Beckman was reduced by addition of 5 µL β-mercaptoethanol and boiled
 10 for five minutes. About 20 µL of reduced sample was then resolved over a bare-fused silica
 capillary (about 70 mm, 50 µm I.D.) according to the method recommended by Beckman
 Coulter.

Figure 18 shows the *N*-glycosylation site occupancy of heavy chains from the CE
 analysis. The figure shows that for both antibodies, the amount of *N*-linked heavy chains species
 15 increased from about 80% to about 94% when the *LmSTT3D* was constitutively expressed to
 about 99% when expression of the *LmSTT3D* was induced at the same time as expression of the
 antibodies was induced.

Table 7 shows *N*-glycosylation site occupancy of anti-HER2 and anti-RSV
 antibodies was increased for compositions in which the antibodies were obtained from host cells
 20 in which the *LmSTT3D* was overexpressed in the presence of the endogenous

oligosaccharyltransferase (OST) complex. To determine *N*-glycosylation site occupancy, antibodies were reduced and the *N*-glycan occupancy of the heavy chains determined. The table shows that in general, overexpression of the *LmSTT3D* under the control of an inducible promoter effected an increase of *N*-glycosylation site occupancy from about 82-83% to about 99% for both antibodies tested (about a 19% increase over the *N*-glycosylation site occupancy in the absence of *LmSTT3D* overexpression). The expression of the *LmSTT3D* and the antibodies were under the control of the same inducible promoter. When overexpression of the *LmSTT3D* was under the control of a constitutive promoter the increase in *N*-glycosylation site occupancy was increased to about 94% for both antibodies tested (about a 13% increase over the *N*-glycosylation site occupancy in the absence of *LmSTT3D* overexpression).

Table 7				
Strain	LmSTT3D		Antibody	Heavy Chain <i>N</i> -glycosylation site occupancy [#] (%)
	AOX1 Prom. (pGLY6301) (inducible)	GAPDH Prom. (pGLY6294) (constitutive)		
YGLY13992	None	None	Anti-HER2	83
YGLY17368	None	overexpressed	Anti-HER2	94
YGLY17351	overexpressed	None	Anti-HER2	99
YGLY14401	None	None	Anti-RSV	82
YGLY17354	None	overexpressed	Anti-RSV	94
YGLY17319	overexpressed	None	Anti-RSV	99
# <i>N</i> -glycosylation site occupancy based upon percent glycosylation site occupancy of total heavy chains from reduced antibodies				

Table 8 shows the *N*-glycosylation site occupancy for compositions comprising whole antibodies obtained from host cells in which the *LmSTT3D* was overexpressed in the presence of the endogenous oligosaccharyltransferase (OST) complex based upon the determination of *N*-glycosylation site occupancy of the individual heavy chains from reduced antibody preparations. The formula $(\text{fraction GHC})^2 \times 100$ will provide an estimate or approximation of the percent fully occupied antibodies based upon the determination of the fraction of heavy chains that are *N*-glycosylated.

Table 8			
	LmSTT3D		Fully

Strain	AOX1 Prom. (pGLY6301) (inducible)	GAPDH Prom. (pGLY6294) (constitutive)	Antibody	Occupied Antibodies [#] (%)
YGLY13992	None	None	Anti-HER2	68.9
YGLY17368	None	overexpressed	Anti-HER2	88.4
YGLY17351	overexpressed	None	Anti-HER2	98.0
YGLY14401	None	None	Anti-RSV	67.2
YGLY17354	None	overexpressed	Anti-RSV	88.4
YGLY17319	overexpressed	None	Anti-RSV	98.0
# based upon results obtained from Table 7.				

Q-TOF Analysis

The high performance liquid chromatography (HPLC) system used consisted of an Agilent 1200 equipped with autoinjector, a column-heating compartment and a UV detector detecting at 210 and 280 nm. All LC-MS experiments performed with this system were running at 1 mL/min. The flow rate was not split for MS detection. Mass spectrometric analysis was carried out in positive ion mode on Accurate-Mass Q-TOF LC/MS 6520 (Agilent technology). The temperature of dual ESI source was set at 350 °C. The nitrogen gas flow rates were set at 13 L/h for the cone and 350 l/h and nebulizer was set at 45 psig with 4500 volt applied to the capillary. Reference mass of 922.009 was prepared from HP-0921 according to API-TOF reference mass solution kit for mass calibration and the protein mass measurements. The data for ion spectrum range from 300-3000 m/z were acquired and processed using Agilent Masshunter.

Sample preparation was as follows. An intact antibody sample (50 µg) was prepared 50 µL 25 mM NH₄HCO₃, pH 7.8. For deglycosylated antibody, a 50 µL aliquot of intact antibody sample was treated with PNGase F (10 units) for 18 hours at 37° C. Reduced antibody was prepared by adding 1 M DTT to a final concentration of 10 mM to an aliquot of either intact antibody or deglycosylated antibody and incubated for 30 min at 37° C.

Three micrograms of intact or deglycosylated antibody sample was loaded onto a Poroshell 300SB-C3 column (2.1 mm × 75 mm, 5 µm) (Agilent Technologies) maintained at 70° C. The protein was first rinsed on the cartridge for 1 minute with 90% solvent A (0.1% HCOOH), 5% solvent B (90% Acetonitrile in 0.1% HCOOH). Elution was then performed using a gradient of 5-100% of B over 26 minutes followed by a three-minute regeneration at 100% B and by a final equilibration period of 10 minute at 5% B.

For reduced antibody, a three microgram sample was loaded onto a Poroshell 300SB-C3 column (2.1 mm × 75 mm, 5 µm) (Agilent Technologies) maintained at 40 °C. The protein was first rinsed on the cartridge for three minutes with 90% solvent A, 5% solvent B.

Elution was then performed using a gradient of 5-80% of B over 20 minutes followed by a seven-minute regeneration at 80% B and by a final equilibration period of 10 minutes at 5% B.

Figure 19 shows the results of a Q-TOF analysis in which the *N*-glycosylation site occupancy of non-reduced anti-Her2 antibody produced in **YGLY17351** was compared to *N*-glycosylation site occupancy of non-reduced commercially available anti-Her2 antibody produced in CHO cells (HERCEPTIN). The figure shows that anti-Her2 antibody produced in strain **YGLY17351** has an *N*-glycosylation site occupancy that is like the *N*-glycosylation site occupancy of an anti-Her2 antibody made in CHO cells. The figure shows that the amount of antibodies in which only one *N*-glycosylation site was occupied decreased and the amount of antibodies in which both *N*-glycosylation sites was occupied increased when the antibodies were produced by strain **YGLY17351**. The results shown for anti-Her2 antibody produced in **YGLY17351** were consistent with the approximated occupancy shown in Table 8.

Figure 20 demonstrates the scalability of *N*-glycosylation site occupancy on anti-Her2 antibodies produced in **YGLY17351**. In order to evaluate scalability of *N*-glycan occupancy, **YGLY17351** was tested in bioreactors ranging from 5 mL through 40 L. In general, *N*-glycosylation site occupancy of glycoproteins in glycoengineered *P. pastoris* has been observed to vary with the process conditions used to produce the glycoproteins. However, the *LmSTT3D* overexpressing strains showed very consistent *N*-glycosylation site occupancy (99%) regardless of scale of bioreactors and process conditions. Thus, the present invention provides a method in which the *N*-glycosylation site occupancy of glycoproteins in glycoengineered *P. pastoris* grown under small scale conditions is maintained when grown under large scale conditions.

Figures 21 and 22 are provided for illustrative purposes. **Figure 21** shows the results of a CE and Q-TOF analysis for a commercial lot of anti-Her2 antibody produced in CHO cells (HERCEPTIN). **Figure 22** shows the results of a CE and Q-TOF analysis for the same commercial lot of anti-Her2 antibody following treatment with PNGase F for a time. The CE shows an increase in non-glycosylated heavy chain and the Q-TOF shows the presence of non-glycosylated antibody following PNGase F treatment (compare **Figure 21** to **Figure 22**).

Table 9 shows the *N*-glycan composition of the anti-Her2 and anti-RSV antibodies produced in strains that overexpress *LmSTT3D* compared to strains that do not overexpress *LmSTT3D*. The Figure confirms that the quality of *N*-glycans of antibodies from *LmSTT3D* overexpressing strains is comparable to that from strains that do not overexpress *LmSTT3D*. Antibodies were produced from SixFors (0.5L bioreactor) and *N*-glycans from protein A-purified antibodies were analyzed with 2AB labeling. Overall, overexpression of *LmSTT3D* did not appear to significantly affect the *N*-glycan composition of the antibodies. The glycosylation composition can vary as a function of fermentation conditions, therefore, the glycosylation composition of antibodies produced in *Pichia pastoris* strains can range from about

50-70 mole%G0, 15-25 mole%G1, 4-12% mole% G2, 5-17 mole% Man5, and 3-15 mole% hybrids.

Table 9						
		N-glycans (%)				
	<i>Lm</i> STT3D	G0	G1	G2	Man5	Hybrids
Anti-Her2	none	58.1±1.8	20.5±0.6	3.0±0.9	14.0±2.1	4.3±1.2
Antibody	overexpressed	53.9±2.0	22.4±3.0	4.5±1.7	14.7±1.5	4.2±1.5
Anti-RSV	none	51.6±1.6	22.9±2.0	5.3±2.2	15.2±1.1	4.9±0.6
Antibody	overexpressed	58.4±5.3	20.9±2.8	3.5±0.3	12.4±0.1	4.7±2.3
G0 - GlcNAc ₂ Man ₃ GlcNAc ₂						
G1 - GalGlcNAc ₂ Man ₃ GlcNAc ₂						
G2 - Gal ₂ GlcNAc ₂ Man ₃ GlcNAc ₂						
Man5 - Man ₅ GlcNAc ₂						
Hybrid - GlcNAcMan ₅ GlcNAc ₂ and/or GalGlcNAcMan ₅ GlcNAc ₂						

- 5 Table 10 shows a comparison of the glycosylation pattern of the anti-RSV antibody produced in strain **YGLY14401** compared to several commercial lots of an anti-RSV antibody produced in CHO cells and marketed as palivizumab under the tradename SYNAGIS.

Table 10			
Glycoform	SYNAGIS (Commercial lot 07A621)	SYNAGIS (Commercial lot 09A621)	Anti-RSV antibody produced in YGLY14401
	% of total	% of total	% of total
Man5	6.4	6.8	9.5
G0	< 1.0	< 1.0	59.9
G0F	33.9	30	0
G1	< 1.0	< 1.0	20
G1F	41.7	48.8	0
G2	0	0	2.8
G2F	10.9	12.3	0
A2	5.1	3.7	0
Hybrid	-	-	7.8
O-glycans occupancy (mol/mol)	0	0	3.0
Mannose (single mannose)	0	0	96

Mannobiose (two mannose residues)	0	0	4

This example shows then that the present invention enables the production of recombinant glycoproteins in *Pichia pastoris* in which the *N*-glycosylation site occupancy of the recombinant glycoproteins is comparable to the *N*-glycosylation site occupancy of recombinant glycoproteins produced in mammalian expression systems such as CHO cells.

EXAMPLE 4

The *Leishmania major* STT3A protein, *Leishmania major* STT3B protein, and *Leishmania major* STT3D protein, are all examples of heterologous single-subunit oligosaccharyltransferases that have been shown to suppress the lethal phenotype of a deletion of the *STT3* locus in *Saccharomyces cerevisiae* (Naseb *et al.*, Molec. Biol. Cell 19: 3758-3768 (2008)). Naseb *et al.* (ibid.) further showed that the *Leishmania major* STT3D protein could suppress the lethal phenotype of a mutation of the *WBP1*, *OST1*, *SWP1*, or *OST2* loci in *Saccharomyces cerevisiae*. Hese *et al.* (Glycobiology 19: 160-171 (2009)) provides data that suggest the *Leishmania major* STT3A, STT3B, and STT3D proteins can functionally complement mutations of the *WBP1*, *OST1*, *SWP1*, and *OST2* loci. Other single-subunit heterologous oligosaccharyltransferases include but are not limited to single-subunit *Giardia* or kinetoplastid STT3 proteins, for example, the *Caenorhabditis elegans* STT3 protein, *Trypanosoma brucei* STT3 protein, *Trypanosoma cruzi* STT3 protein, and *Toxoplasma gondii* STT3 protein. In contrast to the *Leishmania major* STT3D protein, which Naseb *et al.* (op. cit.) teaches does not incorporate into the *Saccharomyces cerevisiae* OTase complex, Castro *et al.* (Proc. Natl. Acad. Sci. USA 103: 14756-14760 (2006)) teaches that the *Trypanosoma cruzi* STT3 appears to integrate into the *Saccharomyces cerevisiae* OTase complex.

In this example, host cells constructed similar to the host cells in the previous example were transformed with plasmid vectors containing expression cassettes encoding the STT3 protein from *Caenorhabditis elegans*, *Trypanosoma cruzi*, and *Leishmania major* STT3C operably linked to the AOX1 promoter. A vector containing an expression cassette encoding the *Pichia pastoris* Stt3p was included in the experiment. As shown in Table 11, expression of the various STT3 proteins concurrently with expression of the anti-Her2 antibody did not appear result in an increase in *N*-glycosylation site occupancy. However, various STT3 proteins can display substrate specificity. For example, the *Leishmania major* STT3A, B, C, and D proteins differ in substrate specificity at the level of glycosylation, which suggests that in addition to the essential N-X-S/T attachment site additional features of the substrate may influence *N*-linked glycosylation at a particular attachment site (Naseb *et al.*, op cit.). The results shown in Table 9 used the anti-Her2 antibody as the substrate. The C_{H2} domain of each heavy chain of an

antibody contains a single site for *N*-linked glycosylation: this is usually at the asparagine residue 297 (Asn-297) (Kabat *et al.*, Sequences of proteins of immunological interest, Fifth Ed., U.S. Department of Health and Human Services, NIH Publication No. 91-3242). Thus, the results shown in Table 9 suggest that the percent *N*-glycosylation site occupancy might be influenced by the substrate specificity of the particular single-subunit oligosaccharyltransferase being used.

Table 11			
STT3 (AOX1 Prom)		Antibody	N-glycosylation site occupancy (%)
<i>C. elegans</i>	overexpressed	Anti-Her2	83
<i>T. cruzi</i>	overexpressed	Anti-Her2	83
<i>L. major</i> (STT3C)	overexpressed	Anti-Her2	82
<i>P. pastoris</i>	overexpressed	Anti-Her2	80

EXAMPLE 5

A strain capable of producing sialylated *N*-glycans was constructed as follows. The strain was transfected with a plasmid vector encoding human GM-CSF and a plasmid vector encoding the *Leishmania major* STT3D. Construction of the strains is illustrated schematically in **Figure 23A-23D**. Briefly, the strains were constructed as follows.

Plasmid pGLY2456 (**Figure 24**) is a KINKO integration vector that targets the *TRP2* locus without disrupting expression of the locus and contains six expression cassettes encoding (1) the mouse CMP-sialic acid transporter (mCMP-Sia Transp), (2) the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase (hGNE), (3) the *Pichia pastoris* *ARG1* gene or transcription unit, (4) the human CMP-sialic acid synthase (hCSS), (5) the human *N*-acetylneuraminate-9-phosphate synthase (hSPS), (6) the mouse α -2,6-sialyltransferase catalytic domain (mST6) fused at the *N*-terminus to *S. cerevisiae* *KRE2* leader peptide (33) to target the chimeric enzyme to the ER or Golgi, and the *P. pastoris* *ARG1* gene or transcription unit. The expression cassette encoding the mouse CMP-sialic acid transporter comprises a nucleic acid molecule encoding the mCMP Sia Transp ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:92), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence. The expression cassette encoding the human UDP-GlcNAc 2-epimerase/*N*-acetylmannosamine kinase comprises a nucleic acid molecule encoding the hGNE ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:93), which

is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The expression cassette encoding the *P. pastoris* *ARG1* gene comprises (SEQ ID NO:94). The expression cassette encoding the human CMP-sialic acid synthase comprises a nucleic acid molecule encoding the hCSS ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:95), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *GAPDH* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae* *CYC* transcription termination sequence. The expression cassette encoding the human N-acetylneuraminase-9-phosphate synthase comprises a nucleic acid molecule encoding the hSIAP S ORF codon optimized for expression in *P. pastoris* (SEQ ID NO:96), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *PMAl* transcription termination sequence. The expression cassette encoding the chimeric mouse α -2,6-sialyltransferase comprises a nucleic acid molecule encoding the mST6 catalytic domain codon optimized for expression in *P. pastoris* (SEQ ID NO:97) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* *KRE2* signal peptide, which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *TEF* promoter and at the 3' end to a nucleic acid molecule comprising the *P. pastoris* *TEF* transcription termination sequence. The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP2* gene ending at the stop codon (SEQ ID NO:98) followed by a *P. pastoris* *ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP2* gene (SEQ ID NO:99). Plasmid pGLY2456 was linearized with *Sfi*I and the linearized plasmid transformed into strain **YGLY7961** to produce a number of strains in which the six expression cassette have been inserted into the *TRP2* locus immediately following the *TRP2* ORF by double-crossover homologous recombination. The strain **YGLY8146** was selected from the strains produced. The strain was then counterselected in the presence of 5-FOA to produce a number of strains now auxotrophic for uridine. Strain **YGLY9296** was selected.

Plasmid pGLY5048 (**Figure 25**) is an integration vector that targets the *STE13* locus and contains expression cassettes encoding (1) the *T. reesei* α -1,2-mannosidase catalytic domain fused at the *N*-terminus to *S. cerevisiae* α MATpre signal peptide (aMATTrMan) to target the chimeric protein to the secretory pathway and secretion from the cell and (2) the *P. pastoris* *URA5* gene or transcription unit. The expression cassette encoding the aMATTrMan comprises a nucleic acid molecule encoding the *T. reesei* catalytic domain (SEQ ID NO:83) fused at the 5' end to a nucleic acid molecule encoding the *S. cerevisiae* α MATpre signal peptide (SEQ ID NO:13), which is operably linked at the 5' end to a nucleic acid molecule comprising the *P. pastoris* *AOX1* promoter and at the 3' end to a nucleic acid molecule comprising the *S. cerevisiae*

CYC transcription termination sequence. The *URA5* expression cassette comprises a nucleic acid molecule comprising the *P. pastoris URA5* gene or transcription unit flanked by nucleic acid molecules comprising *lacZ* repeats. The two tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region of the *STE13* gene (SEQ ID NO:100) and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *STE13* gene (SEQ ID NO:101). Plasmid pGLY5048 was linearized with *SfiI* and the linearized plasmid transformed into strain **YGLY9296** to produce a number of strains. The strain **YGLY9469** was selected from the strains produced. This strain is capable of producing glycoproteins that have single-mannose *O*-glycosylation (See Published U.S. Application No. 20090170159).

Plasmid pGLY5019 (**Figure 26**) is an integration vector that targets the *DAP2* locus and contains an expression cassette comprising a nucleic acid molecule encoding the Nourseothricin resistance (NATR) expression cassette (originally from pAG25 from EROSCARF, Scientific Research and Development GmbH, Daimlerstrasse 13a, D-61352 Bad Homburg, Germany, See Goldstein *et al.*, Yeast 15: 1541 (1999)). The NAT^R expression cassette (SEQ ID NO:34) is operably regulated to the *Ashbya gossypii TEF1* promoter and *A. gossypii TEF1* termination sequences flanked one side with the 5' nucleotide sequence of the *P. pastoris DAP2* gene (SEQ ID NO:102) and on the other side with the 3' nucleotide sequence of the *P. pastoris DAP2* gene (SEQ ID NO:103). Plasmid pGLY5019 was linearized and the linearized plasmid transformed into strain **YGLY9469** to produce a number of strains in which the NATR expression cassette has been inserted into the *DAP2* locus by double-crossover homologous recombination. The strain **YGLY9797** was selected from the strains produced.

Plasmid pGLY5085 (**Figure 27**) is a KINKO plasmid for introducing a second set of the genes involved in producing sialylated *N*-glycans into *P. pastoris*. The plasmid is similar to plasmid YGLY2456 except that the *P. pastoris ARG1* gene has been replaced with an expression cassette encoding hygromycin resistance (HygR) and the plasmid targets the *P. pastoris TRP5* locus. The HYG^R resistance cassette is SEQ ID NO:104. The HYG^R expression cassette (SEQ ID NO:104) is operably regulated to the *Ashbya gossypii TEF1* promoter and *A. gossypii TEF1* termination sequences (See Goldstein *et al.*, Yeast 15: 1541 (1999)). The six tandem cassettes are flanked on one side by a nucleic acid molecule comprising a nucleotide sequence from the 5' region and ORF of the *TRP5* gene ending at the stop codon (SEQ ID NO:105) followed by a *P. pastoris ALG3* termination sequence and on the other side by a nucleic acid molecule comprising a nucleotide sequence from the 3' region of the *TRP5* gene (SEQ ID NO:106). Plasmid pGLY5085 was transformed into strain **YGLY9797** to produce a number of strains of which strain **YGLY1200** was selected.

Plasmid pGLY7240 (**Figure 28**), which targets the *Pichia pastoris TRP2* locus (PpTRP2), encodes a fusion protein comprising the human GM-CSF fused to the *Pichia pastoris*

CWP1 protein via a linker containing a Kex2 cleavage site. The CWP1 protein is removed from GM-CSF in the late Golgi by the Kex2 endoprotease so that free GM-CSF is secreted into the fermentation supernatant. The human GM-CSF has the amino acid sequence shown in SEQ ID NO:108 and is encoded by the nucleotide sequence shown in SEQ ID NO:108. The fusion protein (SEQ ID NO:109) is encoded by the nucleotide sequence shown in SEQ ID NO:110. The CWP1 signal sequence is amino acids 1-18, the CWP1 amino acid sequence is from amino acids 19-289, the GGGSLVKR Kex2 linker amino acid sequence (SEQ ID NO:111) is from amino acids 290-297, and the GM-CSF amino acid sequence is from amino acids 298-424. The expression of the fusion protein is operably linked to the Pp AOX1 promoter and ScCYC termination sequences. Plasmid **pGLY7240** was transformed into strain **YGLY12900** to produce a number of strains of which strain **YGLY15660** was selected. Strain **YGLY15660** was transformed with plasmid **pGLY6301** (encodes *Leishmania major* STT3D) to produce a number of strains of which **YGLY16349** was selected.

Figure 29 shows that LmSTT3D also improved *N*-glycan occupancy of non-antibody glycoprotein, GM-CSF. GM-CSF contains 2 N-linked sites and in wild-type *Pichia* 1 N-linked site on GM-CSF is predominantly glycosylated. To investigate impact of LmSTT3D on *N*-glycan occupancy of GM-CSF, methanol-inducible LmSTT3D was overexpressed in GM-CSF producing strain, yGLY15560. *N*-glycan occupancy was evaluated using Micro24 bioreactor (M24). The cell-free supernatants from M24 were analyzed for *N*-glycan occupancy using Western blot and 15% SDS-PAGE. As shown in Western blot detected with GM-CSF specific antibody, the majority of GM-CSF (Lanes 2-8) is glycosylated with 2N-linked sites in contrast to the control strain (yGLY15560, lane 9) where GM-CSF is predominantly N-glycosylated with 1 site along with the minor portions of 2 N sites and non-glycosylated. Taken together, this indicates that LmSTT3D can improve *N*-glycan occupancy of glycoproteins carrying multiple N-linked sites.

Figure 30 shows Q-TOP analysis of GM-CSF expressed from yGLY15560 (A) and yGLY16349 (B), respectively. This analysis confirms that the majority of GM-CSF is glycosylated with 2N-linked sites in the presence of LmSTT3D as shown in **Figure 29**. Non-glycosylated GM-CSF was not detected.

LC-ESI-TOF

The high performance liquid chromatography (HPLC) system used in this study consisted of an Agilent 1200 equipped with autoinjector, a column-heating compartment and a UV detector detecting at 210 and 280 nm. All LC-MS experiments performed with this system were running at 1 ml/min. The flow rate was not split for MS detection. Mass spectrometric analysis was carried out in positive ion mode on Accurate-Mass Q-TOF LC/MS 6520 (Agilent technology). The temperature of dual ESI source was set at 350°C. The nitrogen gas flow rates were set at 13

l/h for the cone and 350 l/h and nebulizer was set at 45 psig with 4500 volt applied to the capillary. eference mass of 922.009 was prepared from HP-0921 according to API-TOF reference mass solution kit for mass calibration and the protein mass measurements. The data for ion spectrum range from 300-3000 m/z were acquired and processed using Agilent Masshunter.

5

Sample preparation

An intact antibody sample (50 ug) was prepared 50 ul 25 mM NH_4HCO_3 , pH 7.8. For deglycosylated antibody, a 50 ul aliquot of intact antibody sample was treated with PNGase F (10 units) for 18 hr at 37C. Reduced antibody was prepared by adding 1 M DTT to a final concentration of 10 mM to an aliquot of either intact antibody or deglycosylated antibody and incubated for 30 min at 37C.

Three microgram of intact or deglycosylated antibody sample was loaded onto a Poroshell 300SB-C3 column (2.1 mm $\times$ 75 mm, 5 μm) (Agilent Technologies) maintained at 70 °C. The protein was first rinsed on the cartridge for 1 min with 90% solvent A (0.1% HCOOH), 5% solvent B (90% Acetonitrile in 0.1% HCOOH). Elution was then performed using an gradient of 5-100% of B over 26 min followed by a 3 min regeneration at 100% B and by a final equilibration period of 10 min at 5% B.

For reduced antibody, three microgram sample was loaded a Poroshell 300SB-C3 column (2.1 mm $\times$ 75 mm, 5 μm) (Agilent Technologies) maintained at 40 °C. The protein was first rinsed on the cartridge for 3 min with 90% solvent A, 5% solvent B. Elution was then performed using an gradient of 5-80% of B over 20 min followed by a 7 min regeneration at 80% B and by a final equilibration period of 10 min at 5% B.

SEQUENCES

Sequences that were used to produce some of the strains disclosed in Examples 1-4 are provided in Table 12.

Table 12 BRIEF DESCRIPTION OF THE SEQUENCES		
SEQ ID NO:	Description	Sequence
1	PCR primer PpURA6out/UP	CTGAGGAGTCAGATATCAGCTCAATCTCCAT
2	PCR primer Puc19/LP	TCCGGCTCGTATGTTGTGTGGAATTGT
3	PCR primer PpURA6out/LP	CTGGATGTTTGATGGGTTTCAGTTTCAGCTGGA
4	PCR primer ScARR3/UP	GGCAATAGTCGCGAGAATCCTTAAACCAT
5	PCR primer	CCTCGTAAAGATCTGCGGTTTGCAAAGT

	PpTRP1-5'out/UP	
6	PCR primer PpALG3TT/LP	CCTCCCACTGGAACCGATGATATGGAA
7	PCR primer PpTEFTT/UP	GATGCGAAGTTAAGTGCGCAGAAAGTAATATCA
8	PCR primer PpTRP-3'out/LP	CGTGTGTACCTTGAAACGTCAATGATACTTTGA
9	PCR primer LmSTT3D/iUP	CAGACTAAGACTGCTTCTCCACCTGCTAAG
10	PCR primer LmSTT3D/iLP	CAACAGTAGAACCAGAAGCCTCGTAAGTACAG
11	<i>Leishmania major</i> STT3D (DNA)	ATGGGTAAAAGAAAGGGAAACTCCTTGGGAGATTCTG GTTCTGCTGCTACTGCTTCCAGAGAGGCTTCTGCTCAA GCTGAAGATGCTGCTTCCCAGACTAAGACTGCTTCTCC ACCTGCTAAGGTTATCTTGTTGCCAAAGACTTTGACTG ACGAGAAGGACTTCATCGGTATCTTCCCATTTCCATTC TGGCCAGTTCACTTCGTTTTGACTGTTGTTGCTTTGTTT GTTTTGGCTGCTTCCTGTTTCCAGGCTTTCAGTGTTAG AATGATCTCCGTTCAAATCTACGGTACTTGATCCACG AATTTGACCCATGGTTCAACTACAGAGCTGCTGAGTA CATGTCTACTCACGGATGGAGTGCTTTTTTCTCCTGGT TCGATTACATGTCCTGGTATCCATTGGGTAGACCAGTT GGTTCTACTACTTACCCAGGATTGCAGTTGACTGCTGT TGCTATCCATAGAGCTTTGGCTGCTGCTGGAATGCCAA TGTCCTTGAACAATGTTTGTGTTTGGATGCCAGCTTGG TTTGGTGCTATCGCTACTGCTACTTTGGCTTTCTGTACT TACGAGGCTTCTGGTTCTACTGTTGCTGCTGCTGCAGC TGCTTTGTCCTTCTCCATTATCCCTGCTCACTTGATGAG ATCCATGGCTGGTGAGTTTCGACAACGAGTGATTGCT GTTGCTGCTATGTTGTTGACTTTCTACTGTTGGGTTCTGT TCCTTGAGAACTAGATCCTCCTGGCCAATCGGTGTTTT GACAGGTGTTGCTTACGGTTACATGGCTGCTGCTTGGG GAGGTTACATCTTCGTTTTGAACATGGTTGCTATGCAC GCTGGTATCTCTTCTATGGTTGACTGGGCTAGAAACAC TTACAACCCATCCTTGTTGAGAGCTTACACTTTGTTCT ACGTTGTTGGTACTGCTATCGCTGTTTGTGTTCCACCA GTTGGAATGTCTCCATTCAAGTCCTTGGAGCAGTTGGG AGCTTTGTTGGTTTTGGTTTTCTTGTGTGGATTGCAAG TTTGTGAGGTTTTGAGAGCTAGAGCTGGTGTGGAAGTT AGATCCAGAGCTAATTTCAAGATCAGAGTTAGAGTTT TCTCCGTTATGGCTGGTGTGCTGCTTTGGCTATCTCT GTTTTGGCTCCAAGTGGTTACTTTGGTCCATTGTCTGTT AGAGTTAGAGCTTTGTTTGTGAGCACACTAGAACTG GTAACCCATTGGTTGACTCCGTTGCTGAACATCAACCA GCTTCTCCAGAGGCTATGTGGGCTTTCTTGCATGTTTG TGGTGTTACTTGGGGATTGGGTTCCATTGTTTTGGCTG

		TTTCCACTTTCGTTCACTACTCCCCATCTAAGGTTTTCT GGTTGTTGAACTCCGGTGCTGTTTACTACTTCTCCACT AGAATGGCTAGATTGTTGTTGTTGTCCGGTCCAGCTGC TTGTTTGTCCACTGGTATCTTCGTTGGTACTATCTTGG AGGCTGCTGTTCAATTGTCTTTCTGGGACTCCGATGCT ACTAAGGCTAAGAAGCAGCAAAAGCAGGCTCAAAGA CACCAAAGAGGTGCTGGTAAAGGTTCTGGTAGAGATG ACGCTAAGAACGCTACTACTGCTAGAGCTTTCTGTGA CGTTTTCGCTGGTTCTTCTTTGGCTTGGGGTACAGAA TGGTTTTGTCCATTGCTATGTGGGCTTTGGTTACTACT ACTGCTGTTTCTTCTTCTCCTCCGAATTTGCTTCTCAC TCCACTAAGTTCGCTGAACAATCCTCCAACCCAATGAT CGTTTTCGCTGCTGTTGTTTCAAGACAGAGCTACTGGAA AGCCAATGAACCTGTTGGTTGACGACTACTTGAAGGC TTACGAGTGGTTGAGAGACTCTACTCCAGAGGACGCT AGAGTTTTGGCTTGGTGGGACTACGGTTACCAAATCA CTGGTATCGGTAAACAGAACTTCCTTGGCTGATGGTAA CACTTGAACCACGAGCACATTGCTACTATCGGAAAG ATGTTGACTTCCCCAGTTGTTGAAGCTCACTCCCTTGT TAGACACATGGCTGACTACGTTTTGATTTGGGCTGGTC AATCTGGTGACTTGATGAAGTCTCCACACATGGCTAG AATCGGTAACCTCTGTTTACCACGACATTTGTCCAGATG ACCCATTGTGTCAGCAATTCGGTTTCCACAGAAACGA TTACTCCAGACCAACTCCAATGATGAGAGCTTCCTTGT TGTACAACCTTGCACGAGGCTGGAAAAAGAAAGGGTGT TAAGGTTAACCCATCTTTGTTCCAAGAGGTTTACTCCT CCAAGTACGGACTTGTTAGAATCTTCAAGGTTATGAA CGTTTCCGCTGAGTCTAAGAAGTGGGTTGCAGACCCA GCTAACAGAGTTTGTCAACCCACCTGGTTCTTGGATTTG TCCTGGTCAATACCCACCTGCTAAAGAAATCCAAGAG ATGTTGGCTCACAGAGTTCCATTCGACCAGGTTACAA ACGCTGACAGAAAGAACAATGTTGGTTCCTACCAAGA GGAATACATGAGAAGAATGAGAGAGTCCGAGAACAG AAGATAATAG
12	<i>Leishmania</i> major STT3D (protein)	MGKRKGNLSLGDSGSAATASREASAQAEDAASQTKTASP PAKVILLPKTLTDEKDFIGIFPFPPFWPVHFLTVVALFVLA ASCFQAFTVRMISVQIYGyliHEFDPWFNYRAAEYMSH GWSAFFSWFDYMSWYPLGRPVGSTTYPGLQLTAVAIHR ALAAAGMPMSLNNVCVLPWFGAIATATLAFCTYEAS GSTVAAAAAALSFSIIPAHLMRSMAGEFDNECIAVAAML LTFYCWVRSRLRTRSSWPIGVLTGVAYGYMAAAWGGYIF VLNMVAMHAGISSMVDWARNTYNPSLLRAYTLFYVVG TAIAVCVPPVGMSPFKSLEQLGALLVLVFLCGLQVCEVL RARAGVEVRSRANFKIRVRVFSVMAGVAALAIISVLAPTG YFGPLSVRVRALFVEHTRTGNPLVDSVAEHQPASPEAM WAFLHVCVGTWGLGSIVLAVSTFVHYSKVFVWLLNSG AVYYFSTRMARLLLLSGPAACLSTGIFVGITILEAAVQLSF WDSDATKAKKQKQQAQRHQRGAGKGSGRDDAKNATT ARAFCDVFAGSSLAWGHRMVLSIAMWALVTTTAVSFFS

		SEFASHSTKFAEQSSNPMIVFAAVVQNRATGKPMNLLVD DYLKAYEWLRDSTPEDARVLA WWDYGYQITGIGNRTSL ADGNTWNHEHIATIGKMLTSPVVEAHS LVRHMADYVLI WAGQSGDLMKSPH MARIGNSVYHDICPDDPLCQQFGFH RNDYSRPTPMMRASLLYNLHEAGKRKGVKVNPSLFQEV YSSKYGLVRIFKVMNVSAESKKWVADPANRVCHPPGS WICPGQYPPAKEIQEMLAHRVPFDQVTNADRKNNVGSY QEEYMRRMRESENRR
13	<i>Saccharomyces cerevisiae</i> mating factor pre-signal peptide (DNA)	ATGAGATTCCCATCCATCTTCACTGCTGTTTTGTTTCG TGCTTCTTCTGCTTTGGCT
14	<i>Saccharomyces cerevisiae</i> mating factor pre-signal peptide (protein)	MRFPSIFTAVLFAASSALA
15	Anti-Her2 Heavy chain (VH + IgG1 constant region) (DNA)	GAGGTTTCAGTTGGTTGAATCTGGAGGAGGATTGGTTC AACCTGGTGGTTCTTTGAGATTGTCCTGTGCTGCTTCC GGTTTCAACATCAAGGACACTTACATCCACTGGGTTA GACAAGCTCCAGGAAAGGGATTGGAGTGGGTGCTAG AATCTACCCAAC TAACGGTTACACAAGATACGCTGAC TCCGTTAAGGGAAGATTCACTATCTCTGCTGACACTTC CAAGAACACTGCTTACTTGCAGATGAACTCCTTGAGA GCTGAGGATACTGCTGTTTACTACTGTTCCAGATGGGG TGGTGATGGTTTCTACGCTATGGACTACTGGGGTCAA GGAAC TTTGGTTACTGTTTCCTCCGCTTCTACTAAGGG ACCATCTGTTTTCCCATTTGGCTCCATCTTCTAAGTCTA CTTCCGGTGGTACTGCTGCTTTGGGATGTTTGGTTAAA GACTACTTCCCAGAGCCAGTTACTGTTTCTTGGAAC TC CGGTGCTTTGACTTCTGGTGTTACACTTTCCCAGCTG TTTTGCAATCTTCCGGTTTGTACTCTTTGTCCTCCGTTG TACTGTTCCATCCTCTTCCTTGGGTA CTGACTTAC ATCTGTAACGTTAACCACAAGCCATCCAACACTAAGG TTGACAAGAAGGTTGAGCCAAAGTCCTGTGACAAGAC ACATACTTGTCACCATGTCCAGCTCCAGAATTGTTGG GTGGTCCATCCGTTTTCTTGTTCACCAAAGCCAAAG GACACTTTGATGATCTCCAGAACTCCAGAGGTTACAT GTGTTGTTGTTGACGTTTCTCACGAGGACCCAGAGGTT AAGTTCAACTGGTACGTTGACGGTGTTGAAGTTCACA ACGCTAAGACTAAGCCAAGAGAAGAGCAGTACAAC T CCACTTACAGAGTTGTTTCCGTTTTGACTGTTTTGCAC CAGGACTGGTTGAACGGTAAAGAATAACAAGTGTAAGG TTTCCAACAAGGCTTTGCCAGCTCCAATCGAAAAGAC TATCTCCAAGGCTAAGGGTCAACCAAGAGAGCCACAG GTTTACACTTTGCCACCATCCAGAGAAGAGATGACTA AGAACCAGGTTTCTTGACTTGTTTGGTTAAAGGATTC TACCCATCCGACATTGCTGTTGAGTGGGAATCTAACG

		GTCAACCAGAGAACAACACTACAAGACTACTCCACCAGT TTTGGATTCTGATGGTTCCTTCTTCTTGTACTCCAAGTT GACTGTTGACAAGTCCAGATGGCAACAGGGTAACGTT TTCTCCTGTTCCGTTATGCATGAGGCTTTGCACAACCA CTACACTCAAAAGTCCTTGTCTTTGTCCCCTGGTTAA
16	Anti-Her2 Heavy chain (VH + IgG1 constant region) (protein)	EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQ APGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNT AYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGT LVTVSSASTKGPSVFPLAPSSKSTSGGTAAALGCLVKDYFP EPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPS SLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCP APPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED PEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLT VLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREP QVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNG QPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSVMHEALHNHYTQKSLSLSPG
17	Anti-Her2 light chain (VL + Kappa constant region) (DNA)	GACATCCAAATGACTCAATCCCCATCTTCTTTGTCTGC TTCCGTTGGTGACAGAGTTACTATCACTTGTAGAGCTT CCCAGGACGTTAATACTGCTGTTGCTTGGTATCAACAG AAGCCAGGAAAGGCTCCAAAGTTGTTGATCTACTCCG CTTCCTTCTTGTACTCTGGTGTTCATCCAGATTCTCTG GTTCCAGATCCGGTACTGACTTCACTTTGACTATCTCC TCCTTGCAACCAGAAGATTTGCTACTTACTACTGTCA GCAGCACTACACTACTCCACCAACTTTCGGACAGGGT ACTAAGGTTGAGATCAAGAGAACTGTTGCTGCTCCAT CCGTTTTTCATTTTCCCACCATCCGACGAACAGTTGAAG TCTGGTACAGCTTCCGTTGTTTGTGTTGTTGAACAACTT CTACCCAAGAGAGGCTAAGGTTCAAGTGAAGGTTGAC AACGCTTTGCAATCCGGTAACTCCCAAGAATCCGTTA CTGAGCAAGACTCTAAGGACTCCACTTACTCCTTGTCC TCCACTTTGACTTTGTCCAAGGCTGATTACGAGAAGCA CAAGGTTTACGCTTGTGAGGTTACACATCAGGGTTTGT CCTCCCCAGTTACTAAGTCCTTCAACAGAGGAGAGTG TTAA
18	Anti-Her2 light chain (VL + Kappa constant region)	DIQMTQSPSSLSASVGDRTITCRASQDVNTAVAWYQQ KPGKAPKLLIYSASFLYSGVPSRFSGRSGTDFTLTISLQ PEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFP PSDEQLKSGTASVCLLNNFYPREAKVQWKVDNALQSG NSQESVTEQDSKDSYSLSSLTLSKADYEKHKVYACEV THQGLSPVTKSFNRGEC
19	Anti-RSV Heavy chain (VH + IgG1 constant region) (DNA)	CAGGTTACATTGAGAGAATCCGGTCCAGCTTTGGTTA AGCCAACTCAGACTTTGACTTTGACTTGTACTTTCTCC GGTTTCTCCTTGTCTACTTCCGGAATGTCTGTTGGATG GATCAGACAACCACCTGGAAAGGCTTTGGAATGGCTT GCTGACATTTGGTGGGATGACAAGAAGGACTACAACC CATCCTTGAAGTCCAGATTGACTATCTCCAAGGACACT TCCAAGAATCAAGTTGTTTTGAAGGTTACAAACATGG ACCCAGCTGACACTGCTACTTACTACTGTGCTAGATCC

		ATGATCACTAACTGGTACTTCGATGTTTGGGGTGCTGG TACTACTGTTACTGTCTCGAGTGCTTCTACTAAGGGAC CATCCGTTTTTCCATTGGCTCCATCCTCTAAGTCTACTT CCGGTGGAACCGCTGCTTTGGGATGTTTGGTTAAAGA CTACTTCCCAGAGCCAGTTACTGTTTCTTGGAAGTCCG GTGCTTTGACTTCTGGTGTTCACACTTTCCAGCTGTTT TGCAATCTTCCGGTTTGTACTCTTTGTCCTCCGTTGTTA CTGTTCCATCCTCTTCCTTGGGTACTCAGACTTACATC TGTAACGTTAACCACAAGCCATCCAACACTAAGGTTG ACAAGAGAGTTGAGCCAAAGTCCTGTGACAAGACACA TACTTGTCCACCATGTCCAGCTCCAGAATTGTTGGGTG GTCCATCCGTTTTCTTGTTCACCAAAGCCAAAGGAC ACTTTGATGATCTCCAGAACTCCAGAGGTTACATGTGT TGTGTTGACGTTTCTCACGAGGACCCAGAGGTTAAGT TCAACTGGTACGTTGACGGTGTGAAGTTCACAACGC TAAGACTAAGCCAAGAGAAGAGCAGTACAAGTCCACT TACAGAGTTGTTTCCGTTTTGACTGTTTTGCACCAGGA CTGGTTGAACGGTAAAGAATACAAGTGTAAGGTTTCC AACAAGGCTTTGCCAGCTCCAATCGAAAAGACTATCT CCAAGGCTAAGGGTCAACCAAGAGAGCCACAGGTTTA CACTTTGCCACCATCCAGAGAAGAGATGACTAAGAAC CAGGTTTCCTTGACTTGTTTGGTTAAAGGATTCTACCC ATCCGACATTGCTGTTGAGTGGGAATCTAACGGTCAA CCAGAGAACAATAAGACTACTCCACCAGTTTTGG ATTCTGATGGTTCCTTCTTCTGTACTCCAAGTTGACT GTTGACAAGTCCAGATGGCAACAGGGTAACGTTTTCT CCTGTTCCGTTATGCATGAGGCTTTGCACAACCACTAC ACTCAAAAGTCCTTGTCTTTGTCCCCTGGTTAA
20	Anti-RSV Heavy chain (VH + IgG1 constant region) (protein)	QVTLRSGPALVKPTQTLTLCTFSGFSLSTSGMSVGVWIR QPPGKALEWLADIWWDDKKDYNPSLKSRLTISKDTSKN QVVLKVTNMDPADTATYYCARSMITNWFYFDVWGAGTT VTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPE PVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSS LGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPA PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDP EVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTV LHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ VYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQ PENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFSC SVMHEALHNHYTQKSLSLSPG
21	Anti-RSV light chain (VL + Kappa constant region) (DNA)	ATGAGATTCCCATCCATCTTCACTGCTGTTTTGTTCGC TGCTTCTTCTGCTTTGGCTGACATTCAGATGACACAGT CCCCATCTACTTTGTCTGCTTCCGTTGGTGACAGAGTT ACTATCACTTGTAAGTGTGAGTTGTCCGTTGGTTACAT GCACTGGTATCAGCAAAAGCCAGGAAAGGCTCCAAA GTTGTTGATCTACGACACTTCCAAGTTGGCTTCCGGTG TTCCATCTAGATTCTCTGGTTCCGGTTCTGGTACTGAG TTCACTTTGACTATCTCTTCCTTGCAACCAGATGACTT CGCTACTTACTACTGTTTCCAGGGTTCTGGTTACCCAT

		TCACTTTCGGTGGTGGTACTAAGTTGGAGATCAAGAG AACTGTTGCTGCTCCATCCGTTTTTCATTTCCACCAT CCGACGAACAATTGAAGTCCGGTACCGCTTCCGTTGTT TGTTTGTGAACAATTCTACCCACGTGAGGCTAAGGT TCAGTGGAAGGTTGACAACGCTTTGCAATCCGGTAAC TCCCAAGAATCCGTTACTGAGCAGGATTCTAAGGATT CCACTTACTCATTGTCCTCCACTTTGACTTTGTCCAAG GCTGATTACGAGAAGCACAAGGTTTACGCTTGCGAGG TTACACATCAGGGTTTGTCTCCCCAGTTACTAAGTCC TTCAACAGAGGAGAGTGTTAA
22	Anti-RSV light chain (VL + Kappa constant region) (protein)	DIQMTQSPSTLSASVGDRTITCKCQLSVGYMHWYQQK PGKAPKLLIYDTSKLASGVPSRFSGSGSGTEFTLTISLQP DDFATYYCFQGSQYPFTFGGGTKLEIKRTVAAPSVFIFPP SDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGN SQESVTEQDSKDSYSLSTLTLSKADYEKHKVYACEVT HQGLSSPVTKSFNRGEC
23	Pp AOX1 promoter	AACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTG CCATCCGACATCCACAGGTCCATTCTCACACATAAGT GCCAAACGCAACAGGAGGGGATACACTAGCAGCAGA CCGTTGCAAACGCAGGACCTCCACTCCTCTTCTCCTCA ACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATT GGGCTTGATTGGAGCTCGCTCATTCCAATTCCTTCTAT TAGGCTACTAACACCATGACTTTATTAGCCTGTCTATC CTGGCCCCCTGGCGAGGTTTCATGTTTGTATTTCGG AATGCAACAAGCTCCGCATTACACCCGAACATCACTC CAGATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTT CATGTTCCCCAAATGGCCCCAAAACCTGACAGTTTAAAC GCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTC ATCCAAGATGAACTAAGTTTGGTTCGTTGAAATGCTA ACGGCCAGTTGGTCAAAAAGAACTTCCAAAAGTCGG CATACCGTTTGTCTTGTGTTGGTATTGATTGACGAATGC TCAAAAATAATCTCATTAAATGCTTAGCGCAGTCTCTCT ATCGCTTCTGAACCCCGGTGCACCTGTGCCGAAACGC AAATGGGGAAACACCCGCTTTTTTGGATGATTATGCAT TGTCTCCACATTGTATGCTTCCAAGATTCTGGTGGGAA TACTGCTGATAGCCTAACGTTTCATGATCAAAATTTAAC TGTCTAACCCCTACTTGACAGCAATATATAAACAGA AGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATC ATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAAT TGACAAGCTTTTGATTTTAACGACTTTTAACGACAAC TGAGAAGATCAAAAAACAATAATTATTCGAAACG
24	ScCYC TT	ACAGGCCCTTTTCCTTGTGCGATATCATGTAATTAGT TATGTCACGCTTACATTCACGCCCTCCTCCACATCCG CTCTAACCAGAAAAGGAAGGAGTTAGACAACCTGAAGT CTAGGTCCCTATTTATTTTTTTAATAGTTATGTTAGTA TTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTT CTGTACAAACGCGTGTACGCATGTAACATTATACTGA AAACCTTGCTTGAGAAGGTTTGGGACGCTCGAAGGC TTTAATTTGCAAGCTGCCGGCTCTTAAG

25	PpRPL10 promoter	GTTCTTCGCTTGGTCTTGTATCTCCTTACACTGTATCTTCCC ATTTGCGTTTAGGTGGTTATCAAAAACATAAAAGGAAAAAT TTCAGATGTTTATCTCTAAGGTTTTTCTTTTTACAGTATAA CACGTGATGCGTCACGTGGTACTAGATTACGTAAGTTATTT TGGTCCGGTGGGTAAGTGGGTAAGAATAGAAAGCATGAAG GTTTACAAAAACGCAGTCACGAATTATTGCTACTTCGAGCT TGGAACCACCCCAAAGATTATATTGTAAGTACTGACTACCT TCTCGATTTTGCTCCTCCAAGAACCTACGAAAAACATTTCT TGAGCCTTTTCAACCTAGACTACACATCAAGTTATTTAAGG TATGTTCCGTTAACATGTAAGAAAAGGAGAGGATAGATCG TTTATGGGGTACGTCGCCTGATTCAAGCGTGACCATTTCGAA GAATAGGCCTTCGAAAGCTGAATAAAGCAAATGTCAGTTG CGATTGGTATGCTGACAAATTAGCATAAAAAGCAATAGAC TTTCTAACCACCTGTTTTTTTCTTTTACTTTATTTATATTTT GCCACCGTACTAACAAGTTCAGACAAA
26	PpGAPDH promoter	TTTTTGTAGAAATGTCTTGGTGTCTCGTCCAATCAGG TAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCG AACGACCTGCTGGCAACGTAAAATTCTCCGGGGTAAA ACTTAAATGTGGAGTAATGGAACCAGAAACGTCTCTT CCCTTCTCTCTCCTTCCACCGCCCGTTACCGTCCCTAG GAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCC CTTGCAGCAATGCTCTTCCCAGCATTACGTTGCGGGTA AAACGGAGGTCTGTACCCGACCTAGCAGCCCAGGGA TGGAAAAGTCCCGGCCGTCGCTGGCAATAATAGCGGG CGGACGCATGTCATGAGATTATTGGAAACCACCAGAA TCGAATATAAAAGGCGAACACCTTTCCCAATTTTGGTT TCTCCTGACCCAAAGACTTTAAATTTAATTTATTTGTC CCTATTTCAATCAATTGAACAACCTATCAAAACACA
27	PpTEF1 promoter	TTAAGGTTTGGAAACAACACTAAACTACCTTGCGGTAC TACCATTGACACTACACATCCTTAATTCCAATCCTGTC TGGCCTCCTTCACCTTTTAACCATCTTGCCCATTTCAA CTCGTGTCAGATTGCGTATCAAGTGAAAAAAAAAAAAA TTTTAAATCTTTAACCCAATCAGGTAATAACTGTGCGC TCTTTTATCTGCCGCACTGCATGAGGTGTCCCTTAGT GGGAAAGAGTACTGAGCCAACCCTGGAGGACAGCAA GGGAAAAATACCTACAACCTTGCTTCATAATGGTCGTA AAAACAATCCTTGTCGGATATAAGTGTTGTAGACTGT CCCTTATCCTCTGCGATGTTCTTCTCTCAAAGTTTGC GATTTCTCTCTATCAGAATTGCCATCAAGAGACTCAGG ACTAATTTGCGAGTCCACACGCACTCGTACATGATTG GCTGAAATTTCCCTAAAGAATTTCTTTTTACGAAAAAT TTTTTTTTTACACAAGATTTTCAGCAGATATAAAATGG AGAGCAGGACCTCCGCTGTGACTCTTCTTTTTTTCTTT TATTCTCACTACATACATTTTAGTTATTCGCCAAC
28	PpTEF1 TT	ATTGCTTGAAGCTTTAATTTATTTTATTAACATAATAA TAATACAAGCATGATATATTGTATTTTGTTCGTTAAC ATTGATGTTTTCTTCATTTACTGTTATTGTTTGTAACCT TGATCGATTTATCTTTTCTACTTTACTGTAATATGGCTG GCGGGTGAGCCTTGAACCTCCCTGTATTACTTTACCTTG CTATTACTTAATCTATTGACTAGCAGCGACCTCTTCAA

		CCGAAGGGCAAGTACACAGCAAGTTCATGTCTCCGTA AGTGTCATCAACCCTGGAAACAGTGGGCCATGTC
29	PpALG3 TT	ATTTACAATTAGTAATATTAAGGTGGTAAAAACATTC GTAGAATTGAAATGAATTAATATAGTATGACAATGGT TCATGTCTATAAAATCTCCGGCTTCGGTACCTTCTCCCC AATTGAATACATTGTCAAAATGAATGGTTGAACTATT AGGTTTCGCCAGTTTCGTTATTAAGAAAACGTGTTAAAT CAAATTCATATCATCGGTTCCAGTGGGAGGACCAGT TCCATCGCCAAAATCCTGTAAGAATCCATTGTCAGAA CCTGTAAAGTCAGTTTGAGATGAAATTTTCCGGTCTT TGTTGACTTGGAAGCTTCGTTAAGGTTAGGTGAAACA GTTTGATCAACCAGCGGCTCCCGTTTTTCGTCGCTTAGT AG
30	PpTRP1 5' region and ORF	GCGGAAACGGCAGTAAACAATGGAGCTTCATTAGTGGGTG TTATTATGGTCCCTGGCCGGGAACGAACGGTGAAACAAGA GGTTGCGAGGGAAATTCGCAGATGGTGCGGGAAAAGAGA ATTTCAAAGGGCTCAAAATACTTGGATTCCAGACAACTGA GGAAAGAGTGGGACGACTGTCCTCTGGAAGACTGGTTTGA GTACAACGTGAAAGAAATAAACAGCAGTGGTCCATTTTA GTTGGAGTTTTTCGTAATCAAAGTATAGATGAAATCCAGCA AGCTATCCACACTCATGGTTTGGATTTTCGTCCAACACTACATG GGTCTGAGGATTTTGATTTCGTATATACGCAATATCCCAGTT CCTGTGATTACCAGATACACAGATAATGCCGTCGATGGTCT TACCGGAGAAGACCTCGCTATAAATAGGGCCCTGGTGCTA CTGGACAGCGAGCAAGGAGGTGAAGGAAAAACCATCGATT GGGCTCGTGCACAAAAATTTGGAGAACGTAGAGGAAAATA TTTACTAGCCGGAGGTTTGACACCTGATAATGTTGCTCATG CTCGATCTCATACTGGCTGTATTGGTGTGACGTCTCTGGT GGGGTAGAAACAAATGCCTCAAAAGATATGGACAAGATCA CACAATTTATCAGAAACGCTACATAA
31	PpTRP1 3' region	AAGTCAATTAAATACACGCTTGAAAGGACATTACATAGCTT TCGATTTAAGCAGAACCAGAAATGTAGAACCACCTTGTCAA TAGATTGGTCAATCTTAGCAGGAGCGGCTGGGCTAGCAGTT GGAACAGCAGAGGTTGCTGAAGGTGAGAAGGATGGAGTGG ATTGCAAAGTGGTGTGTTAAGTCAATCTCACCAGGGCTG GTTTTGCCAAAAATCAACTTCTCCCAGGCTTCACGGCATT TTGAATGACCTCTTCTGCATACTTCTTGTCTTGCATTACCC AGAGAAAGCAAACCTGGTTCTCAGGTTTTCCATCAGGGATCT TGTAATTTCTGAACCATTCGTTGGTAGCTCTCAACAAGCCC GGCATGTGCTTTTCAACATCCTCGATGTCATTGAGCTTAGG AGCCAATGGGTCGTTGATGTCGATGACGATGACCTTCCAGT CAGTCTCTCCCTCATCCAACAAAGCCATAACACCGAGGACC TTGACTTGCTTGACCTGTCCAGTGTAACCTACGGCTTCACC AATTCGCAAACGTCCAATGGATCATTGTCACCCTTGGCCT TGGTCTCTGGATGAGTGACGTTAGGGTCTTCCCATGTCTGA GGGAAGGCACCGTAGTTGTGAATGTATCCGTGGTGAGGGA AACAGTTACGAACGAAACGAAGTTTTCCCTTCTTTGTGTCC TGAAGAATTGGGTTTCAAGTTTCTCCTCCTTGGAATCTCCAA CTTGGCGTTGGTCCAACGGGGGACTTCAACAACCATGTTGA GAACCTTCTTGGATTCTGTCAGCATAAAGTGGGATGTCGTGG AAAGGAGATACGACTT
32	ScARR3 ORF	ATGTCAGAAGATCAAAAAAGTGAAAATTCCGTACCTTCTA

		AGGTTAATATGGTGAATCGCACCGATATACTGACTACGATC AAGTCATTGTTCATGGCTTGACTTGATGTTGCCATTTACTAT AATTCTCTCCATAATCATTGCAGTAATAATTTCTGTCTATGT GCCTTCTTCCCGTCACTTTTGACGCTGAAGGTCATCCCA ATCTAATGGGAGTGTCCATTCCTTTGACTGTTGGTATGATT GTAATGATGATTCCCCCGATCTGCAAAGTTTCCTGGGAGTC TATTCACAAGTACTTCTACAGGAGCTATATAAGGAAGCAA CTAGCCCTCTCGTTATTTTGAATTGGGTCATCGGTCTTTG TTGATGACAGCATTGGCGTGGATGGCGCTATTCGATTATAA GGAATACCGTCAAGGCATTATTATGATCGGAGTAGCTAGA TGCATTGCCATGGTGCTAATTTGGAATCAGATTGCTGGAGG AGACAATGATCTCTGCGTCGTGCTTGTATTACAACTCGC TTTTACAGATGGTATTATATGCACCATTGCAGATATTTTAC TGTTATGTTATTTCTCATGACCACCTGAATACTTCAAATAG GGTATTATTCGAAGAGGTTGCAAAGTCTGTCTGGAGTTTTTC TCGGCATACCACTGGGAATTGGCATTATCATACGTTTGGGA AGTCTTACCATAGCTGGTAAAAGTAATTATGAAAAATACA TTTTGAGATTTATTTCTCCATGGGCAATGATCGGATTTCAAT ACACTTTATTTGTTATTTTATTAGTAGAGGTTATCAATTTA TCCACGAAATTGGTTCTGCAATATTGTGCTTTGTCCCATTG GTGCTTTACTTCTTTATTGCATGGTTTTTGACCTTCGCATTA ATGAGGTACTTATCAATATCTAGGAGTGATACACAAAGAG AATGTAGCTGTGACCAAGAAGTACTTTTAAAGAGGGTCTG GGGAAGAAAGTCTTGTGAAGCTAGCTTTTCTATTACGATGA CGCAATGTTTCACTATGGCTTCAAATAATTTTGAAGTATCC CTGGCAATTGCTATTTCTTATATGGTAACAATAGCAAGCA AGCAATAGCTGCAACATTTGGGCCGTTGCTAGAAGTTCCAA TTTTATTGATTTTGGCAATAGTCGCGAGAATCCTTAAACCA TATTATATATGGAACAATAGAAATTAA
33	URA6 region	CAAATGCAAGAGGACATTAGAAATGTGTTTGGTAAGAACA TGAAGCCGGAGGCATACAAACGATTACAGATTTGAAGGA GGAAAACAACTGCATCCACCGGAAGTGCCAGCAGCCGTG TATGCCAACCTTGCTCTCAAAGGCATTCCTACGGATCTGAG TGGGAAATATCTGAGATTCACAGACCCACTATTGGAACAG TACCAAACCTAGTTTGGCCGATCCATGATTATGTAATGCAT ATAGTTTTTGTCTGATGCTCACCCTGTTTCGAGTCTGTCTCGTA TCGTCTTACGTATAAGTTCAAGCATGTTTACCAGGTCTGTT AGAACTCCTTTGTGAGGGCAGGACCTATTCGTCTCGGTCC CGTTGTTTCTAAGAGACTGTACAGCCAAGCGCAGAATGGT GGCATTAAACCATAAGAGGATTCTGATCGGACTTGGTCTATT GGCTATTGGAACCACCCTTTACGGGACAACCAACCCTACCA AGACTCCTATTGCATTTGTGGAACCAGCCACGGAAAGAGC GTTTAAGGACGGAGACGTCTCTGTGATTTTTGTTCTCGGAG GTCCAGGAGCTGGAAAAGGTACCCAATGTGCCAACTAGT GAGTAATTACGGATTTGTTACCTGTCAGCTGGAGACTTGT TACGTGCAGAACAGAAGAGGGAGGGGTCTAAGTATGGAGA GATGATTTCCCAGTATATCAGAGATGGACTGATAGTACCTC AAGAGGTCACCATTGCGCTCTTGGAGCAGGCCATGAAGGA AACTTCGAGAAAGGGAAGACACGGTTCTTGATTGATGGA TTCCCTCGTAAGATGGACCAGGCCAAAACCTTTTGAGGAAA AAGTCGCAAAGTCCAAGGTGACACTTTTCTTTGATTGTCCC GAATCAGTGCTCCTTGAGAGATTACTTAAAAGAGGACAGA CAAGCGGAAGAGAGGATGATAATGCGGAGAGTATCAAAA AAAGATTCAAAACATTCGTGGAACTTCGATGCCTGTGGTG

		GACTATTTTCGGGAAGCAAGGACGCGTTTTGAAGGTATCTTG TGACCACCCTGTGGATCAAGTGTATTACAGGTTGTGTCGG TGCTAAAAGAGAAGGGGATCTTTGCCGATAACGAGACGGA GAATAAATAA
34	NatR ORF	ATGGGTACCACTCTTGACGACACGGCTTACCGGTACC GCACCAGTGTCCCGGGGGACGCCGAGGCCATCGAGGC ACTGGATGGGTCTTACCACCGACACCGTCTTCCGC GTCACCGCCACCGGGGACGGCTTACCCTGCGGGAGG TGCCGGTGGACCCGCCCTGACCAAGGTGTTCCCGA CGACGAATCGGACGACGAATCGGACGACGGGGAGGA CGGCGACCCGGACTCCCGGACGTTTCGTCGCGTACGGG GACGACGGCGACCTGGCGGGCTTCGTGGTTCGTCTCGT ACTCCGGCTGGAACCGCCGGCTGACCGTCGAGGACAT CGAGGTCGCCCCGGAGCACCGGGGGCACGGGGTTCGG GCGCGCGTTGATGGGGCTCGCGACGGAGTTCGCCCCG GAGCGGGGCGCCGGGCACCTCTGGCTGGAGGTCACCA ACGTCAACGCACCGGCGATCCACGCGTACCGGCGGAT GGGGTTCACCCTCTGCGGCCTGGACACCGCCCTGTAC GACGGCACCGCCTCGGACGGCGAGCAGGCGCTCTACA TGAGCATGCCCTGCCCTAATCAGTACTG
35	Sequence of the Sh ble ORF (Zeocin resistance marker):	ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCG CGCGCGACGTGCGCGGAGCGGTTCGAGTTCTGGACCGA CCGGCTCGGGTTCTCCCGGACTTCGTGGAGGACGAC TTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTTAT CAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGT ACGCCGAGTGGTCGGAGGTCGTGTCCACGAACCTCCG GGACGCCTCCGGGCGCGCCATGACCGAGATCGGCGAG CAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGG CCGGCAACTGCGTGCACCTTCGTGGCCGAGGAGCAGGA CTGA
36	PpAOX1 TT	TCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAG GCTTCATTTTGATACTTTTTATTTGTAACCTATATAGTATA GGATTTTTTTGTGATTTTGTCTTCTCGTACGAGCTTGCT CCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGTGGTA GGGGTTTGGGAAAATCATTGAGTTTGATGTTTTTCTTGGT ATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGAGAC GTTTCGTTTGTGCA
37	ScTEF1 promoter	GATCCCCACACACCATAGCTTCAAAATGTTTCTACTC CTTTTTTACTCTTCCAGATTTTCTCGGACTCCGCGCATC GCCGTACCACTTCAAAACACCCAAGCACAGCATACTA AATTTCCCCTCTTTCTTCTCTAGGGTGTGTTAATTAC CCGTACTAAAGGTTTGGAAAAGAAAAAGAGACCGC CTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAAT TTTTATCACGTTTCTTTTCTTGAAAATTTTTTTTTTG ATTTTTTCTCTTTCGATGACCTCCCATTTGATATTTAAG TTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCA TTTTTCTTGTTCATTACAACTTTTTTTACTTCTTGCTC ATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTAA ATTACAAA

38	<i>S. cerevisiae</i> invertase gene (ScSUC2) ORF underlined	AGGCCTCGCAACAACCTATAATTGAGTTAAGTGCCTTT CCAAGCTAAAAAGTTTGAGGTTATAGGGGCTTAGCAT CCACACGTCACAATCTCGGGTATCGAGTATAGTATGT AGAATTACGGCAGGAGGTTTCCCAATGAACAAAGGAC AGGGGCACGGTGAGCTGTCGAAGGTATCCATTTTATC ATGTTTCGTTTGTACAAGCACGACATACTAAGACATTT ACCGTATGGGAGTTGTTGTCCTAGCGTAGTTCTCGCTC CCCCAGCAAAGCTCAAAAAAGTACGTCATTTAGAATA GTTTGTGAGCAAATTACCAGTCGGTATGCTACGTTAG AAAGGCCACAGTATTCTTCTACCAAAGGCGTGCCTTT GTTGAACTCGATCCATTATGAGGGCTTCCATTATTCCC CGCATTTTATTACTCTGAACAGGAATAAAAAGAAAA AACCCAGTTTAGGAAATTATCCGGGGGCGAAGAAATA CGCGTAGCGTTAATCGACCCACGTCCAGGGTTTTTCC ATGGAGGTTTCTGGAAAACTGACGAGGAATGTGATT ATAAATCCCTTTATGTGATGTCTAAGACTTTTAAGGTA CGCCCGATGTTTGCCTATTACCATCATAGAGACGTTTC TTTTCGAGGAATGCTTAAACGACTTTGTTTGACAAAAA TGTTGCCTAAGGGCTCTATAGTAAACCATTTGGAAGA AAGATTTGACGACTTTTTTTTTTTTGGATTTCGATCCTAT AATCCTTCCTCCTGAAAAGAAACATATAAATAGATAT GTATTATTCTTCAAAACATTCTCTTGTCTTGTGCTTTT TTTTTACCATATATCTTACTTTTTTTTTTCTCTCAGAGA AACAAGCAAAACAAAAAGCTTTTCTTTTCACTAACGT ATATGATGCTTTTGCAAGCTTTCCTTTTCCTTTTGGCTG <u>GTTTTGCAGCCAAAATATCTGCATCAATGACAAACGA</u> <u>AACTAGCGATAGACCTTTGGTCCACTTCACACCCAAC</u> <u>AAGGGCTGGATGAATGACCCAAATGGGTTGTGGTACG</u> <u>ATGAAAAAGATGCCAAATGGCATCTGTACTTTCAATA</u> <u>CAACCCAAATGACACCGTATGGGGTACGCCATTGTTT</u> <u>TGGGGCCATGCTACTTCCGATGATTTGACTAATTGGGA</u> <u>AGATCAACCCATTGCTATCGCTCCCAAGCGTAACGAT</u> <u>TCAGGTGCTTTCTCTGGCTCCATGGTGGTTGATTACAA</u> <u>CAACACGAGTGGGTTTTTCAATGATACTATTGATCCAA</u> <u>GACAAAGATGCGTTGCGATTTGGACTTATAACACTCC</u> <u>TGAAAGTGAAGAGCAATACATTAGCTATTCTCTTGAT</u> <u>GGTGGTTACACTTTTACTGAATACCAAAAGAACCCTG</u> <u>TTTAGCTGCCAACTCCACTCAATTCAGAGATCCAAAG</u> <u>GTGTTCTGGTATGAACCTTCTCAAAAATGGATTATGAC</u> <u>GGCTGCCAAATCACAAGACTACAAAATTGAAATTTAC</u> <u>TCCTCTGATGACTTGAAGTCCTGGAAGCTAGAATCTGC</u> <u>ATTTGCCAATGAAGGTTTCTTAGGCTACCAATACGAAT</u> <u>GTCCAGGTTTGATTGAAGTCCCAACTGAGCAAGATCC</u> <u>TTCCAAATCTTATTGGGTCATGTTTATTCTATCAACC</u> <u>CAGGTGCACCTGCTGGCGGTTCCCTTCAACCAATATTTT</u> <u>GTTGGATCCTTCAATGGTACTCATTTTGAAGCGTTTGA</u> <u>CAATCAATCTAGAGTGGTAGATTTTGGTAAGGACTAC</u> <u>TATGCCTTGCAAACCTTCTTCAACACTGACCCAACCTA</u> <u>CGGTTCAGCATTAGGTATTGCCTGGGCTTCAAACCTGG</u>
----	--	---

		<u>GAGTACAGTGCCTTTGTCCCACTAACCCATGGAGAT</u> <u>CATCCATGTCTTTGGTCCGCAAGTTTTCTTTGAACACT</u> <u>GAATATCAAGCTAATCCAGAGACTGAATTGATCAATT</u> <u>TGAAAGCCGAACCAATATTGAACATTAGTAATGCTGG</u> <u>TCCCTGGTCTCGTTTTGCTACTAACACAACCTCTAACTA</u> <u>AGGCCAATTCTTACAATGTCGATTTGAGCAACTCGACT</u> <u>GGTACCCTAGAGTTTGAGTTGGTTTACGCTGTTAACAC</u> <u>CACACAAACCATATCCAAATCCGTCTTTGCCGACTTAT</u> <u>CACTTTGGTTCAAGGGTTTAGAAGATCCTGAAGAATA</u> <u>TTTGAGAATGGGTTTTGAAGTCAGTGCTTCTTCCTTCT</u> <u>TTTTGGACCGTGGTAACCTCTAAGGTCAAGTTTGTCAAG</u> <u>GAGAACCCATATTTACAAACAGAATGTCTGTCAACA</u> <u>ACCAACCATTCAAGTCTGAGAACGACCTAAGTTACTA</u> <u>TAAAGTGACGGCCTACTGGATCAAAACATCTTGGA</u> <u>TTGTACTTCAACGATGGAGATGTGGTTTCTACAAATAC</u> <u>CTACTTCATGACCACCGGTAACGCTCTAGGATCTGTGA</u> <u>ACATGACCACTGGTGTGCGATAATTTGTTCTACATTGAC</u> <u>AAGTTCCAAGTAAGGGAAGTAAATAAGAGGTTATAA</u> <u>AACTTATTGTCTTTTTTATTTTTTTTCAAAAGCCATTCTA</u> <u>AAGGGCTTTAGCTAACGAGTGACGAATGTAAACTTT</u> <u>ATGATTTCAAAGAATACCTCCAAACCATTGAAAATGT</u> <u>ATTTTTATTTTTATTTTTCTCCCGACCCAGTTACCTGGA</u> <u>ATTTGTTCTTTATGTACTTTATATAAGTATAATTCTCTT</u> <u>AAAAATTTTTACTACTTTGCAATAGACATCATTTTTTC</u> <u>ACGTAATAAACCCACAATCGTAATGTAGTTGCCTTAC</u> <u>ACTACTAGGATGGACCTTTTTGCCTTTATCTGTTTTGTT</u> <u>ACTGACACAATGAAACCGGGTAAAGTATTAGTTATGT</u> <u>GAAAATTTAAAAGCATTAAAGTAGAAGTATACCATATT</u> <u>GTAAAAAAGCGTTGTCTTCTACGTAAAAGTGT</u> <u>TCTCAAAAAGAAGTAGTGAGGGAAATGGATACCAAG</u> <u>CTATCTGTAAACAGGAGCTAAAAAATCTCAGGGAAAAG</u> <u>CTTCTGGTTTGGGAAACGGTCGAC</u>
39	Sequence of the 5'-Region used for knock out of PpURA5:	ATCGGCCTTTGTTGATGCAAGTTTTACGTGGATCATGG ACTAAGGAGTTTTATTTGGACCAAGTTCATCGTCCTAG ACATTACGGAAAGGGTTCTGCTCCTCTTTTTGGAACT TTTTGGAACCTCTGAGTATGACAGCTTGGTGGATTGTA CCCATGGTATGGCTTCCTGTGAATTTCTATTTTTTCTAC ATTGGATTCACCAATCAAAACAAATTAGTCGCCATGG CTTTTTGGCTTTTGGGTCTATTTGTTTGGACCTTCTTGG AATATGCTTTGCATAGATTTTTGTTCCACTTGGACTAC TATCTTCCAGAGAATCAAATTGCATTTACCATTCAATT CTTATTGCATGGGATACACCACTATTTACCAATGGATA AATACAGATTGGTGATGCCACCTACACTTTTCATTGTA CTTTGCTACCCAATCAAGACGCTCGTCTTTTCTGTTCT ACCATATTACATGGCTTGTCTGGATTTGCAGGTGGAT TCCTGGGCTATATCATGTATGATGTCACTCATTACGTT CTGCATCACTCCAAGCTGCCTCGTTATTTCCAAGAGTT GAAGAAATATCATTTGGAACATCACTACAAGAATTAC GAGTTAGGCTTTGGTGTCACTTCCAATTCTGGGACAA

		AGTCTTTGGGACTTATCTGGGTCCAGACGATGTGTATC AAAAGACAAATTAGAGTATTTATAAAGTTATGTAAGC AAATAGGGGCTAATAGGGAAAGAAAAATTTTGGTTCT TTATCAGAGCTGGCTCGCGCGCAGTGTTTTTCGTGCTC CTTTGTAATAGTCATTTTTGACTACTGTTTCAGATTGAA ATCACATTGAAGATGTCACTCGAGGGGTACCAAAAAA GGTTTTTGGATGCTGCAGTGGCTTCGC
40	Sequence of the 3'-Region used for knock out of PpURA5:	GGTCTTTTCAACAAAGCTCCATTAGTGAGTCAGCTGGC TGAATCTTATGCACAGGCCATCATTAAACAGCAACCTG GAGATAGACGTTGTATTTGGACCAGCTTATAAAGGTA TTCCTTTGGCTGCTATTACCGTGTTGAAGTTGTACGAG CTCGGCGGCAAAAAATACGAAAATGTCGGATATGCGT TCAATAGAAAAGAAAAGAAAAGACCACGGAGAAGGTG GAAGCATCGTTGGAGAAAAGTCTAAAGAATAAAAGAGT ACTGATTATCGATGATGTGATGACTGCAGGTAAGTCT ATCAACGAAGCATTGCTATAATTGGAGCTGAAGGTG GGAGAGTTGAAGGTAGTATTATTGCCCTAGATAGAAT GGAGACTACAGGAGATGACTCAAATACCAGTGCTACC CAGGCTGTTAGTCAGAGATATGGTACCCCTGTCTTGA GTATAGTGACATTGGACCATATTGTGGCCCATTTGGGC GAAACTTTCACAGCAGACGAGAAAATCTCAAATGGAAA CGTATAGAAAAAAGTATTTGCCCAAATAAGTATGAAT CTGCTTCGAATGAATGAATTAATCCAATTATCTTCTCA CCATTATTTCTTCTGTTTCGGAGCTTTGGGCACGGCG GCGGGTGGTGCGGGCTCAGGTTCCCTTTCATAAACAG ATTTAGTACTTGGATGCTTAATAGTGAATGGCGAATG CAAAGGAACAATTTCTGTTTCATCTTTAACCCTTTCCTC GGGGTACACGTTCTGGAATGTACCCGCCCTGTTGCAA CTCAGGTGGACCGGGCAATTCTTGAAGTTTCTGTAACG TTGTTGGATGTTCAACCAGAAATTGTCCTACCAACTGT ATTAGTTTCTTTTGGTCTTATATTGTTTCATCGAGATAC TTCCCACTCTCCTTGATAGCCACTCTCACTCTTCCTGG ATTACCAAAATCTTGAGGATGAGTCTTTTCAGGCTCCA GGATGCAAGGTATATCCAAGTACCTGCAAGCATCTAA TATTGTCTTTGCCAGGGGGTTCTCCACACCATACTCCT TTTGGCGCATGC
41	Sequence of the PpURA5 auxotrophic marker:	TCTAGAGGGACTTATCTGGGTCCAGACGATGTGTATC AAAAGACAAATTAGAGTATTTATAAAGTTATGTAAGC AAATAGGGGCTAATAGGGAAAGAAAAATTTTGGTTCT TTATCAGAGCTGGCTCGCGCGCAGTGTTTTTCGTGCTC CTTTGTAATAGTCATTTTTGACTACTGTTTCAGATTGAA ATCACATTGAAGATGTCACTGGAGGGGTACCAAAAAA GGTTTTTGGATGCTGCAGTGGCTTCGCAGGCCTTGAAG TTTGGAAGTTTACCTTGAAAAGTGGAAGACAGTCTC CATACTTCTTTAACATGGGTCTTTTCAACAAAGCTCCA TTAGTGAGTCAGCTGGCTGAATCTTATGCTCAGGCCAT CATTAAACAGCAACCTGGAGATAGACGTTGTATTTGGA CCAGCTTATAAAGGTATTCCTTTGGCTGCTATTACCGT GTTGAAGTTGTACGAGCTGGGCGGCAAAAAATACGAA

		AATGTCGGATATGCGTTCAATAGAAAAGAAAAGAAAAG ACCACGGAGAAGGTGGAAGCATCGTTGGAGAAAGTCT AAAGAATAAAAGAGTACTGATTATCGATGATGTGATG ACTGCAGGTACTGCTATCAACGAAGCATTTGCTATAA TTGGAGCTGAAGGTGGGAGAGTTGAAGGTTGTATTAT TGCCCTAGATAGAATGGAGACTACAGGAGATGACTCA AATACCAGTGCTACCCAGGCTGTTAGTCAGAGATATG GTACCCCTGTCTTGAGTATAGTGACATTGGACCATATT GTGGCCCATTTGGGCGAACTTTCACAGCAGACGAGA AATCTCAAATGGAAACGTATAGAAAAAAGTATTTGCC CAAATAAGTATGAATCTGCTTCGAATGAATGAATTAA TCCAATTATCTTCTCACCATTATTTTCTTCTGTTTCGGA GCTTTGGGCACGGCGGCGGATCC
42	Sequence of the part of the Ec lacZ gene that was used to construct the PpURA5 blaster (recyclable auxotrophic marker)	CCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTG GCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAG GTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCC GGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTA GTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGC ACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAA CCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCC CGCATCTGACCACCAGCGAAATGGATTTTTGCATCGA GCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCA GGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAAC AACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGC ACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACC CGCATTGACCCTAACGCCTGGGTGCAACGCTGGAAGG CGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCA GTGCACGGCAGATACACTTGCTGATGCGGTGCTGATT ACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCT TATTTATCAGCCGGAACCTACCGGATTGATGGTAG TGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCG AGCGATACACCGCATCCGGCGCGGATTGGCCTGAACT GCCAG
43	Sequence of the 5'-Region used for knock out of PpOCH1:	AAAACCTTTTTCTTATTCAAACACAAGGCATTGCTTC AACACGTGTGCGTATCCTTAACACAGATACTCCATACT TCTAATAATGTGATAGACGAATACAAAGATGTTCACT CTGTGTTGTGTCTACAAGCATTTCTTATTCTGATTGGG GATATTCTAGTTACAGCACTAAACAACCTGGCGATACA AACTTAAATTAAATAATCCGAATCTAGAAAATGAACT TTTGGATGGTCCGCCTGTTGGTTGGATAAATCAATACC GATTAAATGGATTCTATTCCAATGAGAGAGTAATCCA AGACACTCTGATGTCAATAATCATTTGCTTGCAACAAC AAACCCGTCATCTAATCAAAGGGTTTGATGAGGCTTA CCTTCAATTGCAGATAAACTCATTGCTGTCCACTGCTG TATTATGTGAGAATATGGGTGATGAATCTGGTCTTCTC CACTCAGCTAACATGGCTGTTTGGGCAAAGGTGGTAC AATTATACGGAGATCAGGCAATAGTGAAATTGTTGAA TATGGCTACTGGACGATGCTTCAAGGATGTACGTCTA GTAGGAGCCGTGGGAAGATTGCTGGCAGAACCAGTTG

	GCACGTCGCAACAATCCCCAAGAAATGAAATAAGTGA AAACGTAACGTCAAAGACAGCAATGGAGTCAATATTG ATAACACCACTGGCAGAGCGGTTCGTACGTCGTTTTG GAGCCGATATGAGGCTCAGCGTGCTAACAGCACGATT GACAAGAAGACTCTCGAGTGACAGTAGGTTGAGTAAA GTATTCGCTTAGATTCCCAACCTTCGTTTTATTCTTTCG TAGACAAAGAAGCTGCATGCGAACATAGGGACAACCTT TTATAAATCCAATTGTCAAACCAACGTAAAACCCTCT GGCACCATTTTCAACATATATTTGTGAAGCAGTACGC AATATCGATAAATACTCACCGTTGTTTGTAACAGCCCC AACTTGCATACGCCTTCTAATGACCTCAAATGGATAA GCCGCAGCTTGTGCTAACATAACCAGCAGCACCGCCCCG CGGTCAGCTGCGCCACACATATAAAGGCAATCTACG ATCATGGGAGGAATTAGTTTTGACCGTCAGGTCTTCA AGAGTTTTGAACTCTTCTTCTTGAACGTGTGTAACCTTT TAAATGACGGGATCTAAATACGTCATGGATGAGATCA TGTGTGTAAAACTGACTCCAGCATATGGAATCATT CAAAGATTGTAGGAGCGAACCCACGATAAAAGTTTCC CAACCTTGCCAAAGTGTCTAATGCTGTGACTTGAAATC TGGGTTCTCGTTGAAGACCCTGCGTACTATGCCCAAA AACTTTCCTCCACGAGCCCTATTAACCTTCTCTATGAGT TTCAAATGCCAAACGGACACGGATTAGGTCCAATGGG TAAGTGAAAAACACAGAGCAAACCCAGCTAATGAG CCGGCCAGTAACCGTCTTGGAGCTGTTTCATAAGAGT CATTAGGGATCAATAACGTTCTAATCTGTTTCATAACAT ACAAATTTTATGGCTGCATAGGGAAAAATTCTCAACA GGGTAGCCGAATGACCCTGATATAGACCTGCGACACC ATCATACCCATAGATCTGCCTGACAGCCTTAAAGAGC CCGCTAAAAGACCCGGAAAACCGAGAGAACTCTGGAT TAGCAGTCTGAAAAAGAATCTTCACTCTGTCTAGTGG AGCAATTAATGTCTTAGCGGCACTTCCTGCTACTCCGC CAGCTACTCCTGAATAGATCACATACTGCAAAGACTG CTTGTCGATGACCTTGGGGTTATTTAGCTTCAAGGGCA ATTTTTGGGACATTTTGGACACAGGAGACTCAGAAAC AGACACAGAGCGTTCTGAGTCCTGGTGCTCCTGACGT AGGCCTAGAACAGGAATTATTGGCTTTATTTGTTTGTC CATTTTCATAGGCTTGGGGTAATAGATAGATGACAGAG AAATAGAGAAGACCTAATATTTTTTGTTCATGGCAAA TCGCGGGTTCGCGGTCGGGTCACACACGGAGAAGTAA TGAGAAGAGCTGGTAATCTGGGGTAAAAGGGTTCAAA AGAAGGTCGCCTGGTAGGGATGCAATACAAGGTTGTC TTGGAGTTTACATTGACCAGATGATTTGGCTTTTTCTC TGTTCAATTACATTTTTTCAGCGAGAATCGGATTGACG GAGAAATGGCGGGGTGTGGGGTGGATAGATGGCAGA AATGCTCGCAATCACCGCGAAAGAAAGACTTTATGGA ATAGAACTACTGGGTGGTGTAAGGATTACATAGCTAG TCCAATGGAGTCCGTTGGAAAGGTAAGAAGAAGCTAA AACCGGCTAAGTAAGTAAAGGAAGAATGATCAGACTTT GATTTGATGAGGTCTGAAAATACTCTGCTGCTTTTTCA
--	---

		GTTGCTTTTTCCCTGCAACCTATCATTTCCTTTTCATA AGCCTGCCTTTTCTGTTTTCACTTATATGAGTTCGCC GAGACTTCCCCAAATTCTCTCCTGGAACATTCTCTATC GCTCTCCTTCCAAGTTGCGCCCCCTGGCACTGCCTAGT AATATTACCACGCGACTTATATTCACTTCCACAATTC CAGTGTTCGTAGCAAATATCATCAGCCATGGCGAAGG CAGATGGCAGTTTGCTCTACTATAATCCTCACAATCCA CCCAGAAGGTATTACTTCTACATGGCTATATTCGCCGT TTCTGTCATTTGCGTTTTGTACGGACCCTCACAACAAT TATCATCTCCAAAAATAGACTATGATCCATTGACGCTC CGATCACTTGATTTGAAGACTTTGGAAGCTCCTTCACA GTTGAGTCCAGGCACCGTAGAAGATAATCTTCG
44	Sequence of the 3'-Region used for knock out of PpOCH1:	AAAGCTAGAGTAAAATAGATATAGCGAGATTAGAGA ATGAATACCTTCTTCTAAGCGATCGTCCGTCATCATAG AATATCATGGACTGTATAGTTTTTTTTTTGTACATATA ATGATTAAACGGTCATCCAACATCTCGTTGACAGATCT CTCAGTACGCGAAATCCCTGACTATCAAAGCAAGAAC CGATGAAGAAAAAAACAACAGTAACCCAAACACCAC AACAAACACTTTATCTTCTCCCCCCCCAACACCAATCAT CAAAGAGATGTCGGAACCAAACACCAAGAAGCAAAA ACTAACCCCATATAAAAACATCCTGGTAGATAATGCT GGTAACCCGCTCTCCTTCCATATTCTGGGCTACTTCAC GAAGTCTGACCGGTCTCAGTTGATCAACATGATCCTC GAAATGGGTGGCAAGATCGTTCCAGACCTGCCTCCTC TGGTAGATGGAGTGTTGTTTTTGACAGGGGATTACAA GTCTATTGATGAAGATACCCTAAAGCAACTGGGGGAC GTTCCAATATACAGAGACTCCTTCATCTACCAGTGTTT TGTGCACAAGACATCTCTTCCCATTGACACTTTCCGAA TTGACAAGAACGTCGACTTGGCTCAAGATTTGATCAA TAGGGCCCTTCAAGAGTCTGTGGATCATGTCACTTCTG CCAGCACAGCTGCAGCTGCTGCTGTTGTTGTCGCTACC AACGGCCTGTCTTCTAAACCAGACGCTCGTACTAGCA AAATACAGTTCACTCCCGAAGAAGATCGTTTTATTCTT GACTTTGTTAGGAGAAATCCTAAACGAAGAAACACAC ATCAACTGTACACTGAGCTCGCTCAGCACATGAAAAA CCATACGAATCATTCTATCCGCCACAGATTTTCGTCGTA ATCTTTCCGCTCAACTTGATTGGGTTTATGATATCGAT CCATTGACCAACCAACCTCGAAAAGATGAAAACGGGA ACTACATCAAGGTACAAGGCCTTCCA
45	<i>K. lactis</i> UDP- GlcNAc transporter gene (KIMNN2-2) ORF underlined	AAACGTAACGCCTGGCACTCTATTTTCTCAAACCTTCTG GGACGGAAGAGCTAAATATTGTGTTGCTTGAACAAAC CCAAAAAAACAAAAAAATGAACAAACTAAACTACA CCTAAATAAACCGTGTGTAAACGTAGTACCATATTA CTAGAAAAGATCACAAGTGTATCACACATGTGCATCT CATATTACATCTTTTATCCAATCCATTCTCTATCCCG TCTGTTCTGTGAGATTCTTTTCCATAAAAAGAAGAA GACCCCGAATCTCACCGGTACAATGCAAACTGCTGA AAAAAAAAGAAAGTTCACTGGATACGGGAACAGTGC CAGTAGGCTTCACCACATGGACAAAACAATTGACGAT

		AAAATAAGCAGGTGAGCTTCTTTTTCAAGTCACGATC CCTTTATGTCTCAGAAACAATATATACAAGCTAAACC CTTTTGAACCAGTTCTCTCTTCATAGTTATGTTACAT AAATTGCGGGAACAAGACTCCGCTGGCTGTCAGGTAC ACGTTGTAACGTTTTTCGTCCGCCCAATTATTAGCACAA CATTGGCAAAAAGAAAAACTGCTCGTTTTCTCTACAG GTAAATTACAATTTTTTTTCAGTAATTTTCGCTGAAAAA TTTAAAGGGCAGGAAAAAAGACGATCTCGACTTTGC ATAGATGCAAGAACTGTGGTCAAACTTGAAATAGTA ATTTTGCTGTGCGTGAACATAATAATATATATATAT ATATATATATATTTGTGTATTTTGTATATGTAATTGTGC ACGTCTTGGCTATTGGATATAAGATTTTCGCGGGTTGA TGACATAGAGCGTGTACTACTGTAATAGTTGTATATTC AAAAGCTGCTGCGTGGAGAAAGACTAAAATAGATAA AAAGCACACATTTTGACTTCGGTACCGTCAACTTAGTG GGACAGTCTTTTATATTTGGTGTAAAGCTCATTCTGGT ACTATTCGAAACAGAACAGTGTTTTCTGTATTACCGTC CAATCGTTTGTCA <u>ATGAGTTTTGTATTGATTTTGT</u> <u>CGTT</u> <u>AGTGTTCCGAGGATGTTGTTCCAATGTGATTAGTTTCG</u> <u>AGCACATGGTGCAAGGCAGCAATATAAATTTGGGAAA</u> <u>TATTGTTACATTCAC</u> <u>TCAATTCGTGTCTGTGACGCTAA</u> <u>TTCAGTTGCCCAATGCTTTGGACTTCTCTCACTTTCCGT</u> <u>TTAGGTTGCGACCTAGACACATTCCTCTTAAGATCCAT</u> <u>ATGTTAGCTGTGTTTTTGTCTTTACCAGTTCAGTCGCC</u> <u>AATAACAGTGTTTAAATTTGACATTTCCGTTCCGAT</u> <u>TCATATTATCATTAGATTTTCAGGTACCACTTTGACGA</u> <u>TGATAATAGGTTGGGCTGTTTGTAATAAGAGGTA</u> <u>CTC</u> <u>CAAAC</u> <u>TTCA</u> <u>GGTGCAATCTGCCATCATTATGACGCTTG</u> <u>GTGCGATTGT</u> <u>CGCATCATTATACCGTGACAAAGAATT</u> <u>TCAATGGACAGTTTAAAGTTGAATACGGATT</u> <u>CAGTGG</u> <u>GTATGACCCAAAAATCTATGTTTGGTATCTTTGTTGTG</u> <u>CTAGTGGCCACTGCCTTGATGTCATTGTTGTCGTTGCT</u> <u>CAACGAATGGACGTATAACAAGTACGGGAAACATTGG</u> <u>AAAGAACTTTGTTCTATTCGCATTTCTTGGCTCTACC</u> <u>GTTGTTTATGTTGGGGTACACAAGGCTCAGAGACGAA</u> <u>TTCA</u> <u>GAGACCTCTTAATTTCCCTCAGACTCAATGGATAT</u> <u>TCCTATTGTTAAATTACCAATTGCTACGAACTTTTCA</u> <u>TGCTAATAGCAAATAACGTGACCCAGTTCATTGTATC</u> <u>AAAGGTGTTAACATGCTAGCTAGTAACACGGATGCTT</u> <u>TGACACTTTCTGTCGTGCTTCTAGTGCGTAAATTTGTT</u> <u>AGTCTTTTACTCAGTGTCTACATCTACAAGAACGTCTT</u> <u>ATCCGTGACTGCATACCTAGGGACCATCACCGTGTTCC</u> <u>TGGGAGCTGGTTTGTATT</u> <u>CATATGGTTCGGTCAAACT</u> <u>GCACTGCCTCGCTGA</u> <u>AACAATCCACGTCTGTATGATA</u> <u>CTCGTTTCAGAAATTTTTTGATTTTCTGCCGGATATGGT</u> <u>TTCTCATCTTTACAATCGCATTCTTAATTATACCAGAA</u> <u>CGTAATTCAATGATCCAGTGACTCGTAACTCTTATAT</u> <u>GTCAATTTAAGC</u>
46	Sequence of the	GGCCGAGCGGGCCTAGATTTTCACTACAAATTTCAAA

	5'-Region used for knock out of PpBMT2:	ACTACGCGGATTTATTGTCTCAGAGAGCAATTTGGCAT TTCTGAGCGTAGCAGGAGGCTTCATAAGATTGTATAG GACCGTACCAACAAATTGCCGAGGCACAACACGGTAT GCTGTGCACTTATGTGGCTACTTCCCTACAACGGAATG AAACCTTCCTCTTTCCGCTTAAACGAGAAAGTGTGTGCG CAATTGAATGCAGGTGCCTGTGCGCCTTGGTGTATTGT TTTTGAGGGCCCAATTTATCAGGCGCCTTTTTTCTTGG TTGTTTTCCCTTAGCCTCAAGCAAGGTTGGTCTATTTC ATCTCCGCTTCTATAACCGTGCCTGATACTGTTGGATGA GAACACGACTCAACTTCCTGCTGCTCTGTATTGCCAGT GTTTTGTCTGTGATTTGGATCGGAGTCCTCCTTACTTG GAATGATAATAATCTTGGCGGAATCTCCCTAAACGGA GGCAAGGATTCTGCCTATGATGATCTGCTATCATTGGG AAGCTTCAACGACATGGAGGTCGACTCCTATGTCACC AACATCTACGACAATGCTCCAGTGCTAGGATGTACGG ATTTGTCTTATCATGGATTGTTGAAAGTCACCCCAAAG CATGACTTAGCTTGCATTGTTGGAGTTCATAAGAGCTCA GATTTTGGACATTGACGTTTACTCCGCCATAAAAGACT TAGAAGATAAAGCCTTGACTGTAAAACAAAAGGTTGA AAAACACTGGTTTACGTTTTATGGTAGTTCAGTCTTTC TGCCCGAACACGATGTGCATTACCTGGTTAGACGAGT CATCTTTTCGGCTGAAGGAAAGGCGAACTCTCCAGTA ACATC
47	Sequence of the 3'-Region used for knock out of PpBMT2:	CCATATGATGGGTGTTTGCTCACTCGTATGGATCAAAA TTCCATGGTTTCTTCTGTACAACCTTGACACTTATTTGG ACTTTTCTAACGGTTTTTCTGGTGATTTGAGAAGTCCT TATTTTGGTGTTTCGCAGCTTATCCGTGATTGAACCATC AGAAATACTGCAGCTCGTTATCTAGTTTCAGAATGTGT TGTAAGAATAACAATCAATTCTGAGTCTAGTTTGGGTGG GTCTTGGCGACGGGACCGTTATATGCATCTATGCAGT GTTAAGGTACATAGAATGAAAATGTAGGGGTTAATCG AAAGCATCGTTAATTTCAGTAGAACGTAGTTCTATTCC CTACCCAAATAATTTGCCAAGAATGCTTCGTATCCACA TACGCAGTGGACGTAGCAAATTTCACTTTGGACTGTG ACCTCAAGTCGTTATCTTCTACTTGGACATTGATGGTC ATTACGTAATCCACAAAGAATTGGATAGCCTCTCGTTT TATCTAGTGCACAGCCTAATAGCACTTAAGTAAGAGC AATGGACAAATTTGCATAGACATTGAGCTAGATACGT AACTCAGATCTTGTTCACTCATGGTGTACTCGAAGTAC TGCTGGAACCGTTACCTCTTATCATTTTCGCTACTGGCT CGTGAAACTACTGGATGAAAAAAAAAAAAAGAGCTGA AAGCGAGATCATCCCATTTTGTTCATCATACAAATTCAC GCTTGCAGTTTTGCTTCGTTAACAAGACAAGATGTCTT TATCAAAGACCCGTTTTTTCTTCTTGAAGAATACTTCC CTGTTGAGCACATGCAAACCATATTTATCTCAGATTTT ACTCAACTTGGGTGCTTCCAAGAGAAGTAAAATTTCTT CCCCTGCATCAACTTCCAAGAAACCCGTAGACCAGT TTCTCTTCAGCCAAAAGAAGTTGCTCGCCGATCACCGC

		GGTAACAGAGGAGTCAGAAGGTTTCACACCCTTCCAT CCCGATTTCAAAGTCAAAGTGCTGCGTTGAACCAAGG TTTTCAAGTTGCCAAAGCCCAGTCTGCAAAAAGTAGTT CCAAATGGCCTATTAATTCCCATAAAAAGTGTGGCTAC GTATGTATCGGTACCTCCATTCTGGTATTTGCTATTGT TGTCGTTGGTGGGTTGACTAGACTGACCGAATCCGGT CTTCCATAACGGAGTGGAAACCTATCACTGGTTCGGT TCCCCCACTGACTGAGGAAGACTGGAAGTTGGAATTT GAAAAATACAAACAAAGCCCTGAGTTTCAGGAACTAA ATTCTCACATAACATTGGAAGAGTTCAAGTTTATATTT TCCATGGAATGGGGACATAGATTGTTGGGAAGGGTCA TCGGCCTGTCGTTTGTCTTCCCACGTTTTACTTCATTG CCCGTCGAAAGTGTTCCAAAGATGTTGCATTGAAACT GCTTGCAATATGCTCTATGATAGGATTCCAAGGTTTCA TCGGCTGGTGGATGGTGTATTCCGGATTGGACAAACA GCAATTGGCTGAACGTAACCTCAAACCAACTGTGTCT CCATATCGCTTAACTACCCATCTTGGAAGTGCATTTGT TATTTACTGTTACATGATTTACACAGGGCTTCAAGTTT TGAAGAACTATAAGATCATGAAACAGCCTGAAGCGTA TGTTCAAATTTTCAAGCAAATTGCGTCTCCAAAATTGA AACTTTCAAGAGACTCTCTTCAGTTCTATTAGGCCTG GTG
48	DNA encodes MmSLC35A3 UDP-GlcNAc transporter	ATGTCTGCCAACCTAAAATATCTTTCCTTGGAATTTT GGTGTTCAGACTACCAGTCTGGTTCTAACGATGCGGT ATTCTAGGACTTTAAAAGAGGAGGGGCCTCGTTATCT GTCTTCTACAGCAGTGGTTGTGGCTGAATTTTGAAGA TAATGGCCTGCATCTTTTGTAGTCTACAAAGACAGTAAG TGTAAGTGTGAGAGCACTGAATAGAGTACTGCATGATG AAATTCTTAATAAGCCCATGGAAACCCTGAAGCTCGC TATCCCGTCAGGGATATATACTCTTCAGAACAACTTAC TCTATGTGGCACTGTCAAACCTAGATGCAGCCACTTAC CAGGTTACATATCAGTTGAAAATACTTACAACAGCAT TATTTTCTGTGTCTATGCTTGGTAAAAAATTAGGTGTG TACCAGTGGCTCTCCCTAGTAATTCTGATGGCAGGAGT TGCTTTTGTACAGTGGCCTTCAGATTCTCAAGAGCTGA ACTCTAAGGACCTTTCAACAGGCTCACAGTTTGTAGG CCTCATGGCAGTTCTCACAGCCTGTTTTTCAAGTGGCT TTGCTGGAGTTTATTTTGAGAAAATCTTAAAAGAAAC AAAACAGTCAGTATGGATAAGGAACATTCAACTTGGT TTCTTTGGAAGTATATTGGGATTAATGGGTGTATACGT TTATGATGGAGAATTGGTCTCAAAGAATGGATTTTTTC AGGGATATAATCAACTGACGTGGATAGTTGTTGCTCT GCAGGCACTTGGAGGCCTTGTAATAGCTGCTGTCATC AAATATGCAGATAACATTTTAAAAGGATTTGCGACCT CCTTATCCATAATATTGTCAACAATAATATCTTATTTT TGGTTGCAAGATTTTGTGCCAACCAGTGTCTTTTTCCT TGGAGCCATCCTTGTAATAGCAGCTACTTTCTTGTATG GTTACGATCCCAAACCTGCAGGAAATCCCCTAAAGC ATAG

49	Sequence of the 5'-Region used for knock out of PpMNN4L1:	GATCTGGCCATTGTGAACTTGACACTAAAGACAAAA CTCTTAGAGTTTCCAATCACTTAGGAGACGATGTTTCC TACAACGAGTACGATCCCTCATTGATCATGAGCAATTT GTATGTGAAAAAAGTCATCGACCTTGACACCTTGGAT AAAAGGGCTGGAGGAGGTGGAACCACCTGTGCAGGC GGTCTGAAAGTGTTCAAGTACGGATCTACTACCAAT ATACATCTGGTAACCTGAACGGCGTCAGGTTAGTATA CTGGAACGAAGGAAAAGTTGCAAAGCTCCAAATTTGTG GTTTCGATCCTCTAATTACTCTCAAAGCTTGGAGGAA ACAGCAACGCCGAATCAATTGACAACAATGGTGTGGG TTTTGCCTCAGCTGGAGACTCAGGCGCATGGATTCTTT CCAAGCTACAAGATGTTAGGGAGTACCAGTCATTAC TGAAAAGCTAGGTGAAGCTACGATGAGCATTTCGAT TTCCACGGTCTTAAACAGGAGACTTCTACTACAGGGC TTGGGGTAGTTGGTATGATTCACTTACGACGGTGAG TTCAAACAGTTTGGTTTGTTCCTCCAATGACATCTAT TCTACAAAGACTTCAACGAGTGACCAATGTAGAATGG TGTGTAGCGGGTTGCGAAGATGGGGATGTGGACACTG AAGGAGAACACGAATTGAGTGATTTGGAACAACTGCA TATGCATAGTGATTCCGACTAGTCAGGCAAGAGAGAG CCCTCAAATTTACCTCTCTGCCCCCTCCTCACTCCTTTT GTACGCATAATTGCAGTATAAAGAACTTGCTGCCAGC CAGTAATCTTATTTTCATACGCAGTTCTATATAGCACAT AATCTTGCTTGTATGTATGAAATTTACCGCGTTTTAGT TGAAATTGTTTATGTTGTGTGCCTTGCATGAAATCTCT CGTTAGCCCTATCCTTACATTTAACTGGTCTCAAACC TCTACCAATTCCATTGCTGTACAACAATATGAGGCGG CATTACTGTAGGGTTGGAAAAAAATTGTCATTCCAGC TAGAGATCACACGACTTCATCACGCTTATTGCTCCTCA TTGCTAAATCATTTACTCTTGACTTCGACCCAGAAAAG TTCGCC
50	Sequence of the 3'-Region used for knock out of PpMNN4L1:	GCATGTCAAACCTTGAACACAACGACTAGATAGTTGTT TTTTCTATATAAAACGAAACGTTATCATCTTTAATAAT CATTGAGGTTTACCCTTATAGTTCCGTATTTTCGTTTCC AACTTAGTAATCTTTTGGAAATATCATCAAAGCTGGT GCCAATCTTCTTGTGTTGAAGTTTCAAACCTGCTCCACCA AGCTACTTAGAGACTGTTCTAGGTCTGAAGCAACTTC GAACACAGAGACAGCTGCCGCCGATTGTTCTTTTTTGT GTTTTTCTTCTGGAAGAGGGGCATCATCTTGTATGTCC AATGCCCGTATCCTTTCTGAGTTGTCCGACACATTGTC CTTCGAAGAGTTTCCTGACATTGGGCTTCTTCTATCCG TGTATTAATTTTGGGTTAAGTTCCTCGTTTGCATAGCA GTGGATACCTCGATTTTTTTGGCTCCTATTACCTGAC ATAATATTCTACTATAATCCAACCTTGACGCGTCATCT ATGATAACTAGGCTCTCCTTTGTTCAAAGGGGACGTCT TCATAATCCACTGGCACGAAGTAAGTCTGCAACGAGG CGGCTTTTGCAACAGAACGATAGTGTCGTTTCGTA GGACTATGCTAAACAAAAGGATCTGTCAAACATTTCA ACCGTGTTTCAAGGCACTCTTTACGAATTATCGACCAA

		GACCTTCCTAGACGAACATTTCAACATATCCAGGCTA CTGCTTCAAGGTGGTGCAAATGATAAAGGTATAGATA TTAGATGTGTTTGGGACCTAAACAGTTCTTGCCTGAA GATTCCCTTGAGCAACAGGCTTCAATAGCCAAGTTAG AGAAGCAGTACCAAATCGGTAACAAAAGGGGGAAGC ATATAAAACCTTTACTATTGCGACAAAATCCATCCTTG AAAGTAAAGCTGTTTGTTCATGTAAAGCATAACGAAA CGAAGGAGGTAGATCCTAAGATGGTTAGAGAACTTAA CGGGACATACTCCAGCTGCATCCCATATTACGATCGCT GGAAGACTTTTTTCATGTACGTATCGCCACCAACCTT TCAAAGCAAGCTAGGTATGATTTTGACAGTTCTCACA ATCCATTGGTTTTTCATGCAACTTGAAAAACCCAACTC AACTTCATGGGGATCCATACAATGTAAATCATTACG AGAGGGCGAGGTTGAAAAGTTTCCATTGCAATCACGT CGCATCATGGCTACTGAAAGGCCTTAAC
51	Sequence of the 5'-Region used for knock out of PpPNO1 and PpMNN4:	TCATTCTATATGTTCAAGAAAAGGGTAGTGAAAGGAA AGAAAAGGCATATAGGCGAGGGAGAGTTAGCTAGCA TACAAGATAATGAAGGATCAATAGCGGTAGTTAAAGT GCACAAGAAAAGAGCACCTGTTGAGGCTGATGATAAA GCTCCAATTACATTGCCACAGAGAAACACAGTAACAG AAATAGGAGGGGATGCACCACGAGAAGAGCATTTCAG TGAACAACCTTTGCCAAATTCATAACCCCAAGCGCTAA TAAGCCAATGTCAAAGTCGGCTACTAACATTAATAGT ACAACAACCTATCGATTTTCAACCAGATGTTTGCAAGG ACTACAAACAGACAGGTTACTGCGGATATGGTGACAC TTGTAAGTTTTTGCACCTGAGGGATGATTTCAAACAGG GATGGAAATTAGATAGGGAGTGGGAAAATGTCCAAA AGAAGAAGCATAATACTCTCAAAGGGGTTAAGGAGAT CCAAATGTTTAATGAAGATGAGCTCAAAGATATCCCG TTAAATGCATTATATGCAAAGGAGATTACAAATCAC CCGTGAAAACCTTCTTGCAATCATTATTTTTCGAACAA TGTTTCCTGCAACGGTCAAGAAGAAAACCAAATTGTA TTATATGTGGCAGAGACACTTTAGGAGTTGCTTTACCA GCAAAGAAGTTGTCCCAATTTCTGGCTAAGATACATA ATAATGAAAGTAATAAAGTTTAGTAATTGCATTGCGTT GACTATTGATTGCATTGATGTCGTGTGATACTTTCACC GAAAAAAAACACGAAGCGCAATAGGAGCGGTTGCAT ATTAGTCCCCAAAGCTATTTAATTGTGCCTGAAACTGT TTTTTAAGCTCATCAAGCATAATTGTATGCATTGCGAC GTAACCAACGTTTAGGCGCAGTTTAATCATAGCCAC TGCTAAGCC
52	Sequence of the 3'-Region used for knock out of PpPNO1 and PpMNN4:	CGGAGGAATGCAAATAATAATCTCCTTAATTACCCAC TGATAAGCTCAAGAGACGCGGTTTGAAAACGATATAA TGAATCATTTGGATTTTATAATAAACCTGACAGTTTT TCCACTGTATTGTTTTAACACTCATTGGAAGCTGTATT GATTCTAAGAAGCTAGAAATCAATACGGCCATACAAA AGATGACATTGAATAAGCACCGGCTTTTTTGATTAGC ATATACCTTAAAGCATGCATTCATGGCTACATAGTTGT TAAAGGGCTTCTTCCATTATCAGTATAATGAATTACAT

		AATCATGCACTTATATTTGCCCATCTCTGTTCTCTCACT CTTGCCCTGGGTATATTCTATGAAATTGCGTATAGCGTG TCTCCAGTTGAACCCCAAGCTTGGCGAGTTTGAAGAG AATGCTAACCTTGCGTATTCCTTGCTTCAGGAAACATT CAAGGAGAAACAGGTCAAGAAGCCAAACATTTTGATC CTTCCCGAGTTAGCATTGACTGGCTACAATTTTCAAAG CCAGCAGCGGATAGAGCCTTTTTTGGAGGAAACAACC AAGGGAGCTAGTACCCAATGGGCTCAAAAAGTATCCA AGACGTGGGATTGCTTTACTTTAATAGGATACCCAGA AAAAAGTTTAGAGAGCCCTCCCGTATTTACAACAGT GCGGTACTTGTATCGCCTCAGGGAAAAGTAATGAACA ACTACAGAAAGTCCTTCTTGTATGAAGCTGATGAACA TTGGGGATGTTTCGGAATCTTCTGATGGGTTTCAAACAG TAGATTTATTAATTGAAGGAAAGACTGTAAAGACATC ATTTGGAATTTGCATGGATTTGAATCCTTATAAATTTG AAGCTCCATTCACAGACTTCGAGTTCAGTGGCCATTGC TTGAAAACCGGTACAAGACTCATTTTGTGCCCAATGG CCTGGTTGTCCCCTCTATCGCCTTCATTAAAAAGGAT CTTAGTGATATAGAGAAAAGCAGACTTCAAAAAGTTCT ACCTTGAAAAAATAGATACCCCGGAATTTGACGTAA TTACGAATTGAAAAAAGATGAAGTATTGCCACCCGT ATGAATGAAACGTTGGAAACAATTGACTTTGAGCCTT CAAAACCGGACTACTCTAATATAAATTATTGGTAACT AAGGTTTTTCCCTTTCTGACTCATGTCTATAAACGAG ATGTGCTCAAAGAGAATGCAGTTGCAGTCTTATGCAA CCGAGTTGGCATTGAGAGTGATGTCTTGTACGGAGGA TCAACCACGATTCTAAACTTCAATGGTAAGTTAGCATC GACACAAGAGGAGCTGGAGTTGTACGGGCAGACTAAT AGTCTCAACCCAGTGTGGAAGTATTGGGGGCCCTTG GCATGGGTCAACAGGGAATTCTAGTACGAGACATTGA ATTAACATAATATACAATATACAATAAACACAAATAA AGAATACAAGCCTGACAAAAATTCACAAATTATTGCC TAGACTTGTCTGTTATCAGCAGCGACCTTTTCCAATGC TCAATTTACGATATGCCTTTTCTAGCTCTGCTTTAAG CTTCTCATTGGAATTGGCTAACTCGTTGACTGCTTGGT CAGTGATGAGTTTCTCCAAGGTCCATTTCTCGATGTTG TTGTTTTCGTTTTCTTTAATCTCTTGATATAATCAACA GCCTTCTTTAATATCTGAGCCTTGTTTCGAGTCCCCTGT TGGCAACAGAGCGGCCAGTTCCTTTATTCCGTGGTTTA TATTTTCTCTTCTACGCCTTTCTACTTCTTTGTGATTCT CTTTACGCATCTTATGCCATTCTTCAGAACCAGTGGCT GGCTTAACCGAATAGCCAGAGCCTGAAGAAGCCGCAC TAGAAGAAGCAGTGGCATTGTTGACTATGG
53	DNA encodes human GnTI catalytic domain (NA) Codon-	TCAGTCAGTGCTCTTGATGGTGACCCAGCAAGTTTGAC CAGAGAAGTGATTAGATTGGCCCAAGACGCAGAGGTG GAGTTGGAGAGACAACGTGGACTGCTGCAGCAAATCG GAGATGCATTGTCTAGTCAAAGAGGTAGGGTGCCTAC CGCAGCTCCTCCAGCACAGCCTAGAGTGCATGTGACC CCTGCACCAGCTGTGATTCTTATCTTGGTCATCGCCTG

	optimized	TGACAGATCTACTGTTAGAAGATGTCTGGACAAGCTG TTGCATTACAGACCATCTGCTGAGTTGTTCCCTATCAT CGTTAGTCAAGACTGTGGTCACGAGGAGACTGCCAA GCCATCGCCTCCTACGGATCTGCTGTCACTCACATCAG ACAGCCTGACCTGTCATCTATTGCTGTGCCACCAGACC ACAGAAAGTTCCAAGGTTACTACAAGATCGCTAGACA CTACAGATGGGCATTGGGTCAAGTCTTCAGACAGTTT AGATTCCCTGCTGCTGTGGTGGTGGAGGATGACTTGG AGGTGGCTCCTGACTTCTTTGAGTACTTTAGAGCAACC TATCCATTGCTGAAGGCAGACCCATCCCTGTGGTGTGT CTCTGCCTGGAATGACAACGGTAAGGAGCAAATGGTG GACGCTTCTAGGCCTGAGCTGTTGTACAGAACCGACT TCTTTCCTGGTCTGGGATGGTTGCTGTTGGCTGAGTTG TGGGCTGAGTTGGAGCCTAAGTGGCCAAAGGCATTCT GGGACGACTGGATGAGAAGACCTGAGCAAAGACAGG GTAGAGCCTGTATCAGACCTGAGATCTCAAGAACCAT GACCTTTGGTAGAAAGGGAGTGTCTCACGGTCAATTC TTTGACCAACACTTGAAGTTTATCAAGCTGAACCAGC AATTTGTGCACTTCACCCAACCTGGACCTGTCTTACTTG CAGAGAGAGGCCTATGACAGAGATTTCCCTAGCTAGAG TCTACGGAGCTCCTCAACTGCAAGTGGAGAAAGTGAG GACCAATGACAGAAAGGAGTTGGGAGAGGTGAGAGT GCAGTACACTGGTAGGGACTCCTTTAAGGCTTTCGCTA AGGCTCTGGGTGTCATGGATGACCTTAAGTCTGGAGT TCCTAGAGCTGGTTACAGAGGTATTGTCACCTTTCAAT TCAGAGGTAGAAGAGTCCACTTGGCTCCTCCACCTAC TTGGGAGGGTTATGATCCTTCTTGGAATTAG
54	DNA encodes Pp SEC12 (10) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest.	ATGCCCAGAAAAATATTTAACTACTTCATTTTGACTGT ATTCATGGCAATTCTTGCTATTGTTTTACAATGGTCTA TAGAGAATGGACATGGGCGCGCC
55	Sequence of the PpSEC4 promoter:	GAAGTAAAGTTGGCGAAACTTTGGGAACCTTTGGTTA AAACTTTGTAAATTTTGTGCTACCCATTAGGCAGAAT CTGCATCTTGGGAGGGGGATGTGGTGGCGTTCTGAGA TGTACGCGAAGAATGAAGAGCCAGTGGTAACAACAG GCCTAGAGAGATACGGGCATAATGGGTATAACCTACA AGTTAAGAATGTAGCAGCCCTGGAAACCAGATTGAAA CGAAAAACGAAATCATTTAACTGTAGGATGTTTTGG CTCATTGTCTGGAAGGCTGGCTGTTTATTGCCCTGTTC TTTGCATGGGAATAAGCTATTATATCCCTCACATAATC CCAGAAAATAGATTGAAGCAACGCGAAATCCTTACGT ATCGAAGTAGCCTTCTTACACATTACGTTGTACGGAT

		AAGAAAAC TACTCAAACGAACAATC
56	Sequence of the PpOCH1 terminator:	AATAGATATAGCGAGATTAGAGAATGAATACCTTCTT CTAAGCGATCGTCCGTCATCATAGAATATCATGGACT GTATAGTTTTTTTTTTGTACATATAATGATTAAACGGT CATCCAACATCTCGTTGACAGATCTCTCAGTACGCGA AATCCCTGACTATCAAAGCAAGAACCGATGAAGAAAA AAACAACAGTAACCCAAACACCACAACAAACACTTTA TCTTCTCCCCCACAACCAATCATCAAAGAGATGTC GGAACACAAACACCAAGAAGCAAAAACCTAACCCCAT ATAAAAACATCCTGGTAGATAATGCTGGTAACCCGCT CTCCTTCCATATTCTGGGCTACTTCACGAAGTCTGACC GGTCTCAGTTGATCAACATGATCCTCGAAATGG
57	DNA encodes Mm ManI catalytic domain (FB)	GAGCCCGCTGACGCCACCATCCGTGAGAAGAGGGCAA AGATCAAAGAGATGATGACCCATGCTTGGAATAATTA TAAACGCTATGCGTGGGGCTTGAACGAACTGAAACCT ATATCAAAAGAAGGCCATTCAAGCAGTTTGTGTTGGCA ACATCAAAGGAGCTACAATAGTAGATGCCCTGGATAC CCTTTTCATTATGGGCATGAAGACTGAATTTCAAGAA GCTAAATCGTGGATTAAAAAATATTTAGATTTTAATGT GAATGCTGAAGTTTCTGTTTTTGAAGTCAACATACGCT TCGTCGGTGGACTGCTGTCAGCCTACTATTTGTCCGGA GAGGAGATATTTCGAAAGAAAGCAGTGGAACCTGGGG TAAAATTGCTACCTGCATTTTCACTCCCTCTGGAATA CCTTGGGCATTGCTGAATATGAAAAGTGGGATCGGGC GGAAC TGGCCCTGGGCCTCTGGAGGCAGCAGTATCCT GGCCGAATTTGGAAC TCTGCATTTAGAGTTTATGCACT TGTCCTCACTTATCAGGAGACCCAGTCTTTGCCGAAAA GGTTATGAAAATTCGAACAGTGTTGAACAAACTGGAC AAACCAGAAGGCCTTTATCCTA ACTATCTGAACCCCA GTAGTGGACAGTGGGGTCAACATCATGTGTGCGTTGG AGGACTTGGAGACAGCTTTTATGAATATTTGCTTAAG GCGTGGTTAATGTCTGACAAGACAGATCTCGAAGCCA AGAAGATGTATTTTGATGCTGTTTCAGGCCATCGAGAC TCACTTGATCCGCAAGTCAAGTGGGGGACTAACGTAC ATCGCAGAGTGGAAGGGGGGCCTCCTGGAACACAAG ATGGGCCACCTGACGTGCTTTGCAGGAGGCATGTTTG CACTTGGGGCAGATGGAGCTCCGGAAGCCCGGGCCCA ACACTACCTTGAAC TCGGAGCTGAAATTGCCCGCACT TGTCATGAATCTTATAATCGTACATATGTGAAGTTGGG ACCGGAAGCGTTTCGATTTGATGGCGGTGTGGAAGCT ATTGCCACGAGGCAAAATGAAAAGTATTACATCTTAC GGCCCGAGGTCATCGAGACATACATGTACATGTGGCG ACTGACTCACGACCCCAAGTACAGGACCTGGGCCTGG GAAGCCGTGGAGGCTCTAGAAAGTCACTGCAGAGTGA ACGGAGGCTACTCAGGCTTACGGGATGTTTACATTGC CCGTGAGAGTTATGACGATGTCCAGCAAAGTTTCTTCC TGGCAGAGACACTGAAGTATTTGTACTTGATATTTTCC GATGATGACCTTCTTCCACTAGAACACTGGATCTTCAA CACCGAGGCTCATCCTTTCCCTATACTCCGTGAACAGA

		AGAAGGAAATTGATGGCAAAGAGAAATGA
58	DNA encodes ScSEC12 (8) The last 9 nucleotides are the linker containing the AscI restriction site used for fusion to proteins of interest	ATGAACACTATCCACATAATAAAATTACCGCTTA ACGCCAACTACACCTCAATGAAACAAAAAATCTCTAA ATTTTTCACCAACTTCATCCTTATTGTGCTGCTTTCTTA CATTTTACAGTTCTCCTATAAGCACAATTTGCATTCCA TGCTTTTCAATTACGCGAAGGACAATTTTCTAACGAAA AGAGACACCATCTCTTCGCCCTACGTAGTTGATGAAG ACTTACATCAAACAACCTTTGTTTGGCAACCACGGTAC AAAAACATCTGTACCTAGCGTAGATTCCATAAAAGTG CATGGCGTGGGGCGCGCC
59	Sequence of the 5'-region that was used to knock into the PpADE1 locus:	GAGTCGGCCAAGAGATGATAACTGTTACTAAGCTTCT CCGTAATTAGTGGTATTTTGTAACCTTTACCAATAATC GTTTATGAATACGGATATTTTTCGACCTTATCCAGTGC CAAATCACGTAACCTTAATCATGGTTTAAATACTCCACT TGAACGATTTCATTATTCAGAAAAAAGTCAGGTTGGCA GAAACACTTGGGCGCTTTGAAGAGTATAAGAGTATTA AGCATTAACATCTGAACTTTCACCGCCCCAATATACT ACTCTAGGAACTCGAAAAATTCCTTTCCATGTGTCAT CGCTTCCAACACACTTTGCTGTATCCTTCCAAGTATGT CCATTGTGAACACTGATCTGGACGGAATCCTACCTTTA ATCGCCAAAGGAAAGGTTAGAGACATTTATGCAGTCG ATGAGAACAACTTGCTGTTTCGTCGCAACTGACCGTAT CTCCGCTTACGATGTGATTATGACAAACGGTATTCCTG ATAAGGGAAAGATTTTGACTCAGCTCTCAGTTTCTGG TTTGATTTTTTGGCACCTACATAAAGAATCATTTGGT TGCTTCTAATGACAAGGAAGTCTTTGCTTTACTACCAT CAAACTGTCTGAAGAAAAATACAAATCTCAATTAGA GGGACGATCCTTGATAGTAAAAAAGCACAGACTGATA CCTTTGGAAGCCATTGTCAGAGGTTACATCACTGGAA GTGCATGGAAAGAGTACAAGAACTCAAAAACGTCCA TGGAGTCAAGGTTGAAAACGAGAACCTTCAAGAGAGC GACGCCTTTCCAACCTCCGATTTTCACACCTTCAACGAA AGCTGAACAGGGTGAACACGATGAAAACATCTCTATT GAACAAGCTGCTGAGATTGTAGGTAAAGACATTTGTG AGAAGGTCGCTGTCAAGGCGGTCGAGTTGTATTCTGC TGCAAAAAACCTCGCCCTTTTGAAGGGGATCATTATT GCTGATACGAAATTCGAATTTGGACTGGACGAAAACA ATGAATTGGTACTAGTAGATGAAGTTTTAACTCCAGA TTCTTCTAGATTTTGGAAATCAAAAGACTTACCAAGTGG GTAAATCGCAAGAGAGTTACGATAAGCAGTTTCTCAG AGATTGGTTGACGGCCAACGGATTGAATGGCAAAGAG GGCGTAGCCATGGATGCAGAAATTGCTATCAAGAGTA AAGAAAAGTATATTGAAGCTTATGAAGCAATTACTGG CAAGAAATGGGCTTGA
60	Sequence of the 3'-region that was used to	ATGATTAGTACCCTCCTCGCCTTTTTCAGACATCTGAA ATTTCCCTTATTCTTCCAATTCCATATAAAATCCTATTT AGGTAATTAGTAAACAATGATCATAAAGTGAAATCAT

	knock into the PpADE1 locus:	TCAAGTAACCATTCCGTTTATCGTTGATTTAAAATCAA TAACGAATGAATGTCGGTCTGAGTAGTCAATTTGTTGC CTTGGAGCTCATTGGCAGGGGGTCTTTTGGCTCAGTAT GGAAGGTTGAAAGGAAAACAGATGGAAAGTGGTTCG TCAGAAAAGAGGTATCCTACATGAAGATGAATGCCAA AGAGATATCTCAAGTGATAGCTGAGTTCAGAATTCTT AGTGAGTTAAGCCATCCCAACATTGTGAAGTACCTTC ATCACGAACATATTTCTGAGAATAAAACTGTCAATTT ATACATGGAATACTGTGATGGTGGAGATCTCTCCAAG CTGATTCGAACACATAGAAGGAACAAAGAGTACATTT CAGAAGAAAAAATATGGAGTATTTTTACGCAGGTTTT ATTAGCATTGTATCGTTGTCATTATGGAACGTATTTC CGGCTTCAAAGGAGTTTGAATCGCTCAATAAAGGTAA TAGACGAACCCAGAATCCTTCGTGGGTAGACTCGACA AGAGTTATTATTACAGGGATATAAAACCCGACAACA TCTTTCTGATGAACAATTCAAACCTTGTCAAACGTGGGA GATTTTGGATTAGCAAAAATTCTGGACCAAGAAAACG ATTTTGCCAAAACATACGTCGGTACGCCGTATTACATG TCTCCTGAAGTGCTGTTGGACCAACCCTACTCACCATT ATGTGATATATGGTCTCTTGGGTGCGTCATGTATGAGC TATGTGCATTGAGGCCTCCTT
61	DNA encodes ScGAL10	ATGACAGCTCAGTTACAAAGTGAAAGTACTTCTAAAA TTGTTTTGGTTACAGGTGGTGGTGGTGGTGGTGGTGGT CACACTGTGGTAGAGCTAATTGAGAATGGATATGACT GTGTTGTTGCTGATAACCTGTGCAATTCAACTTATGAT TCTGTAGCCAGGTTAGAGGTCTTGACCAAGCATCACA TTCCCTTCTATGAGGTTGATTTGTGTGACCGAAAAGGT CTGGAAAAGGTTTTCAAAGAATATAAAATTGATTCGG TAATTCACTTTGCTGGTTTAAAGGCTGTAGGTGAATCT ACACAAATCCCGCTGAGATACTATCACAATAACATTT TGGGAACGTGCTGTTTTATTAGAGTTAATGCAACAATAC AACGTTTCCAAATTTGTTTTTTCATCTTCTGCTACTGTC TATGGTGATGCTACGAGATTCCCAAATATGATTCCTAT CCCAGAAGAATGTCCCTTAGGGCCTACTAATCCGTAT GGTCATACGAAATACGCCATTGAGAATATCTTGAATG ATCTTTACAATAGCGACAAAAAAAGTTGGAAGTTTGC TATCTTGCGTTATTTTAACCCAATTGGCGCACATCCCT CTGGATTAATCGGAGAAGATCCGCTAGGTATACCAAA CAATTTGTTGCCATATATGGCTCAAGTAGCTGTTGGTA GGCGCGAGAAGCTTTACATCTTCGGAGACGATTATGA TTCCAGAGATGGTACCCCGATCAGGGATTATATCCAC GTAGTTGATCTAGCAAAAGGTCATATTGCAGCCCTGC AATACCTAGAGGCCTACAATGAAAATGAAGGTTTGTG TCGTGAGTGGAACCTTGGGTTCGGTAAAGGTTCTACA GTTTTTGAAGTTTATCATGCATTCTGCAAAGCTTCTGG TATTGATCTTCCATACAAAGTTACGGGCAGAAGAGCA GGTGATGTTTTGAACCTTGACGGCTAAACCAGATAGGG CCAAACGCGAAGTGAATGGCAGACCGAGTTGCAGGT TGAAGACTCCTGCAAGGATTTATGGAAATGGACTACT

		GAGAATCCTTTTGGTTACCAGTTAAGGGGTGTCGAGG CCAGATTTTCCGCTGAAGATATGCGTTATGACGCAAG ATTTGTGACTATTGGTGCCGGCACCAGATTTCAAGCCA CGTTTGCCAATTTGGGCGCCAGCATTGTTGACCTGAAA GTGAACGGACAATCAGTTGTTCTTGGCTATGAAAATG AGGAAGGGTATTTGAATCCTGATAGTGCTTATATAGG CGCCACGATCGGCAGGTATGCTAATCGTATTTTCAAG GGTAAGTTTAGTTTATGCAACAAAGACTATCAGTTAA CCGTAAATAACGGCGTTAATGCGAATCATAGTAGTAT CGGTTCTTTCCACAGAAAAAGATTTTTGGGACCCATCA TTCAAAATCCTTCAAAGGATGTTTTTACCGCCGAGTAC ATGCTGATAGATAATGAGAAGGACACCGAATTTCCAG GTGATCTATTGGTAACCATAACAGTATACTGTGAACGTT GCCCAAAAAAGTTTGGAAATGGTATATAAAGGTAAAT TGACTGCTGGTGAAGCGACGCCAATAAATTTAACAAA TCATAGTTATTTCAATCTGAACAAGCCATATGGAGAC ACTATTGAGGGTACGGAGATTATGGTGCGTTCAAAAA AATCTGTTGATGTCGACAAAAACATGATTCCTACGGG TAATATCGTCGATAGAGAAATTGCTACCTTTAACTCTA CAAAGCCAACGGTCTTAGGCCCAAAAAATCCCCAGTT TGATTGTTGTTTTGTGGTGGATGAAAATGCTAAGCCAA GTCAAATCAATACTCTAAACAATGAATTGACGCTTATT GTCAAGGCTTTTCATCCCGATTCCAATATTACATTAGA AGTTTTAAGTACAGAGCCAACCTTATCAATTTTATACCG GTGATTTCTTGTCTGCTGGTTACGAAGCAAGACAAGG TTTTGCAATTGAGCCTGGTAGATACATTGATGCTATCA ATCAAGAGAACTGGAAAGATTGTGTAACCTTGAAAAA CGGTGAAACTTACGGGTCCAAGATTGTCTACAGATTTT CCTGA
62	Sequence of the PpPMA1 terminator:	TAAGCTTCACGATTTGTGTTCCAGTTTATCCCCCTTT ATATACCGTTAACCCTTTCCCTGTTGAGCTGACTGTTG TTGTATTACCGCAATTTTCCAAGTTTGCCATGCTTTTC GTGTTATTTGACCGATGTCTTTTTTCCCAAATCAAAC ATATTGTTACCATTTAAACCAAGTTATCTTTTGTATT AAGAGTCTAAGTTTGTTCAGGCTTCATGTGAGAGT GATAACCATCCAGACTATGATTCTTGTTTTTTATTGGG TTTGTGTTGTGTGATACATCTGAGTTGTGATTCGTAAAG TATGTCAGTCTATCTAGATTTTAAATAGTTAATTGGTA ATCAATGACTTGTGTTTAACTTTTAAATTGTGGGT CGTATCCACGCGTTTAGTATAGCTGTTTCATGGCTGTTA GAGGAGGGCGATGTTTATATACAGAGGACAAGAATGA GGAGGCGGCGTGTATTTTAAAATGGAGACGCGACTC CTGTACACCTTATCGGTTGG
63	hGalT codon optimized (XB)	GGTAGAGATTTGTCTAGATTGCCACAGTTGGTTGGTGT TTCCACTCCATTGCAAGGAGGTTCTAACTCTGCTGCTG CTATTGGTCAATCTTCCGGTGAGTTGAGAACTGGTGG AGCTAGACCACCTCCACCATTGGGAGCTTCCTCTCAA CCAAGACCAGGTGGTGATTCTTCTCCAGTTGTTGACTC TGGTCCAGGTCCAGCTTCTAACTTGACTTCCGTTCCAG

		TTCCACACACTACTGCTTTGTCCTTGCCAGCTTGTCCA GAAGAATCCCCATTGTTGGTTGGTCCAATGTTGATCGA GTTCAACATGCCAGTTGACTTGGAGTTGGTTGCTAAGC AGAACCCAAACGTTAAGATGGGTGGTAGATACGCTCC AAGAGACTGTGTTTCCCCACACAAAGTTGCTATCATC ATCCCATTTCAGAAACAGACAGGAGCACTTGAAGTACT GGTGTACTACTTGCACCCAGTTTTTGCAAAGACAGCA GTTGGACTACGGTATCTACGTTATCAACCAGGCTGGT GACACTATTTTCAACAGAGCTAAGTTGTTGAATGTTGG TTTCCAGGAGGCTTTGAAGGATTACGACTACACTTGT TCGTTTTCTCCGACGTTGACTTGATTCCAATGAACGAC CACAACGCTTACAGATGTTTCTCCCAGCCAAGACACA TTTCTGTTGCTATGGACAAGTTCGGTTTCTCCTTGCCA TACGTTCAATACTTCGGTGGTGTTCGCTTTGTCCAA GCAGCAGTTCTTGACTATCAACGGTTTCCCAAACAATT ACTGGGGATGGGGTGGTGAAGATGACGACATCTTTAA CAGATTGGTTTTTCAGAGGAATGTCCATCTCTAGACCA AACGCTGTTGTTGGTAGATGTAGAATGATCAGACACT CCAGAGACAAGAAGAACGAGCCAAACCCACAAAGAT TCGACAGAATCGCTCACACTAAGGAACTATGTTGTC CGACGGATTGAACTCCTTGACTTACCAGGTTTTGGACG TTCAGAGATACCCATTGTACACTCAGATCACTGTTGAC ATCGGTACTCCATCCTAG
64	DNA encodes ScMnt1 (Kre2) (33)	ATGGCCCTCTTCTCAGTAAGAGACTGTTGAGATTTAC CGTCATTGCAGGTGCGGTATTGTTCTCCTCCTAACAT TGAATTCCAACAGTAGAACTCAGCAATATATTCCGAG TTCCATCTCCGCTGCATTTGATTTTACCTCAGGATCTA TATCCCCTGAACAACAAGTCATCGGGCGCGCC
65	DNA encodes DmUGT	ATGAATAGCATAACATGAACGCCAATACGCTGAAGT ACATCAGCCTGCTGACGCTGACCCTGCAGAATGCCAT CCTGGGCCTCAGCATGCGCTACGCCCGCACCCGGCCA GCGGACATCTTCTCAGCTCCACGGCCGTACTCATGG CAGAGTTCGCCAAACTGATCACGTGCCTGTTCTTGGTC TTCAACGAGGAGGGCAAGGATGCCCAGAAGTTTGTAC GCTCGCTGCACAAGACCATCATTGCGAATCCCATGGA CACGCTGAAGGTGTGCGTCCCCTCGCTGGTCTATATCG TTCAAAACAATCTGCTGTACGTCTCTGCCTCCCATTTG GATGCGGCCACCTACCAGGTGACGTACCAGCTGAAGA TTCTCACCACGGCCATGTTGCGGGTTGTCATTCTGCGC CGCAAGCTGCTGAACACGCAGTGGGGTGCCTGCTGC TCCTGGTGATGGGCATCGTCCTGGTGCAGTTGGCCCA AACGGAGGGTCCGACGAGTGGCTCAGCCGGTGGTGCC GCAGCTGCAGCCACGGCCGCCTCCTCTGGCGGTGCTC CCGAGCAGAACAGGATGCTCGGACTGTGGGCCGCACT GGGCGCCTGCTTCTCTCCGGATTGCGGGGCATCTACT TTGAGAAGATCCTCAAGGGTGCCGAGATCTCCGTGTG GATGCGGAATGTGCAGTTGAGTCTGCTCAGCATTCCT TCGGCCTGCTCACCTGTTTCGTAAACGACGGCAGTAGG ATCTTCGACCAGGGATTCTTCAAGGGCTACGATCTGTT

		TGTCTGGTACCTGGTCCTGCTGCAGGCCGGCGGTGGA TTGATCGTTGCCGTGGTGGTCAAGTACGCGGATAACA TTCTCAAGGGCTTCGCCACCTCGCTGGCCATCATCATC TCGTGCGTGGCCTCCATATACATCTTCGACTTCAATCT CACGCTGCAGTTCAGCTTCGGAGCTGGCCTGGTCATC GCCTCCATATTTCTCTACGGCTACGATCCGGCCAGGTC GGCGCCGAAGCCAACATATGCATGGTCCTGGCGGCGAT GAGGAGAAGCTGCTGCCGCGCGTCTAG
66	Sequence of the PpOCH1 promoter:	TGGACACAGGAGACTCAGAAACAGACACAGAGCGTT CTGAGTCCTGGTGCTCCTGACGTAGGCCTAGAACAGG AATTATTGGCTTTATTTGTTTGTCCATTTTCATAGGCTTG GGTAATAGATAGATGACAGAGAAATAGAGAAGACC TAATATTTTTTGTTCATGGCAAATCGCGGGTTCGCGGT CGGGTCACACACGGAGAAGTAATGAGAAGAGCTGGT AATCTGGGGTAAAAGGGTTCAAAAGAAGGTCGCCTGG TAGGGATGCAATACAAGGTTGTCTTGGAGTTTACATT GACCAGATGATTTGGCTTTTTCTCTGTTCAATTCACAT TTTTCAGCGAGAATCGGATTGACGGAGAAATGGCGGG GTGTGGGGTGGATAGATGGCAGAAATGCTCGCAATCA CCGCGAAAGAAAGACTTTATGGAATAGAACTACTGGG TGGTGTAAGGATTACATAGCTAGTCCAATGGAGTCCG TTGGAAAGGTAAGAAGAAGCTAAAACCGGCTAAGTA ACTAGGGAAGAATGATCAGACTTTGATTTGATGAGGT CTGAAAATACTCTGCTGCTTTTTTCAGTTGCTTTTTCCCT GCAACCTATCATTTTTCCTTTTCATAAGCCTGCCTTTTCT GTTTTCACTTATATGAGTTCCGCCGAGACTTCCCCAAA TTCTCTCCTGGAACATTCTCTATCGCTCTCCTTCCAAG TTGCGCCCCCTGGCACTGCCTAGTAATATTACCACGCG ACTTATATTCAGTTCCACAATTTCCAGTGTTTCGTAGCA AATATCATCAGCC
67	Sequence of the PpALG12 terminator:	AATATATACCTCATTTGTTCAATTTGGTGTAAGAGTG TGGCGGATAGACTTCTTGTAATCAGGAAAGCTACAA TTCCAATTGCTGCAAAAAATACCAATGCCCATAAACC AGTATGAGCGGTGCCTTCGACGGATTGCTTACTTTCCG ACCCTTTGTCGTTTGATTCTTCTGCCTTTGGTGAGTCA GTTTGTTTCGACTTTATATCTGACTCATCAACTTCCTTT ACGGTTGCGTTTTTAATCATAATTTAGCCGTTGGCTT ATTATCCCTTGAGTTGGTAGGAGTTTGTATGATGCTG
68	Sequence of the 5'-Region used for knock out of PpHIS1:	TAACTGGCCCTTTGACGTTTCTGACAATAGTTCTAGAG GAGTCGTCCAAAACTCAACTCTGACTTGGGTGACAC CACCACGGGATCCGGTTCTTCCGAGGACCTTGATGAC CTTGGCTAATGTAAGTGGAGTTTTAGTATCCATTTTAA GATGTGTGTTTCTGTAGGTTCTGGGTGGAATAAATT TTAGACACCAGAAGAGAGGAGTGAAGTGGTTTGCGTG GGTTTAGACTGTGTAAGGCACTACTCTGTCAAGTTTT AGATAGGGGTTACCCGCTCCGATGCATGGGAAGCGAT TAGCCCGGCTGTTGCCCGTTTGGTTTTTGAAGGGTAAT TTTCAATATCTCTGTTTGAGTCATCAATTTTCATATTCA

		AAGATTCAAAAACAAAATCTGGTCCAAGGAGCGCATT TAGGATTATGGAGTTGGCGAATCACTTGAACGATAGA CTATTATTTGC
69	Sequence of the 3'-Region used for knock out of PpHIS1:	GTGACATTCTTGTCTTTGAGATCAGTAATTGTAGAGCA TAGATAGAATAATATTCAAGACCAACGGCTTCTCTTC GGAAGCTCCAAGTAGCTTATAGTGATGAGTACCGGCA TATATTTATAGGCTTAAAATTTGAGGGGTTCACTATAT TCGTTTAGTGGGAAGAGTTCCTTTCACTCTTGTTATCT ATATTGTCAGCGTGGACTGTTTATAACTGTACCAACTT AGTTTCTTTCAACTCCAGGTAAAGAGACATAAATGTCC TTTGATGCTGACAATAATCAGTGGAATTCAAGGAAGG ACAATCCCGACCTCAATCTGTTCATTAATGAAGAGTTC GAATCGTCCTTAAATCAAGCGCTAGACTCAATTGTCA ATGAGAACCCTTTCTTTGACCAAGAACTATAAATAG ATCGAATGACAAAGTTGGAAATGAGTCCATTAGCTTA CATGATATTGAGCAGGCAGACCAAAATAAACCGTCCT TTGAGAGCGATATTGATGGTTCGGCGCCGTTGATAAG AGACGACAAATTGCCAAAGAAACAAAGCTGGGGGCT GAGCAATTTTTTTTCAAGAAGAAATAGCATATGTTTAC CACTACATGAAAATGATTCAAGTGTTGTTAAGACCGA AAGATCTATTGCAGTGGAACACCCCATCTTCAATAC TGCTTCAATGGAATCTCCAATGCCAAGTACAATGCATT TACCTTTTTCCAGTCATCCTATACGAGCAATTCAAAT TTTTTTTCAATTTATACTTTACTTTAGTGGCTCTCTCTC AAGCGATACCGCAACTTCGCATTGGATATCTTCTTCG TATGTCGTCCCACTTTTGTGTGTACTCATAGTGACCAT GTCAAAAGAGGCGATGGATGATATTCAACGCCGAAGA AGGGATAGAGAACAGAACAAATGAACCATATGAGGTT CTGTCCAGCCCATCACCAGTTTTGTCCAAAACTTAAA ATGTGGTCACTTGGTTCGATTGCATAAGGGAATGAGA GTGCCCCGAGATATGGTTCTTGTCCAGTCAAGCGAAT CCACCGGAGAGTCATTTATCAAGACAGATCAGCTGGA TGGTGAGACTGATTGGAAGCTTCGGATTGTTTCTCCAG TTACACAATCGTTACCAATGACTGAACTTCAAAATGTC GCCATCACTGCAAGCGCACCCCTCAAAATCAATTCACT CCTTTCTTGAAGATTGACCTACAATGGGCAATCATAT GGTCTTACGATAGACAACACAATGTGGTGTAATACTG TATTAGCTTCTGGTTCAGCAATTGGTTGTATAATTTAC ACAGGTAAAGATACTCGACAATCGATGAACACAACCTC AGCCCAAACCTGAAAACGGGCTTGTTAGAACTGGAAAT CAATAGTTTGTCCAAGATCTTATGTGTTTGTGTGTTTG CATTATCTGTCACTTAGTGCTATTCCAAGGAATAGCT GATGATTGGTACGTCGATATCATGCGGTTTCTCATTCT ATTCTCCACTATTATCCCAGTGTCTCTGAGAGTTAACC TTGATCTTGGAAGTCAAGTCCATGCTCATCAAAATAGA AACTGATAGCTCAATACCTGAAACCGTTGTTAGAACT AGTACAATACCGGAAGACCTGGGAAGAATTGAATACC TATTAAGTGACAAAACCTGGAACCTTACTCAAAATGA

		TATGGAAATGAAAAAACTACACCTAGGAACAGTCTCT TATGCTGGTGATACCATGGATATTATTTCTGATCATGT TAAAGGTCTTAATAACGCTAAAACATCGAGGAAAGAT CTTGGTATGAGAATAAGAGATTTGGTTACAACCTCTGG CCATCTG
70	DNA encodes <i>Drosophila</i> <i>melanogaster</i> ManII codon- optimized (KD)	AGAGACGATCCAATTAGACCTCCATTGAAGGTTGCTA GATCCCCAAGACCAGGTCAATGTCAAGATGTTGTTCA GGACGTCCCAAACGTTGATGTCCAGATGTTGGAGTTG TACGATAGAATGTCCTTCAAGGACATTGATGGTGGTG TTTGGAAGCAGGGTTGGAACATTAAGTACGATCCATT GAAGTACAACGCTCATCACAAGTTGAAGGTCTTCGTT GTCCACACTCCCACAACGATCCTGGTTGGATTGAGA CCTTCGAGGAATACTACCAGCACGACACCAAGCACAT CTTGTCCAACGCTTTGAGACATTTGCACGACAACCCA GAGATGAAGTTCATCTGGGCTGAAATCTCCTACTTCGC TAGATTCTACCACGATTTGGGTGAGAACAAGAAGTTG CAGATGAAGTCCATCGTCAAGAACGGTCAGTTGGAAT TCGTCACTGGTGGATGGGTCATGCCAGACGAGGCTAA CTCCCACTGGAGAAACGTTTTTGTTCAGTTGACCGAA GGTCAAACCTTGGTTGAAGCAATTCATGAACGTCACTC CAACTGCTTCCTGGGCTATCGATCCATTCGGACACTCT CCAACATATGCCATACATTTTGCAGAAGTCTGGTTTCAA GAATATGTTGATCCAGAGAACCCACTACTCCGTTAAG AAGGAGTTGGCTCAACAGAGACAGTTGGAGTTCTTGT GGAGACAGATCTGGGACAACAAAGGTGACACTGCTTT GTTACCCACATGATGCCATTCTACTCTTACGACATTC CTCATACCTGTGGTCCAGATCCAAAGGTTTGTTCAG TTCGATTTCAAAGAATGGGTTCTTCGGTTTGTCTTG TCCATGGAAGGTTCCACCTAGAACTATCTCTGATCAA AATGTTGCTGCTAGATCCGATTTGTTGGTTGATCAGTG GAAGAAGAAGGCTGAGTTGTACAGAACCAACGTCTTG TTGATTCCATTGGGTGACGACTTCAGATTCAAGCAGA ACACCGAGTGGGATGTTTCAGAGAGTCAACTACGAAAG ATTGTTTCGAACACATCAACTCTCAGGCTCACTTCAATG TCCAGGCTCAGTTCGGTACTTTGCAGGAATACTTCGAT GCTGTTACACAGGCTGAAAGAGCTGGACAAGCTGAGT TCCCAACCTTGTCTGGTGACTTCTTCACTTACGCTGAT AGATCTGATAACTACTGGTCTGGTTACTACACTTCCAG ACCATAACCATAAGAGAATGGACAGAGTCTTGATGCAC TACGTTAGAGCTGCTGAAATGTTGTCCGCTTGGCACTC CTGGGACGGTATGGCTAGAATCGAGGAAAGATTGGAG CAGGCTAGAAGAGAGTTGTCCTTGTTCAGCACCACG ACGGTATTACTGGTACTGCTAAAACTCACGTTGTCGTC GACTACGAGCAAAGAATGCAGGAAGCTTTGAAAGCTT GTCAAATGGTCATGCAACAGTCTGTCTACAGATTGTTG ACTAAGCCATCCATCTACTCTCCAGACTTCTCCTTCTC CTACTTCACTTTGGACGACTCCAGATGGCCAGGTTCTG GTGTTGAGGACTCTAGAACTACCATCATCTTGGGTGA GGATATCTTGCCATCCAAGCATGTTGTCATGCACAAC

		ACCTTGCCACACTGGAGAGAGCAGTTGGTTGACTTCT ACGTCTCCTCTCCATTCGTTTCTGTTACCGACTTGGCT AACAAATCCAGTTGAGGCTCAGGTTTCTCCAGTTTGGTC TTGGCACCACGACACTTTGACTAAGACTATCCACCCA CAAGGTTCCACCACCAAGTACAGAATCATCTTCAAGG CTAGAGTTCCACCAATGGGTTTGGCTACCTACGTTTTG ACCATCTCCGATTCCAAGCCAGAGCACACCTCCTACG CTTCCAATTTGTTGCTTAGAAAGAACCCAACCTTCCTTG CCATTGGGTCAATACCCAGAGGATGTCAAGTTCGGTG ATCCAAGAGAGATCTCCTTGAGAGTTGGTAACGGTCC AACCTTGGCTTTCTCTGAGCAGGGTTTGTGAAGTCCA TTCAGTTGACTCAGGATTCTCCACATGTTCCAGTTCAC TTCAAGTTCTTGAAGTACGGTGTTAGATCTCATGGTGA TAGATCTGGTGCTTACTTGTTCTTGCCAAATGGTCCAG CTTCTCCAGTCGAGTTGGGTCAGCCAGTTGTCTTGGTC ACTAAGGGTAAATTGGAGTCTTCCGTTTCTGTTGGTTT GCCATCTGTGCTTACCAGACCATCATGAGAGGTGGT GCTCCAGAGATTAGAAATTTGGTCGATATTGGTTCTTT GGACAACACTGAGATCGTCATGAGATTGGAGACTCAT ATCGACTCTGGTGATATCTTCTACACTGATTTGAATGG ATTGCAATTCATCAAGAGGAGAAGATTGGACAAGTTG CCATTGCAGGCTAACTACTACCCAATTCCATCTGGTAT GTTCAATTGAGGATGCTAATACCAGATTGACTTTGTTGA CCGGTCAACCATTGGGTGGATCTTCTTTGGCTTCTGGT GAGTTGGAGATTATGCAAGATAGAAGATTGGCTTCTG ATGATGAAAGAGGTTTGGGTCAGGGTGTTTTGGACAA CAAGCCAGTTTTGCATATTTACAGATTGGTCTTGGAGA AGGTAAACAACCTGTGTCAGACCATCTAAGTTGCATCC AGCTGGTTACTTGACTTCTGCTGCTCACAAAGCTTCTC AGTCTTTGTTGGATCCATTGGACAAGTTCATCTTCGCT GAAAATGAGTGGATCGGTGCTCAGGGTCAATTCGGTG GTGATCATCCATCTGCTAGAGAGGATTTGGATGTCTCT GTCATGAGAAGATTGACCAAGTCTTCTGCTAAAACCC AGAGAGTTGGTTACGTTTTGCACAGAACCAATTTGAT GCAATGTGGTACTCCAGAGGAGCATACTCAGAAGTTG GATGTCTGTCACTTGTGTCGCAATGTTGCTAGATGTGA GAGAACTACCTTGACTTTCTTGCAGAATTTGGAGCACT TGGATGGTATGGTTGCTCCAGAAGTTTGTCCAATGGA AACCGCTGCTTACGTCTCTTCTCACTCTTCTTGA
71	DNA encodes Mnn2 leader (53)	ATGCTGCTTACCAAAGGTTTTCAAAGCTGTTCAAGCT GACGTTCATAGTTTTGATATTGTGCGGGCTGTTCTGTC TTACAAACAAATACATGGATGAGAACACGTCG
72	Sequence of the PpHIS1 auxotrophic marker:	CAAGTTGCGTCCGGTATACGTAACGTCTCACGATGAT CAAAGATAATACTTAATCTTCATGGTCTACTGAATAAC TCATTTAAACAATTGACTAATTGTACATTATATTGAAC TTATGCATCCTATTAACGTAATCTTCTGGCTTCTCTCTC AGACTCCATCAGACACAGAATATCGTTCTCTCTAACTG GTCCTTTGACGTTTCTGACAATAGTTCTAGAGGAGTCG TCCAAAAACTCAACTCTGACTTGGGTGACACCACCAC

		GGGATCCGGTTCTTCCGAGGACCTTGATGACCTTGGCT AATGTAAGTGGAGTTTATAGTATCCATTTAAGATGTGT GTTTCTGTAGGTTCTGGGTTGGAAAAAATTTTAGACA CCAGAAGAGAGGAGTGAAGTGGTTTGCGTGGGTTTAG ACTGTGTAAGGCACTACTCTGTCTGAAGTTTATAGATAG GGGTTACCCGCTCCGATGCATGGGAAGCGATTAGCCC GGCTGTTGCCCGTTTGGTTTTTGAAGGGTAATTTTCAA TATCTCTGTTTGAGTCATCAATTTTCATATTCAAAGATT CAAAAACAAAATCTGGTCCAAGGAGCGCATTTAGGAT TATGGAGTTGGCGAATCACTTGAACGATAGACTATTA TTTGCTGTTCCCTAAAGAGGGCAGATTGTATGAGAAAT GCGTTGAATTACTTAGGGGATCAGATATTCAGTTTCGA AGATCCAGTAGATTGGATATAGCTTTGTGCACTAACCT GCCCCTGGCATTGGTTTTCTTCCAGCTGCTGACATTC CCACGTTTGTAGGAGAGGGTAAATGTGATTTGGGTAT AACTGGTATTGACCAGGTTTCAGGAAAGTGACGTAGAT GTCATACCTTTATTAGACTTGAATTTTCGGTAAGTGCAA GTTGCAGATTCAAGTTCCCGAGAATGGTGACTTGAAA GAACCTAAACAGCTAATTGGTAAAGAAATTGTTTCCT CCTTTACTAGCTTAACCACCAGGTACTTTGAACAACCTG GAAGGAGTTAAGCCTGGTGAGCCACTAAAGACAAAA ATCAAATATGTTGGAGGGTCTGTTGAGGCCTCTGTGTC CCTAGGAGTTGCCGATGCTATTGTGGATCTTGTTGAGA GTGGAGAAACCATGAAAGCGGCAGGGCTGATCGATAT TGAAACTGTTCTTTCTACTTCCGCTTACCTGATCTCTTC GAAGCATCCTCAACACCCAGAACTGATGGATACTATC AAGGAGAGAATTGAAGGTGTACTGACTGCTCAGAAGT ATGTCTTGTGTAATTACAACGCACCTAGAGGTAACCTT CCTCAGCTGCTAAAACTGACTCCAGGCAAGAGAGCTG CTACCGTTTCTCCATTAGATGAAGAAGATTGGGTGGG AGTGTCTCGATGGTAGAGAAGAAAGATGTTGGAAGA ATCATGGACGAATTAAAGAAACAAGGTGCCAGTGACA TTCTTGTCTTTGAGATCAGTAATTGTAGAGCATAGATA GAATAATATTCAAGACCAACGGCTTCTCTTCGGAAGC TCCAAGTAGCTTATAGTGATGAGTACCGGCATATATTT ATAGGCTTAAAATTTTCGAGGGTTCACTATATTCGTTTA GTGGGAAGAGTTCCTTTCACTCTTGTTATCTATATTGT CAGCGTGGACTGTTTATAACTGTACCAACTAGTTTCT TTCAACTCCAGGTTAAGAGACATAAATGTCCTTTGATG C
73	DNA encodes Rat GnT II (TC) Codon- optimized	TCCTTGGTTTACCAATTGAACTTCGACCAGATGTTGAG AAACGTTGACAAGGACGGTACTTGGTCTCCTGGTGAG TTGGTTTTGGTTGTTTCAGGTTTCAACAGACCAGAGTA CTTGAGATTGTTGATCGACTCCTTGAGAAAGGCTCAA GGTATCAGAGAGGTTTTGGTTATCTTCTCCACGATTT CTGGTCTGCTGAGATCAACTCCTTGATCTCCTCCGTTG ACTTCTGTCCAGTTTTGCAGGTTTTCTTCCATTCTCCA TCCAATTGTACCCATCTGAGTTCCAGGTTCTGATCCA AGAGACTGTCCAAGAGACTTGAAGAAGAACGCTGCTT

		TGAAGTTGGGTTGTATCAACGCTGAATACCCAGATTCT TTCGGTCACTACAGAGAGGCTAAGTTCTCCCAAATA AGCATCATTGGTGGTGGAAAGTTGCACTTTGTTTGGGA GAGAGTTAAGGTTTTGCAGGACTACACTGGATTGATC TTGTTCTTGGAGGAGGATCATTACTTGGCTCCAGACTT CTACCACGTTTTCAAGAAGATGTGGAAGTTGAAGCAA CAAGAGTGTCCAGGTTGTGACGTTTTGTCCTTGGGAAC TTACACTACTATCAGATCCTTCTACGGTATCGCTGACA AGGTTGACGTTAAGACTTGGAAGTCCACTGAACACAA CATGGGATTGGCTTTGACTAGAGATGCTTACCAGAAG TTGATCGAGTGTACTGACACTTTCTGTACTTACGACGA CTACAACCTGGGACTGGACTTTGCAGTACTTGACTTTGG CTTGTTTGCCAAAAGTTTGGAAGGTTTTGGTTCCACAG GCTCCAAGAATTTCCACGCTGGTGACTGTGGAATGC ACCACAAGAAAACCTTGTAGACCATCCACTCAGTCCGC TCAAATTGAGTCCTTGTGGAACAACAACAAGCAGTAC TTGTTCCAGAGACTTTGGTTATCGGAGAGAAGTTTCC AATGGCTGCTATTTCCCAAGAAAGAATGGTGGA TGGGGTGATATTAGAGACCACGAGTTGTGTAAATCCT ACAGAAGATTGCAGTAG
74	DNA encodes Mnn2 leader (54) The last 9 nucleotides are the linker containing the AseI restriction site)	ATGCTGCTTACCAAAAGGTTTTCAAAGCTGTTCAAGCT GACGTTTCATAGTTTTGATATTGTGCGGGCTGTTTCGTCA TTACAAACAAATACATGGATGAGAACACGTCGGTCAA GGAGTACAAGGAGTACTTAGACAGATATGTCCAGAGT TACTCCAATAAGTATTCATCTTCCTCAGACGCCGCCAG CGCTGACGATTCAACCCCATTTGAGGGACAATGATGAG GCAGGCAATGAAAAGTTGAAAAGCTTCTACAACAACG TTTTCAACTTTCTAATGGTTGATTCGCCCCGGGCGCGCC
75	Sequence of the 5'-Region used for knock out of PpARG1:	GATCTGGCCTTCCTGAATTTTTACGTCCAGCTATACG ATCCGTTGTGACTGTATTTCTGAAATGAAGTTTCAAC CTAAAGTTTTGGTTGTACTTGCTCCACCTACCACGGAA ACTAATATCGAAACCAATGAAAAAGTAGAACTGGAAT CGTCAATCGAAATTCGCAACCAAGTGGAACCCAAAGA CTTGAATCTTTCTAAAGTCTATTCTAGTGACACTAATG GCAACAGAAGATTTGAGCTGACTTTTCAAATGAATCT CAATAATGCAATATCAACATCAGACAATCAATGGGCT TTGTCTAGTGACACAGGATCAATTATAGTAGTGTCTTC TGCAGGAAGAATAACTTCCCGATCCTAGAAGTCGGG GCATCCGTCTGTGTCTTAAGATCGTACAACGAACACCT TTTGGCAATAACTTGTGAAGGAACATGCTTTTCATGGA ATTTAAAGAAGCAAGAATGTGTTCTAAACAGCATTTTC ATTAGCACCTATAGTCAATTCACACATGCTAGTTAAG AAAGTTGGAGATGCAAGGAACATTTCTATTGTATCTG CCGAAGGAGACAACAATCCGTTACCCAGATTCTAGA CTGCGAACTTTCCAAAAATGGCGCTCCAATTGTGGCTC TTAGCACGAAAGACATCTACTCTTATTCAAAGAAAAT GAAATGCTGGATCCATTTGATTGATTCGAAATACTTTG AATTGTTGGGTGCTGACAATGCACTGTTTGAGTGTGTG

		GAAGCGCTAGAAGGTCCAATTGGAATGCTAATTCATA GATTGGTAGATGAGTTCTTCCATGAAAACACTGCCGG TAAAAAACTCAAACCTTTACAACAAGCGAGTACTGGAG GACCTTTCAAATTCACCTTGAAGAACTAGGTGAAAATG CGTCTCAATTAAGAGAGAACTTGACAACTCTATGG TGATGAGGTTGAGGCTTCTTGACCTCTTCTCTCTATCT GCGTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGTTG AGCCAGACCGCGCTAAACGCATACCAATTGCCAAATC AGGCAATTGTGAGACAGTGGTAAAAAAGATGCCTGCA AAGTTAGATTCACACAGTAAGAGAGATCCTACTCATA AATGAGGCGCTTATTTAGTAGCTAGTGATAGCCACTG CGTTTCTGCTTTATGCTATTTGTTGTATGCCTTACTATC TTTGTGTTGGCTCCTTTTTCTTGACGTTTTCCGTTGGAGG GACTCCCTATTCTGAGTCATGAGCCGCACAGATTATCG CCCAAATTGACAAAATCTTCTGGCGAAAAAAGTATA AAAGGAGAAAAAAGCTCACCTTTTCCAGCGTAGAAA GTATATATCAGTCATTGAAGAC
76	Sequence of the 3'-Region used for knock out of PpARG1:	GGGACTTTAACTCAAGTAAAAGGATAGTTGTACAATT ATATATACGAAGAATAAATCATTACAAAAAGTATTCCG TTTCTTTGATTCTTAACAGGATTCAATTTCTGGGTGTCA TCAGGTACAGCGCTGAATATCTTGAAGTTAACATCGA GCTCATCATCGACGTTTCATCACACTAGCCACGTTTCCG CAACGGTAGCAATAATTAGGAGCGGACCACACAGTGA CGACATCTTTCTCTTTGAAATGGTATCTGAAGCCTTCC ATGACCAATTGATGGGCTCTAGCGATGAGTTGCAAGT TATTAATGTGGTTGAACTCACGTGCTACTCGAGCACCG AATAACCAGCCAGCTCCACGAGGAGAAACAGCCCAA CTGTGCACTTCATCTGGGTCAGACCAAACCAAGTCAC AAAATCCTCCTTCATGAGGGACCTCTTGCGCTCGGCTG AGAACTCTGATTTGATCTAACATGCGAATATCGGGAG AGAGACCACCATGGATACATAATATTTTACCATCAAT GATGGCACTAAGGGTTAAAAAAGTCGAACACCTGGCAA CAGTACTTCCAGACAGTGGTGGAACCATATTTATTGA GACATTCCTCATAAAATCCATAAACCTGAGTGATCTGT CTGGATTCATGATTTCCCCTTACCAATGTGATATGTTG AGGAACTTAATTTTTTAAATCATGAGTAACGTGAAC GTCTCCAACGAGAAATAGCCTCTATCCACATAGTCTCC TAGGAAGATATAGTTCTGTTTTATTCCATTAGAGGAGG ATCCGGGAAACCCACCACTAATCTTGAAAAGTTCCAG TAGATCGTGAAATTGGCCGTGAATATCTCCGCATACT GTCACTGGACTCTGCACTGGCTGTATATTGGATTCTC CATCAGCAAATCCTTCACCCGTTGCAAAGATGCTTCA TATCATTTTCACTTAAAGCCTTGACGCTTTTGACTTCTT CAAACCACTGATCTGGTCTCTTTCTGGCATGATTAAG GTCTATAATATTTCTGAGCTGAGATGTAAAAAAAAT AATAAAAATGGGGAGTGAAAAAGTGTGTAGCTTTTAG GAGTTTGGGATTGATACCCCAAATGATCTTTATGAG AATTAAAAGGTAGATACGCTTTTAATAAGAACACCTA TCTATAGTACTTTGTGGTCTTGAGTAATTGAGATGTTC

		AGCTTCTGAGGTTTGCCGTTATTCTGGGATAGTAGTGC GCGACCAAACAACCCGCCAGGCAAAGTGTGTTGTGCT CGAAGACGATTGCCAGAAGAGTAAGTCCGTCCTGCCT CAGATGTTACACACTTTCTTCCCTAGACAGTCGATGCA TCATCGGATTTAAACCTGAAACTTTGATGCCATGATAC GCCTAGTCACGTCGACTGAGATTTTAGATAAAGCCCCG ATCCCTTTAGTACATTCTGTTATCCATGGATGGAATG GCCTGATA
77	Sequence of the 5'-Region used for knock out of BMT4	AAGCTTGTTCACCGTTGGGACTTTTCCGTGGACAATGT TGACTACTCCAGGAGGGATTCCAGCTTTCTCTACTAGC TCAGCAATAATCAATGCAGCCCCAGGCGCCCGTTCTG ATGGCTTGATGACCGTTGTATTGCCTGTCACTATAGCC AGGGGTAGGGTCCATAAAGGAATCATAGCAGGGAAA TTAAAAGGGCATATTGATGCAATCACTCCCAATGGCT CTCTTGCCATTGAAGTCTCCATATCAGCACTAACTTCC AAGAAGGACCCCTTCAAGTCTGACGTGATAGAGCACG CTTGCTCTGCCACCTGTAGTCCTCTCAAAACGTCACCT TGTGCATCAGCAAAGACTTTACCTTGCTCCAATACTAT GACGGAGGCAATTCTGTCAAAATTCTCTCTCAGCAATT CAACCAACTTGAAAGCAAATTGCTGTCTCTTGATGAT GGAGACTTTTTTCCAAGATTGAAATGCAATGTGGGAC GACTCAATTGCTTCTTCCAGCTCCTCTTCGGTTGATTG AGGAACTTTTGAAACCACAAAATTGGTCGTTGGGTCA TGTACATCAAACCATTCTGTAGATTTAGATTCGACGAA AGCGTTGTTGATGAAGGAAAAGGTTGGATACGGTTTG TCGGTCTCTTTGGTATGGCCGGTGGGGTATGCAATTGC AGTAGAAGATAATTGGACAGCCATTGTTGAAGGTAGA GAAAAGGTCAGGGAACCTTGGGGGTATTATACCATT TTACCCACAAATAACAACTGAAAAGTACCCATTCCA TAGTGAGAGGTAACCGACGGAAAAAGACGGGCCCAT GTTCTGGGACCAATAGAACTGTGTAATCCATTGGGAC TAATCAACAGACGATTGGCAATATAATGAAATAGTTC GTTGAAAAGCCACGTCAGCTGTCTTTTCATTAACTTTG GTCGGACACAACATTTTCTACTGTTGTATCTGTCTAC TTTGCTTATCATCTGCCACAGGGCAAGTGGATTTCCTT CTCGCGCGGCTGGGTGAAAACGGTTAACGTGAA
78	Sequence of the 3'-Region used for knock out of BMT4	GCCTTGGGGGACTTCAAGTCTTTGCTAGAACTAGAT GAGGTCAGGCCCTCTTATGGTTGTGTCCCAATTGGGCA ATTTCACTCACCTAAAAAGCATGACAATTATTTAGCG AAATAGGTAGTATATTTTCCCTCATCTCCCAAGCAGTT TCGTTTTTGCATCCATATCTCTCAAATGAGCAGCTACG ACTCATTAGAACCAGAGTCAAGTAGGGGTGAGCTCAG TCATCAGCCTTCGTTTCTAAAACGATTGAGTCTTTTG TTGCTACAGGAAGCGCCCTAGGGAACTTTCGCACTTT GGAAATAGATTTTGATGACCAAGAGCGGGAGTTGATA TTAGAGAGGCTGTCCAAAGTACATGGGATCAGGCCGG CCAAATTGATTGGTGTGACTAAACCATTTGTGTACTTGG ACACTCTATTACAAAAGCGAAGATGATTTGAAGTATT ACAAGTCCCGAAGTGTTAGAGGATTCTATCGAGCCCA

		GAATGAAATCATCAACCGTTATCAGCAGATTGATAAA CTCTTGGAAGCGGTATCCCATTTTCATTATTGAAGAA CTACGATAATGAAGATGTGAGAGACGGCGACCCTCTG AACGTAGACGAAGAAACAAATCTACTTTTGGGGTACA ATAGAGAAAGTGAATCAAGGGAGGTATTTGTGGCCAT AATACTCAACTCTATCATTAAATG
79	Sequence of the 5'-Region used for knock out of BMT1	CATATGGTGAGAGCCGTTCTGCACAACCTAGATGTTTT GAGCTTCGCATTGTTTCCTGCAGCTCGACTATTGAATT AAGATTTCCGGATATCTCCAATCTCACAAAACCTTATG TTGACCACGTGCTTTCTGAGGCGAGGTGTTTTATATG CAAGCTGCCAAAATGGAAAACGAATGGCCATTTTTC GCCAGGCAAATTATTCGATTACTGCTGTCATAAAGA CAGTGTTGCAAGGCTCACATTTTTTTTTAGGATCCGAG ATAAAGTGAATACAGGACAGCTTATCTCTATATCTTGT ACCATTCGTGAATCTTAAGAGTTCGGTTAGGGGGACT CTAGTTGAGGGTTGGCACTCACGTATGGCTGGGCGCA GAAATAAAATTCAGGCGCAGCAGCACTTATCGATG
80	Sequence of the 3'-Region used for knock out of BMT1	GAATTCACAGTTATAAATAAAAAACAAAACCTCAAAA GTTTGGGCTCCACAAAATAACTTAATTTAAATTTTGT CTAATAAATGAATGTAATTCCAAGATTATGTGATGCA AGCACAGTATGCTTCAGCCCTATGCAGCTACTAATGTC AATCTCGCCTGCGAGCGGGCCTAGATTTTCACTACAA ATTTCAAACTACGCGGATTTATTGTCTCAGAGAGCA ATTTGGCATTCTGAGCGTAGCAGGAGGCTTCATAAG ATTGTATAGGACCGTACCAACAAATTGCCGAGGCACA ACACGGTATGCTGTGCACTTATGTGGCTACTTCCCTAC AACGGAATGAAACCTTCCTCTTTCCGCTTAAACGAGA AAGTGTGTCGCAATTGAATGCAGGTGCCTGTGCGCCT TGGTGTATTGTTTTTGAGGGCCCAATTTATCAGGCGCC TTTTTCTTGTTGTTTTCCCTTAGCCTCAAGCAAGGTT GGTCTATTTTCATCTCCGCTTCTATACCGTGCCTGATAC TGTTGGATGAGAACACGACTCAACTTCCTGCTGCTCTG TATTGCCAGTGTTTTGTCTGTGATTGATCGGAGTCC TCCTTACTTGGAATGATAATAATCTTGGCGGAATCTCC CTAAACGGAGGCAAGGATTCTGCCTATGATGATCTGC TATCATTGGGAAGCTT
81	Sequence of the 5'-Region used for knock out of BMT3	GATATCTCCCTGGGGACAATATGTGTTGCAACTGTTTCG TTGTTGGTGCCCCAGTCCCCAACCGGTACTAATCGGT CTATGTTCCCGTAACTCATATTCGGTTAGAACTAGAAC AATAAGTGCATCATTGTTCAACATTGTGGTTCAATTGT CGAACATTGCTGGTGCTTATATCTACAGGGAAGACGA TAAGCCTTTGTACAAGAGAGGTAACAGACAGTTAATT GGTATTTCTTTGGGAGTCGTTGCCCTCTACGTTGTCTC CAAGACATACTACATTCTGAGAAACAGATGGAAGACT CAAAAATGGGAGAAGCTTAGTGAAGAAGAGAAAGTT GCCTACTTGGACAGAGCTGAGAAGGAGAACCTGGGTT CTAAGAGGCTGGACTTTTTGTTCGAGAGTTAAACTGC ATAATTTTTCTAAGTAAATTTCATAGTTATGAAATTT CTGCAGCTTAGTGTTTACTGCATCGTTTACTGCATCAC

		CCTGTAAATAATGTGAGCTTTTTTCCTTCCATTGCTTG GTATCTTCCTTGCTGCTGTTT
82	Sequence of the 3'-Region used for knock out of BMT3	ACAAAACAGTCATGTACAGAACTAACGCCTTTAAGAT GCAGACCACTGAAAAGAATTGGGTCCCATTTTTCTTG AAAGACGACCAGGAATCTGTCCATTTTGTTTACTCGTT CAATCCTCTGAGAGTACTCAACTGCAGTCTTGATAAC GGTGCATGTGATGTTCTATTTGAGTTACCACATGATTT TGGCATGTCTTCCGAGCTACGTGGTGCCACTCCTATGC TCAATCTTCCTCAGGCAATCCCGATGGCAGACGACAA AGAAATTTGGGTTTCATTCCCAAGAACGAGAATATCA GATTGCGGGTGTCTGAAACAATGTACAGGCCAATGT TAATGCTTTTTGTTAGAGAAGGAACAACTTTTTTGCT GAGC
83	DNA encodes Tr ManI catalytic domain	CGCGCCGGATCTCCCAACCCTACGAGGGCGGCAGCAG TCAAGGCCGCATTCCAGACGTCGTGGAACGCTTACCA CCATTTTGCCTTTCCCATGACGACCTCCACCCGGTCA GCAACAGCTTTGATGATGAGAGAAACGGCTGGGGCTC GTCGGCAATCGATGGCTTGGACACGGCTATCCTCATG GGGGATGCCGACATTGTGAACACGATCCTTCAGTATG TACCGCAGATCAACTTCACCACGACTGCGGTTGCCAA CCAAGGCATCTCCGTGTTTCGAGACCAACATTTCGGTAC CTCGGTGGCCTGCTTTCTGCCTATGACCTGTTGCGAGG TCCTTTCAGCTCCTTGGCGACAAACCAGACCCTGGTAA ACAGCCTTCTGAGGCAGGCTCAAACACTGGCCAACGG CCTCAAGGTTGCGTTCACCACTCCAGCGGTGTCCCG GACCCTACCGTCTTCTTCAACCCTACTGTCCGGAGAAG TGGTGCATCTAGCAACAACGTCGCTGAAATTGGAAGC CTGGTGCTCGAGTGGACACGGTTGAGCGACCTGACGG GAAACCCGCAGTATGCCAGCTTGCGCAGAAGGGCGA GTCGTATCTCCTGAATCCAAAGGGAAGCCCGGAGGCA TGGCCTGGCCTGATTGGAACGTTTGTGACGACGAGCA ACGGTACCTTTCAGGATAGCAGCGGCAGCTGGTCCGG CCTCATGGACAGCTTCTACGAGTACCTGATCAAGATG TACCTGTACGACCCGGTTGCGTTTGCACACTACAAGG ATCGCTGGGTCTTGTGCGGACTCGACCATTGCGCAT CTCGCCTCTACCCGTCGACGCGCAAGGACTTGACCTT TTTGTCTTCGTACAACGGACAGTCTACGTCGCCAAACT CAGGACATTTGGCCAGTTTTGCCGGTGGCAACTTCATC TTGGGAGGCATTCTCCTGAACGAGCAAAAGTACATTG ACTTTGGAATCAAGCTTGCCAGCTCGTACTTTGCCACG TACAACCAGACGGCTTCTGGAATCGGCCCCGAAGGCT TCGCGTGGGTGGACAGCGTGACGGGCGCCGGCGGCTC GCCGCCCTCGTCCAGTCCGGGTTCTACTCGTCGGCAG GATTCTGGGTGACGGCACCGTATTACATCCTGCGGCC GGAGACGCTGGAGAGCTTGTAACGACATACCGCGTC ACGGGCGACTCCAAGTGGCAGGACCTGGCGTGGGAA GCGTTCAGTGCCATTGAGGACGCATGCCGCGCCGGCA GCGCGTACTCGTCCATCAACGACGTGACGCAGGCCAA CGGCGGGGGTGCCTCTGACGATATGGAGAGCTTCTGG

		TTTGCCGAGGCGCTCAAGTATGCGTACCTGATCTTTGCGGAGGAGTCGGATGTGCAGGTGCAGGCCAACGGCGGGAACAAATTTGTCTTTAACACGGAGGCGCACCCCTTTAGCATCCGTTTCATCATCACGACGGGGCGGCCACCTTGCTTAA
84	5'ARG1 and ORF	TACCAATTGCCAAATCAGGCAATTGTGAGACAGTGGTAA AAAGATGCCTGCAAAGTTAGATTCACACAGTAAGAGAGAT CCTACTCATAAATGAGGCGCTTATTTAGTAGCTAGTGATAG CCACTGCGGTTCTGCTTTATGCTATTTGTTGTATGCCTTACT ATCTTTGTTTGGCTCCTTTTCTTGACGTTTTCCGTTGGAGG GACTCCCTATTCTGAGTCATGAGCCGCACAGATTATCGCCC AAAATTGACAAAATCTTCTGGCGAAAAAAGTATAAAAGGA GAAAAAAGCTCACCCCTTTTCCAGCGTAGAAAGTATATATCA GTCATTGAAGACTATTATTTAAATAACACAATGTCTAAAGG AAAAGTTTGTTTGGCCTACTCCGGTGGTTTGGATACCTCCA TCATCCTAGCTTGGTTGTTGGAGCAGGGATACGAAGTCGTT GCCTTTTTAGCCAACATTGGTCAAGAGGAAGACTTTGAGGC TGCTAGAGAGAAAGCTCTGAAGATCGGTGCTACCAAGTTT ATCGTCAGTGACGTTAGGAAGGAATTTGTTGAGGAAGTTT GTTCCCAGCAGTCCAAGTTAACGCTATCTACGAGAACGTCT ACTTACTGGGTACCTCTTTGGCCAGACCAGTCATTGCCAAG GCCCAAATAGAGGTTGCTGAACAAGAAGGTTGTTTTGCTGT TGCCACCGTTGTACCGGAAAGGGTAACGATCAGGTTAGA TTTGAGCTTTCCTTTTATGCTCTGAAGCCTGACGTTGTCTGT ATCGCCCCATGGAGAGACCCAGAATTCTTCGAAAGATTCTG CTGGTAGAAATGACTTGCTGAATTACGCTGCTGAGAAGGA TATCCAGTTGCTCAGACTAAAGCCAAGCCATGGTCTACTG ATGAGAACATGGCTCACATCTCCTTCGAGGCTGGTATTCTA GAAGATCCAAACACTACTCCTCCAAGGACATGTGGAAGC TCACTGTTGACCCAGAAGATGCACCAGACAAGCCAGAGTT CTTGACGTCCACTTTGAGAAGGGTAAGCCAGTTAAATTAG TTCTCGAGAACAAAACACTGAGGTCACCGATCCGGTTGAGAT CTTTTGTACTGCTAACGCCATTGCTAGAAGAAACGGTGTTG GTAGAATTGACATTGTGCGAGAACAGATTCATCGGAATCAA GTCCAGAGGTTGTTATGAAACTCCAGGTTTGACTCTACTGA GAACCACTCACATCGACTTGGAAGGTCTTACCGTTGACCGT GAAGTTAGATCGATCAGAGACACTTTTGTACCCCAACCTA CTCTAAGTTGTTATACAACGGGTTGTACTTTACCCCAAGAG GTGAGTACGTCAGAACTATGATTCAGCCTTCTCAAAACACC GTCAACGGTGTTGTTAGAGCCAAGGCCTACAAAGGTAATG TGTATAACCTAGGAAGATACTCTGAAACCGAGAAATTGTA CGATGCTACCGAATCTTCCATGGATGAGTTGACCGGATTCC ACCCTCAAGAAGCTGGAGGATTTATCACAACACAAGCCAT CAGAATCAAGAAGTACGGAGAAAGTGTGAGAGAGAAGGG AAAGTTTTTGGGACTTTAACTCAAGTAAAAGGATAGTTGTA CAATTATATATACGAAGAATAAATCATTACAAAAAGTATT CGTTTCTTTGATTCTTAACAGGATTCATTTTCTGGGTGTCAT CAGGTACAGCGCTGAATATCTTGAAGTTAACATCGAGCTCA TCATCGACGTTTCATCACTAGCCACGTTTCCGCAACGGTA G
85	PpCITI TT	CCGGCCATTTAAATATGTGACGACTGGGTGATCCGGGTTAG TGAGTTGTTCTCCCATCTGTATATTTTTCATTTACGATGAAT ACGAAATGAGTATTAAGAAATCAGGCGTAGCAATATGGGC

		AGTGTTTCAGTCCTGTCATAGATGGCAAGCACTGGCACATCC TTAATAGGTTAGAGAAAATCATTGAATCATTTGGGTGGTGA AAAAAAATTGATGTAAACAAGCCACCCACGCTGGGAGTCG AACCCAGAATCTTTTGATTAGAAAGTCAAACGCGTTAACCAT TACGCTACGCAGGCATGTTTCACGTCCATTTTGTATTGCTTT CTATCATAATCTAAAGATGTGAACTCAATTAGTTGCAATTT GACCAATTCTTCCATTACAAGTCGTGCTTCCTCCGTTGATG CAAC
86	<i>Ashbya gossypii</i> <i>TEF1</i> promoter	GATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCG GCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACA GTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTG TCGCCCGTACATTTAGCCCATACATCCCATGTATAAT CATTTGCATCCATACATTTTGTATGGCCGCACGGCGCGA AGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGC AGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCC CCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAG GATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTT AAAATCTTGCTAGGATACAGTTCTCACATCACATCCG AACATAAACAACC
87	<i>Ashbya gossypii</i> <i>TEF1</i> termination sequence	TAATCAGTACTGACAATAAAAAGATTCTTGTTTTCAAG AACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTT CTATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTT CGCCTCGACATCATCTGCCCAGATGCGAAGTTAAGTG CGCAGAAAGTAATATCATGCGTCAATCGTATGTGAAT GCTGGTCGCTATACTGCTGTCGATTTCGATACTAACGCC GCCATCCAGTGTCGAAAAC
88	Sequence of the PpPMA1 promoter:	AAATGCGTACCTCTTCTACGAGATTCAAGCGAATGAG AATAATGTAATATGCAAGATCAGAAAGAATGAAAGG AGTTGAAAAAAAACCGTTGCGTTTTGACCTTGAAT GGGGTGGAGGTTTCCATTCAAAGTAAAGCCTGTGTCT TGGTATTTTCGGCGGCACAAGAAATCGTAATTTTCATC TTCTAAACGATGAAGATCGCAGCCCAACCTGTATGTA GTTAACCGGTCGGAATTATAAGAAAGATTTTCGATCA ACAAACCCTAGCAAATAGAAAGCAGGGTTACAACCTT AAACCGAAGTCACAAACGATAAACCCTCAGCTCCCA CCCAAATTCATTCCCCTAGCAGAAAGGAATTATTTA ATCCCTCAGGAAACCTCGATGATTCTCCCGTTCTTCCA TGGGCGGGTATCGCAAAATGAGGAATTTTCAAATTT CTCTATTGTCAAGACTGTTTATTATCTAAGAAATAGCC CAATCCGAAGCTCAGTTTTTGAAAAAATCACTTCCGCG TTTCTTTTTTACAGCCCGATGAATATCCAAATTTGGAA TATGGATTACTCTATCGGGACTGCAGATAATATGACA ACAACGCAGATTACATTTTAGGTAAGGCATAAACACC AGCCAGAAATGAAACGCCCACTAGCCATGGTCGAATA GTCCAATGAATTCAGATAGCTATGGTCTAAAAGCTGA TGTTTTTTTATTGGGTAATGGCGAAGAGTCCAGTACGAC TTCCAGCAGAGCTGAGATGGCCATTTTGGGGGTATT AGTAACTTTTTGAGCTCTTTTCACTTCGATGAAGTGTC CCATTCGGGATATAATCGGATCGCGTCGTTTTCTCGAA AATACAGCTTAGCGTCGTCCGCTTGTTGTAAAAGCAG

		CACCACATTCCTAATCTCTTATATAAAACAAAACAACCC AAATTATCAGTGCTGTTTTCCACAGATATAAGTTTC TTTTCTCTCCGCTTTTTGATTTTTTATCTCTTTCCTTAA AAACTTCTTTACCTTAAAGGGCGGCC
89	Sequence of the 5'-region that was used to knock into the PpPRO1 locus:	GAAGGGCCATCGAATTGTCATCGTCTCCTCAGGTGCC ATCGCTGTGGGCATGAAGAGAGTCAACATGAAGCGGA AACCAAAAAAGTTACAGCAAGTGCAGGCATTGGCTGC TATAGGACAAGGCCGTTTGATAGGACTTTGGGACGAC CTTTTCCGTCAGTTGAATCAGCCTATTGCGCAGATTTT ACTGACTAGAACGGATTTGGTCGATTACACCCAGTTT AAGAACGCTGAAAATACATTGGAACAGCTTATTA AAA TGGGTATTATTCCTATTGTCAATGAGAATGACACCCTA TCCATTCAAGAAATCAAATTTGGTGACAATGACACCT TATCCGCCATAACAGCTGGTATGTGTCATGCAGACTA CCTGTTTTTGGTGACTGATGTGGACTGTCTTTACACGG ATAACCCTCGTACGAATCCGGACGCTGAGCCAATCGT GTTAGTTAGAAATATGAGGAATCTAAACGTCAATACC GAAAGTGGAGGTTCCGCCGTAGGAACAGGAGGAATG ACAATAAATTGATCGCAGCTGATTTGGGTGTATCTGC AGGTGTTACAACGATTATTTGCAAAGTGAACATCCC GAGCAGATTTTGGACATTGTAGAGTACAGTATCCGTG CTGATAGAGTCGAAAATGAGGCTAAATATCTGGTCAT CAACGAAGAGGAACTGTGGAACAATTTCAAGAGATC AATCGGTCAGAACTGAGGGAGTTGAACAAGCTGGACA TTCCTTTGCATACACGTTTCGTTGGCCACAGTTTTAAT GCTGTTAATAACAAAGAGTTTTTGGTTACTCCATGGACT AAAGGCCAACGGAGCCATTATCATTGATCCAGGTTGT TATAAGGCTATCACTAGAAAAACAAAGCTGGTATTC TTCCAGCTGGAATTATTTCCGTAGAGGGTAATTTCCAT GAATACGAGTGTGTTGATGTTAAGGTAGGACTAAGAG ATCCAGATGACCCACATTCACTAGACCCCAATGAAGA ACTTTACGTCGTTGGCCGTGCCCGTTGTAATTACCCCA GCAATCAAATCAACAAAATTAAGGGTCTACAAAGCTC GCAGATCGAGCAGGTTCTAGGTTACGCTGACGGTGAG TATGTTGTTACAGGGACAACCTTGGCTTTCCAGTATT TGCCGATCCAGAACTGTTGGATGTTGTTGAGAGTACC CTGTCTGAACAGGAGAGAGAATCCAAACCAAATAAAT AG
90	Sequence of the 3'-region that was used to knock into the PpPRO1 locus:	AATTTACATATGCTGCTTGATTATGTAATTATACCTT GCGTTCGATGGCATCGATTTCCTCTTCTGTCAATCGCG CATCGCATTAAAAGTATACTTTTTTTTTTTTCTATAGT ACTATTCGCCTTATTATAAACTTTGCTAGTATGAGTTC TACCCCAAGAAAGAGCCTGATTTGACTCCTAAGAAG AGTCAGCCTCCAAAGAATAGTCTCGGTGGGGGTAAAG GCTTTAGTGAGGAGGGTTTCTCCAAGGGGACTTCAG CGCTAAGCATATACTAAATCGTCGCCCTAACACCGAA GGCTCTTCTGTGGCTTCGAACGTCATCAGTTCGTCATC ATTGCAAAGGTTACCATCCTCTGGATCTGGAAGCGTT GCTGTGGGAAGTGTGTTGGGATCTTCGCCATTAACTCT

		TTCTGGAGGGTTCCACGGGCTTGATCCAACCAAGAAT AAAATAGACGTTCCAAAGTCGAAACAGTCAAGGAGA CAAAGTGTTCTTTCTGACATGATTTCCACTTCTCATGC AGCTAGAAATGATCACTCAGAGCAGCAGTTACAAACT GGACAACAATCAGAACAAAAAGAAGAAGATGGTAGT CGATCTTCTTTTTCTGTTTCTTCCCCCGCAAGAGATATC CGGCACCCAGATGTACTGAAAACGTGTCGAGAAACATC TTGCCAATGACAGCGAGATCGACTCATCTTTACAACCTT CAAGGTGGAGATGTCAGTAGAGGCATTTATCAATGGG TAACTGGAGAAAGTAGTCAAAAAGATAACCCGCCTTT GAAACGAGCAAATAGTTTTAATGATTTTTCTTCTGTGC ATGGTGACGAGGTAGGCAAGGCAGATGCTGACCACG ATCGTGAAAGCGTATTCGACGAGGATGATATCTCCAT TGATGATATCAAAGTTCGGGGAGGGATGCGTCGAAGT TTTTTATTACAAAAGCATAGAGACCAACAACCTTTCTGG ACTGAATAAAACGGCTCACCAACCAAAAACAACCTTACT AAACCTAATTTCTTCACGAACAACCTTTATAGAGTTTTT GGCATTGTATGGGCATTTTGCAGGTGAAGATTTGGAG GAAGACGAAGATGAAGATTTAGACAGTGGTTCCGAAT CAGTCGCAGTCAGTGATAGTGAGGGAGAATTCAGTGA GGCTGACAACAATTTGTTGTATGATGAAGAGTCTCTCC TATTAGCACCTAGTACCTCCAACCTATGCGAGATCAAG AATAGGAAGTATTCGTA CTCTACTTATGGATCTTTCA GTTCAAATGTTGGTTCTTCGTCTATTCATCAGCAGTTA ATGAAAAGTCAAATCCCGAAGCTGAAGAAACGTGGA CAGCACAAGCATAAAACACAATCAAAAATACGCTCGA AGAAGCAAACCTACCACCGTAAAAGCAGTGTTGCTGCT ATTAAA
91	Sequence of the PpTRP2 gene integration locus:	GGTTTCTCAATTACTATATACTACTAACCATTACCTG TAGCGTATTTCTTTTCCCTCTTCGCGAAAGCTCAAGGG CATCTTCTTGACTCATGAAAAATATCTGGATTTCTTCT GACAGATCATCACCTTGAGCCCAACTCTCTAGCCTAT GAGTGTAAGTGATAGTCATCTTGCAACAGATTATTTTG GAACGCAACTAACAAGCAGATACACCCTTCAGCAGA ATCCTTTCTGGATATTGTGAAGAATGATCGCCAAAGTC ACAGTCCTGAGACAGTTCTTAATCTTTACCCCATTTAC AAGTTCATCCAATCAGACTTCTTAACGCCTCATCTGGC TTATATCAAGCTTACCAACAGTTCAGAACTCCAGTC CAAGTTTCTTGCTTGAAAGTGCGAAGAATGGTGACAC CGTTGACAGGTACACCTTTATGGGACATTCCCCCAGA AAAATAATCAAGACTGGGCCTTTAGAGGGTGCTGAAG TTGACCCCTTGGTGCTTCTGGAAAAAGAACTGAAGGG CACCAGACAAGCGCAACTTCCTGGTATTCTCGTCTAA GTGGTGGTGCCATAGGATACATCTCGTACGATTGTATT AAGTACTTTGAACCAAAAACCTGAAAGAAAACCTGAAA GATGTTTTGCAACTTCCGGAAGCAGCTTTGATGTTGTT CGACACGATCGTGGCTTTTGACAATGTTTATCAAAGAT TCCAGGTAATTGGAAACGTTTCTCTATCCGTTGATGAC TCGGACGAAGCTATTCTTGAGAAATATTATAAGACAA

		GAGAAGAAGTGGAAAAGATCAGTAAAGTGGTATTTGA CAATAAACTGTTCCCTACTATGAACAGAAAGATATT ATTCAAGGCCAAACGTTACCTCTAATATTGGTCAGG AAGGGTATGAAAACCATGTTTCGCAAGCTGAAAGAACA TATTCTGAAAGGAGACATCTTCCAAGCTGTTCCCTCTC AAAGGGTAGCCAGGCCGACCTCATTGCACCCTTTCAA CATCTATCGTCATTTGAGAACTGTCAATCCTTCTCCAT ACATGTTCTATATTGACTATCTAGACTTCCAAGTTGTT GGTGCTTCACCTGAATTACTAGTTAAATCCGACAACA ACAACAAAATCATCACACATCCTATTGCTGGAACCTT CCCAGAGGTAAACTATCGAAGAGGACGACAATTATG CTAAGCAATTGAAGTCGTCTTTGAAAGACAGGGCCGA GCACGTCATGCTGGTAGATTTGGCCAGAAATGATATT AACCGTGTGTGTGAGCCCACCAGTACCACGGTTGATC GTTTATTGACTGTGGAGAGATTTTCTCATGTGATGCAT CTTGTGTCAGAAGTCAGTGGAACATTGAGACCAAACA AGACTCGCTTCGATGCTTTTCAGATCCATTTTCCCAGCA GGAACCGTCTCCGGTGCTCCGAAGGTAAGAGCAATGC AACTCATAGGAGAATTGGAAGGAGAAAAGAGAGGTG TTTATGCGGGGGCCGTAGGACACTGGTCGTACGATGG AAAATCGATGGACACATGTATTGCCTTAAGAACAATG GTCGTCAAGGACGGTGTGCTTACCTTCAAGCCGGAG GTGGAATTGTCTACGATTCTGACCCCTATGACGAGTAC ATCGAAACCATGAACAAAATGAGATCCAACAATAACA CCATCTTGGAGGCTGAGAAAATCTGGACCGATAGGTT GGCCAGAGACGAGAATCAAAGTGAATCCGAAGAAAA CGATCAATGAACGGAGGACGTAAGTAGGAATTTATG
92	Human UDP- GlcNAc 2- epimerase/N- acetylmannosam- ine kinase (HsGNE) codon optimized	ATGGAAAAGAACGGTAACAACAGAAAGTTGAGAGTTT GTGTTGCTACTTGTAACAGAGCTGACTACTCCAAGTTG GCTCCAATCATGTTCCGGTATCAAGACTGAGCCAGAGT TCTTCGAGTTGGACGTTGTTGTTTTGGGTTCCTTCTG ATTGATGACTACGGTAACACTTACAGAATGATCGAGC AGGACGACTTCGACATCAACACTAGATTGCACACTAT TGTTAGAGGAGAGGACGAAGCTGCTATGGTTGAATCT GTTGGATTGGCTTTGGTTAAGTTGCCAGACGTTTTGAA CAGATTGAAGCCAGACATCATGATTGTTACGGTGAC AGATTCGATGCTTTGGCTTTGGCTACTTCCGCTGCTTT GATGAACATTAGAATCTTGACATCGAGGGTGGTGAA GTTTCTGGTACTATCGACGACTCCATCAGACACGCTAT CACTAAGTTGGCTCACTACCATGTTTGTGTACTAGAT CCGCTGAGCAACACTTGATTTCCATGTGTGAGGACCA CGACAGAATTTTGTGGCTGGTTGTCCATCTTACGACA AGTTGTTGTCCGCTAAGAACAAGGACTACATGTCCAT CATCAGAATGTGGTTGGGTGACGACGTTAAGTCTAAG

	GACTACATCGTTGCTTTGCAGCACCCAGTTACTACTGA CATCAAGCACTCCATCAAGATGTTTCGAGTTGACTTTGG ACGCTTTGATCTCCTTCAACAAGAGAACTTTGGTTTTG TCCCCAAACATTGACGCTGGTTCCAAAGAGATGGTTA GAGTTATGAGAAAGAAGGGTATCGAACACCACCCAA ACTTCAGAGCTGTTAAGCACGTTCCATTCGACCAATTC ATCCAGTTGGTTGCTCATGCTGGTTGTATGATCGGTAA CTCCTCCTGTGGTGTAGAGAAGTTGGTGCTTTTCGGTA CTCCAGTTATCAACTTGGGTACTAGACAGATCGGTAG AGAGACTGGAGAAAACGTTTTTGCATGTTAGAGATGCT GACACTCAGGACAAGATTTTGCAGGCTTTGCACTTGC AATTCGGAAAGCAGTACCCATGTTCCAAAATCTACGG TGACGGTAACGCTGTTCCAAGAATCTTGAAGTTTTTGA AGTCCATCGACTTGCAAGAGCCATTGCAGAAGAAGTT CTGTTTCCCACCAGTTAAGGAGAACATCTCCCAGGAC ATTGACCACATCTTGGAGACATTGTCCGCTTTGGCTGT TGATTTGGGTGGAACCTTGAAGAGTTGCTATCGTTT CCATGAAGGGAGAGATCGTTAAGAAGTACACTCAGTT CAACCCAAAGACTTACGAGGAGAGAATCAACTTGATC TTGCAGATGTGTGTTGAAGCTGCTGCTGAGGCTGTAA GTTGAACTGTAGAATCTTGGGTGTTGGTATCTCTACTG GTGGTAGAGTTAATCCAAGAGAGGGTATCGTTTTGCA CTCCACTAAGTTGATTCAGGAGTGGAACCTCCGTTGATT TGAGAACTCCATTGTCCGACACATTGCACTTGCCAGTT TGGGTTGACAACGACGGTAATTGTGCTGCTTTGGCTG AGAGAAAGTTCGGTCAAGGAAAGGGATTGGAGAACTT CGTTACTTTGATCACTGGTACTGGTATTGGTGGTGGTA TCATTCACCAGCACGAGTTGATTCACGGTTCTTCCTTC TGTGCTGCTGAATTGGGACACTTGGTTGTTTCTTTGGA CGGTCCAGACTGTTCTTGTGGTTCCCACGGTTGTATTG AAGCTTACGCATCAGGAATGGCATTGCAGAGAGAGGC TAAGAAGTTGCACGACGAGGACTTGTTGTTGGTTGAG GGAATGTCTGTTCCAAAGGACGAGGCTGTTGGTGCTT TGCATTTGATCCAGGCTGCTAAGTTGGGTAATGCTAA GGCTCAGTCCATCTTGAGAACTGCTGGTACTGCTTTGG GATTGGGTGTTGTTAATATCTTGCACACTATGAACCCA TCCTTGGTTATCTTGTCCGGTGTTTTGGCTTCTCACTAC
--	--

		ATCCACATCGTTAAGGACGTTATCAGACAGCAAGCTT TGTCTCCGTTCAAGACGTTGATGTTGTTTCCGAC TTGGTTGACCCAGCTTTGTTGGGTGCTGCTTCCATGGT TTTGGACTACACTACTAGAAGAATCTACTAATAG
93	Sequence of the PpARG1 auxotrophic marker:	CAGTTGAGCCAGACCGCGCTAAACGCATACCAATTGC CAAATCAGGCAATTGTGAGACAGTGGTAAAAAAGATG CCTGCAAAGTTAGATTCACACAGTAAGAGAGATCCTA CTCATAAATGAGGCGCTTATTTAGTAGCTAGTGATAG CCACTGCGGTTCTGCTTTATGCTATTTGTTGTATGCCTT ACTATCTTTGTTTGGCTCCTTTTTCTTGACGTTTTCCGT TGGAGGGACTCCCTATTCTGAGTCATGAGCCGCACAG ATTATCGCCCAAATTGACAAAATCTTCTGGCGAAAA AAGTATAAAAGGAGAAAAAAGCTCACCCTTTTCCAGC GTAGAAAGTATATATCAGTCATTGAAGACTATTATTTA AATAACACAATGTCTAAAGGAAAAGTTTGTTTGGCCT ACTCCGGTGGTTTGGATACCTCCATCATCCTAGCTTGG TTGTTGGAGCAGGGATACGAAGTCGTTGCCTTTTTAGC CAACATTGGTCAAGAGGAAGACTTTGAGGCTGCTAGA GAGAAAGCTCTGAAGATCGGTGCTACCAAGTTTATCG TCAGTGACGTTAGGAAGGAATTTGTTGAGGAAGTTTT GTTCCCAGCAGTCCAAGTTAACGCTATCTACGAGAAC GTCTACTTACTGGGTACCTCTTTGGCCAGACCAGTCAT TGCCAAGGCCCAAATAGAGGTTGCTGAACAAGAAGGT TGTTTTGCTGTTGCCACGGTTGTACCGGAAAGGGTAA CGATCAGGTTAGATTTGAGCTTTCCTTTTATGCTCTGA AGCCTGACGTTGTCTGTATCGCCCCATGGAGAGACCC AGAATTCTTCGAAAGATTCGCTGGTAGAAATGACTTG CTGAATTACGCTGCTGAGAAGGATATTCCAGTTGCTC AGACTAAAGCCAAGCCATGGTCTACTGATGAGAACAT GGCTCACATCTCCTTCGAGGCTGGTATTCTAGAAGATC CAAACACTACTCCTCCAAAGGACATGTGGAAGCTCAC TGTTGACCCAGAAGATGCACCAGACAAGCCAGAGTTC TTTGACGTCCACTTTGAGAAGGGTAAGCCAGTTAAAT TAGTTCTCGAGAACAAAACCTGAGGTCACCGATCCGGT TGAGATCTTTTTGACTGCTAACGCCATTGCTAGAAGAA ACGGTGTTGGTAGAATTGACATTGTCGAGAACAGATT CATCGGAATCAAGTCCAGAGGTTGTTATGAAACTCCA

		GGTTTGACTCTACTGAGAACCCTCACATCGACTTGG AAGGTCTTACCGTTGACCGTGAAGTTAGATCGATCAG AGACACTTTTGTACCCCAACCTACTCTAAGTTGTTAT ACAACGGGTTGTACTTTACCCCAAGGTGAGTACGT CAGAACTATGATTCAGCCTTCTCAAAACACCGTCAAC GGTGTGTTAGAGCCAAGGCCTACAAAGGTAATGTGT ATAACCTAGGAAGATACTCTGAAACCGAGAAATTGTA CGATGCTACCGAATCTTCCATGGATGAGTTGACCGGA TTCCACCCTCAAGAAGCTGGAGGATTTATCACAACAC AAGCCATCAGAATCAAGAAGTACGGAGAAAGTGTC GAGAGAAGGGAAAGTTTTTGGGACTTTAACTCAAGTA AAAGGATAGTTGTACAATTATATATACGAAGAATAAA TCATTACAAAAAGTATTCGTTTCTTTGATTCTTAACAG GATTCATTTTCTGGGTGTCATCAGGTACAGCGCTGAAT ATCTTGAAGTTAACATCGAGCTCATCATCGACGTTTAT CACACTAGCCACGTTTCCGCAACGGTAGCAATAATTA GGAGCGGACCACACAGTGACGACATC
94	Human CMP-sialic acid synthase (HsCSS) codon optimized	ATGGACTCTGTTGAAAAGGGTGCTGCTACTTCTGTTTC CAACCCAAGAGGTAGACCATCCAGAGGTAGACCTCCT AAGTTGCAGAGAACTCCAGAGGTGGTCAAGGTAGA GGTGTGAAAAGCCACCACACTTGGCTGCTTTGATCTT GGCTAGAGGAGGTTCTAAGGGTATCCCATTGAAGAAC ATCAAGCACTTGGCTGGTGTTCATTGATTGGATGGGT TTTGAGAGCTGCTTTGGACTCTGGTGCTTTCCAATCTG TTTGGGTTTCCACTGACCACGACGAGATTGAGAACGT TGCTAAGCAATTCCGGTGCTCAGGTTACAGAAGATCC TCTGAGGTTTCCAAGGACTCTTCTACTTCCTTGGACGC TATCATCGAGTTCTTGAACCTACCACAACGAGGTTGAC ATCGTTGGTAACATCCAAGCTACTTCCCCATGTTTGCA CCCAACTGACTTGCAAAAAGTTGCTGAGATGATCAGA GAAGAGGGTTACGACTCCGTTTTCTCCGTTGTTAGAAG GCACCAGTTCAGATGGTCCGAGATTGAGAAGGGTGTT AGAGAGGTTACAGAGCCATTGAACTTGAACCCAGCTA AAAGACCAAGAAGGCAGGATTGGGACGGTGAATTGT ACGAAAACGGTTCCTTCTACTTCGCTAAGAGACACTT GATCGAGATGGGATACTTGCAAGGTGGAAAGATGGCT TACTACGAGATGAGAGCTGAACACTCCGTTGACATCG

		ACGTTGATATCGACTGGCCAATTGCTGAGCAGAGAGT TTTGAGATACGGTTACTTCGGAAAGGAGAAGTTGAAG GAGATCAAGTTGTTGGTTTGTAAACATCGACGGTTGTTT GACTAACGGTCACATCTACGTTTCTGGTGACCAGAAG GAGATTATCTCCTACGACGTTAAGGACGCTATTGGTAT CTCCTTGTTGAAGAAGTCCGGTATCGAAGTTAGATTG ATCTCCGAGAGAGCTTGTTCCAAGCAAACATTGTCCTC TTTGAAGTTGGACTGTAAGATGGAGGTTTCCGTTTCTG ACAAGTTGGCTGTTGTTGACGAATGGAGAAAGGAGAT GGGTTTGTGTTGGAAGGAAGTTGCTTACTTGGGTAAC GAAGTTTCTGACGAGGAGTGTTTGAAGAGAGTTGGTT TGTCTGGTGCTCCAGCTGATGCTTGTTCCACTGCTCAA AAGGCTGTTGGTTACATCTGTAAGTGTAACGGTGGTA GAGGTGCTATTAGAGAGTTCGCTGAGCACATCTGTTT GTTGATGGAGAAAGTTAATAACTCCTGTCAGAAGTAG TAG
95	Human N- acetylneuraminat e-9-phosphate synthase (HsSPS) codon optimized	ATGCCATTGGAATTGGAGTTGTGTCCTGGTAGATGGG TTGGTGGTCAACACCCATGTTTCATCATCGCTGAGATC GGTCAAAACCACCAAGGAGACTTGGACGTTGCTAAGA GAATGATCAGAATGGCTAAGGAATGTGGTGCTGACTG TGCTAAGTTCCAGAAGTCCGAGTTGGAGTTCAAGTTC AACAGAAAGGCTTTGGAAAGACCATACACTTCCAAGC ACTCTTGGGGAAAGACTTACGGAGAACACAAGAGAC ACTTGGAGTTCTCTCACGACCAATACAGAGAGTTGCA GAGATACGCTGAGGAAGTTGGTATCTTCTTCACTGCTT CTGGAATGGACGAAATGGCTGTTGAGTTCTTGACGA GTTGAACGTTCCATTCTTCAAAGTTGGTTCCGGTGACA CTAACAACTTCCCATACTTGGAAAAGACTGCTAAGAA AGGTAGACCAATGGTTATCTCCTCTGGAATGCAGTCT ATGGACACTATGAAGCAGGTTTACCAGATCGTTAAGC CATTGAACCCAACTTTTGTTTCTTGCAGTGTACTTCC GCTTACCCATTGCAACCAGAGGACGTTAATTTGAGAG TTATCTCCGAGTACCAGAAGTTGTTCCCAGACATCCCA ATTGGTTACTCTGGTCACGAGACTGGTATTGCTATTTT CGTTGCTGCTGTTGCTTTGGGTGCTAAGGTTTGGAGA GACACATCACTTTGGACAAGACTTGGAAAGGGTTCTGA TCACTCTGCTTCTTTGGAACCTGGTGAGTTGGCTGAAC

		TTGTTAGATCAGTTAGATTGGTTGAGAGAGCTTTGGGT TCCCCAACTAAGCAATTGTTGCCATGTGAGATGGCTTG TAACGAGAAGTTGGGAAAGTCCGTTGTTGCTAAGGTT AAGATCCCAGAGGGTACTATCTTGACTATGGACATGT TGACTGTTAAAGTTGGAGAGCCAAAGGGTTACCCACC AGAGGACATCTTTAACTTGGTTGGTAAAAAGGTTTTG GTTACTGTTGAGGAGGACGACACTATTATGGAGGAGT TGGTTGACAACCACGGAAGAAGATCAAGTCCTAG
96	Mouse alpha- 2,6-sialyl transferase catalytic domain (MmmST6) codon optimized	GTTTTTCAAATGCCAAAGTCCCAGGAGAAAGTTGCTG TTGGTCCAGCTCCACAAGCTGTTTTCTCCAACCTCCAAG CAAGATCCAAAGGAGGGTGTTCAAATCTTGTCCTACC CAAGAGTTACTGCTAAGGTAAAGCCACAACCATCCTT GCAAGTTTGGGACAAGGACTCCACTTACTCCAAGTTG AACCCAAGATTGTTGAAGATTTGGAGAACTACTTGA ACATGAACAAGTACAAGGTTTCCTACAAGGGTCCAGG TCCAGGTGTTAAGTTCTCCGTTGAGGCTTTGAGATGTC ACTTGAGAGACCACGTAAACGTTTCCATGATCGAGGC TACTGACTTCCCATTCAACACTACTGAATGGGAGGGA TACTTGCCAAAGGAGAACTTCAGAACTAAGGCTGGTC CATGGCATAAGTGTGCTGTTGTTTCTTCTGCTGGTTCC TTGAAGAACTCCCAGTTGGGTAGAGAAATTGACAACC ACGACGCTGTTTTGAGATTCAACGGTGCTCCAACCTGA CAACTTCCAGCAGGATGTTGGTACTAAGACTACTATC AGATTGGTTAACTCCCAATTGGTTACTACTGAGAAGA GATTCTTGAAGGACTCCTTGTACACTGAGGGAATCTTG ATTTTGTGGGACCCATCTGTTTACCACGCTGACATTCC ACAATGGTATCAGAAGCCAGACTACAACCTTCTTCGAG ACTTACAAGTCCTACAGAAGATTGCACCCATCCCAGC CATTCTACATCTTGAAGCCACAAATGCCATGGGAATT GTGGGACATCATCCAGGAAATTTCCCCAGACTTGATC CAACCAAACCCACCATCTTCTGGAATGTTGGGTATCAT CATCATGATGACTTTGTGTGACCAGGTTGACATCTACG AGTTCTTGCCATCCAAGAGAAAGACTGATGTTTGTTAC TACCACCAGAAGTTCTTCGACTCCGCTTGTAATATGGG AGCTTACCACCCATTGTTGTTTCGAGAAGAACATGGTT AAGCACTTGAACGAAGGTACTGACGAGGACATCTACT TGTTCCGGAAGGCTACTTTGTCCGTTTCAGAAACAA

		CAGATGTTAG
97	Human UDP- GlcNAc 2- epimerase/N- acetylmannosam- ine kinase (HsGNE) codon optimized	ATGGAAAAGAACGGTAACAACAGAAAGTTGAGAGTTT GTGTTGCTACTTGTAAACAGAGCTGACTACTCCAAGTTG GCTCCAATCATGTTTCGGTATCAAGACTGAGCCAGAGT TCTTCGAGTTGGACGTTGTTGTTTTGGGTTCCCACTTG ATTGATGACTACGGTAACACTTACAGAATGATCGAGC AGGACGACTTCGACATCAACACTAGATTGCACACTAT TGTTAGAGGAGAGGACGAAGCTGCTATGGTTGAATCT GTTGGATTGGCTTTGGTTAAGTTGCCAGACGTTTTGAA CAGATTGAAGCCAGACATCATGATTGTTACGGTGAC AGATTCGATGCTTTGGCTTTGGCTACTTCCGCTGCTTT GATGAACATTAGAATCTTGCACATCGAGGGTGGTGAA GTTTCTGGTACTATCGACGACTCCATCAGACACGCTAT CACTAAGTTGGCTCACTACCATGTTTGTTGTACTAGAT CCGCTGAGCAACACTTGATTTCCATGTGTGAGGACCA CGACAGAATTTTGTGGCTGGTTGTCCATCTTACGACA AGTTGTTGTCCGCTAAGAACAAGGACTACATGTCCAT CATCAGAATGTGGTTGGGTGACGACGTTAAGTCTAAG GACTACATCGTTGCTTTGCAGCACCCAGTTACTACTGA CATCAAGCACTCCATCAAGATGTTTCGAGTTGACTTTGG ACGCTTTGATCTCCTTCAACAAGAGAACTTTGGTTTTG TTCCCAAACATTGACGCTGGTTCCAAAGAGATGGTTA GAGTTATGAGAAAGAAGGGTATCGAACACCACCCAA ACTTCAGAGCTGTTAAGCACGTTCCATTCGACCAATTC ATCCAGTTGGTTGCTCATGCTGGTTGTATGATCGGTAA CTCCTCCTGTGGTGTTAGAGAAGTTGGTGCTTTTCGGTA CTCCAGTTATCAACTTGGGTACTAGACAGATCGGTAG AGAGACTGGAGAAAACGTTTTGCATGTTAGAGATGCT GACACTCAGGACAAGATTTTGCAGGCTTTGCACTTGC AATTCGGAAGCAGTACCCATGTTCCAAAATCTACGG TGACGGTAACGCTGTTCCAAGAATCTTGAAGTTTTTGA AGTCCATCGACTTGCAAGAGCCATTGCAGAAGAAGTT CTGTTTCCCACCAGTTAAGGAGAACATCTCCAGGAC ATTGACCACATCTTGGAGACATTGTCCGCTTTGGCTGT TGATTTGGGTGGAACAACTTGAGAGTTGCTATCGTTT CCATGAAGGGAGAGATCGTTAAGAAGTACACTCAGTT CAACCCAAAGACTTACGAGGAGAGAATCAACTTGATC

		TTGCAGATGTGTGTTGAAGCTGCTGCTGAGGCTGTAA GTTGAACTGTAGAATCTTGGGTGTTGGTATCTCTACTG GTGGTAGAGTTAATCCAAGAGAGGGTATCGTTTTGCA CTCCACTAAGTTGATTCAGGAGTGGAACCTCCGTTGATT TGAGAACTCCATTGTCCGACACATTGCACTTGCCAGTT TGGGTGACAACGACGGTAATTGTGCTGCTTTGGCTG AGAGAAAAGTTCGGTCAAGGAAAGGGATTGGAGAACTT CGTTACTTTGATCACTGGTACTGGTATTGGTGGTGGTA TCATTCACCAGCACGAGTTGATTCACGGTTCTTCCTTC TGTGCTGCTGAATTGGGACACTTGGTTGTTTCTTTGGA CGGTCCAGACTGTTCTTGTGGTTCCACGGTTGTATTG AAGCTTACGCATCAGGAATGGCATTGCAGAGAGAGGC TAAGAAGTTGCACGACGAGGACTTGTTGTTGGTTGAG GGAATGTCTGTTCCAAAGGACGAGGCTGTTGGTGCTT TGCATTTGATCCAGGCTGCTAAGTTGGGTAATGCTAA GGCTCAGTCCATCTTGAGAACTGCTGGTACTGCTTTGG GATTGGGTGTTGTTAATATCTTGACACTATGAACCCA TCCTTGGTTATCTTGTCCGGTGTTTTGGCTTCTCACTAC ATCCACATCGTTAAGGACGTTATCAGACAGCAAGCTT TGTCTCCGTTCAAGACGTTGATGTTGTTGTTTCCGAC TTGGTTGACCCAGCTTTGTTGGGTGCTGCTTCCATGGT TTTGGACTACACTACTAGAAGAATCTACTAATAG
98	Pp TRP2: 5' and ORF	ACTGGGCCTTTAGAGGGTGCTGAAGTTGACCCCTTGG TGCTTCTGGAAAAAGAACTGAAGGGCACCAGACAAGC GCAACTTCCTGGTATTCCTCGTCTAAGTGGTGGTGCCA TAGGATACATCTCGTACGATTGTATTAAGTACTTTGAA CCAAAACTGAAAGAAAAGTGAAGATGTTTTGCAAC TTCCGGAAGCAGCTTTGATGTTGTTTCGACACGATCGTG GCTTTTGACAATGTTTATCAAAGATTCCAGGTAATTGG AAACGTTTCTCTATCCGTTGATGACTCGGACGAAGCTA TTCTTGAGAAATATTATAAGACAAGAGAAGAAGTGGA AAAGATCAGTAAAGTGGTATTTGACAATAAACTGTT CCCTACTATGAACAGAAAGATATTATTCAAGGCCAAA CGTTCACCTCTAATATTGGTCAGGAAGGGTATGAAAA CCATGTTTCGCAAGCTGAAAGAACATATTCTGAAAGGA GACATCTTCCAAGCTGTTCCCTCTCAAAGGGTAGCCA GGCCGACCTCATTGCACCCTTTCAACATCTATCGTCAT

		TTGAGAACTGTCAATCCTTCTCCATACATGTTCTATAT TGACTATCTAGACTTCCAAGTTGTTGGTGCTTCACCTG AATTACTAGTTAAATCCGACAACAACAACAAAATCAT CACACATCCTATTGCTGGAACCTCTCCAGAGGTA ACTATCGAAGAGGACGACAATTATGCTAAGCAATTGA AGTCGTCTTTGAAAGACAGGGCCGAGCACGTCATGCT GGTAGATTTGGCCAGAAATGATATTAACCGTGTGTGT GAGCCCACCAGTACCACGGTTGATCGTTTATTGACTGT GGAGAGATTTTCTCATGTGATGCATCTTGTGTCAGAAG TCAGTGGAACATTGAGACCAAACAAGACTCGCTTCGA TGCTTTCAGATCCATTTTCCCAGCAGGTACCGTCTCCG GTGCTCCGAAGGTAAGAGCAATGCAACTCATAGGAGA ATTGGAAGGAGAAAAGAGAGGTGTTTATGCGGGGGCC GTAGGACACTGGTCGTACGATGGAAAATCGATGGACA CATGTATTGCCTTAAGAACAATGGTCGTCAAGGACGG TGTCGCTTACCTTCAAGCCGGAGGTGGAATTGTCTACG ATTCTGACCCCTATGACGAGTACATCGAAACCATGAA CAAAATGAGATCCAACAATAACACCATCTTGAGGGCT GAGAAAATCTGGACCGATAGGTTGGCCAGAGACGAG AATCAAAGTGAATCCGAAGAAAACGATCAATGA
99	PpTRP2 3' region	ACGGAGGACGTAAGTAGGAATTTATGTAATCATGCCA ATACATCTTTAGATTTCTTCTCTTCTTTTAAACGAAAG ACCTCCAGTTTTGCACTCTCGACTCTCTAGTATCTTCC CATTCTGTGTGCTGCAACCTCTTGCCTTCTGTTTCCTTC AATTGTTCTTCTTTCTTCTGTTGCACTTGGCCTTCTTCC TCCATCTTTCGTTTTTTTTCAAGCCTTTTCAGCAGTTCT TCTTCCAAGAGCAGTTCTTTGATTTTCTCTCTCCAATCC ACCAAAAACTGGATGAATTCAACCGGGCATCATCAA TGTTCCACTTTCTTTCTTATCAATAATCTACGTGCTT CGGCATACGAGGAATCCAGTTGCTCCCTAATCGAGTC ATCCACAAGGTTAGCATGGGCCTTTTTCAGGGTGTCA AAAGCATCTGGAGCTCGTTTATTCGGAGTCTTGTCTGG ATGGATCAGCAAAGACTTTTTCGGGAAAGTCTTTCTTA TATCTTCCGGAGAACAACCTGGTTTCAAATCCAAGAT GGCATAGCTGTCCAATTTGAAAGTGGAAGAATCCTG CCAATTTCTTCTCTCGTGTCTCAGCTCGTTCTCTCTCTT TGCAACAGGTCCACTTCATCTGGCATTCTTTCTTTATGT

		TAACTTTAATTATTATTAATTATAAAAGTTGATTATCGT TATCAAAAATAATCATATTCGAGAAATAATCCGTCCAT GCAATATATAAATAAGAATTCATAATAATGTAATGAT AACAGTACCTCTGATGACCTTTGATGAACCGCAATTTT CTTTCCAATGACAAGACATCCCTATAATAACAATTATAC AGTTTATATATCACAAATAATCACCTTTTTATAAGAAA ACCGTCCTCTCCGTAACAGAACTTATTATCCGCACGTT ATGGTTAACACACTACTAATACCGATATAGTGTATGA AGTCGCTACGAGATAGCCATCCAGGAACTTACCAAT TCATCAGCACTTTCATGATCCGATTGTTGGCTTTATTC TTTGCGAGACAGATACTTGCCAATGAAATAACTGATC CCACAGATGAGAATCCGGTGCTCGT
100	Sequence of the 5'-Region used for knock out of STE13	TTGGGGGCCTCCAGGACTTGCTGAAATTTGCTGACTCA TCTTCGCCATCCAAGGATAATGAGTTAGCTAATGTGA CAGTTAATGAGTCGTCTTGACTAACGGGGAACATTTC ATTATTTATATCCAGAGTCAATTTGATAGCAGAGTTTG TGGTTGAAATACCTATGATTCGGGAGACTTTGTTGTAA CGACCATTATCCACAGTTTGGACCGTGAAAATGTCAT CGAAGAGAGCAGACGACATATTATCTATTGTGGTAAG TGATAGTTGGAAGTCCGACTAAGGCATGAAAATGAGA AGACTGAAAATTTAAAGTTTTTGAAAACACTAATCGG GTAATAACTTGGAATTACGTTTACGTGCCTTTAGCTC TTGTCCTTACCCCTGATAATCTATCCATTTCGAGAG ACAATGACATCTCGGACAGCTGAGAACCCGTTTCGATA TAGAGCTTCAAGAGAATCTAAGTCCACGTTCTTCCAAT TCGTCCATATTGGAAAACATTAATGAGTATGCTAGAA GACATCGCAATGATTCGCTTTCCCAAGAATGTGATAA TGAAGATGAGAACGAAAATCTCAATTATACTGATAAC TTGGCCAAGTTTTCAAAGTCTGGAGTATCAAGAAAGA GCTGTATGCTAATATTTGGTATTTGCTTTGTTATCTGG CTGTTTCTCTTTGCCTTGATGCGAGGGACAATCGATT TTCCAATTTGAACGAGTACGTTCCAGATTCAAACAG
101	Sequence of the 3'-Region used for knock out of STE13	CTACTGGGAACCACGAGACATCACTGCAGTAGTTTCC AAGTGGATTTTCAGATCACTCATTTGTGAATCCTGACAA AACTGCGATATGGGGGTGGTCTTACGGTGGGTTCCT ACGCTTAAGACATTGGAATATGATTCTGGAGAGGTTT TCAAATATGGTATGGCTGTTGCTCCAGTAACTAATTGG

		CTTTTGTATGACTCCATCTACACTGAAAGATACATGAA CCTTCCAAAGGACAATGTTGAAGGCTACAGTGAACAC AGCGTCATTAAGAAGGTTTCCAATTTTAAGAATGTAA ACCGATTCTTGTTTTGTCACGGGACTACTGATGATAAC GTGCATTTTCAGAACACACTAACCTTACTGGACCAGTT CAATATTAATGGTGTGTTGTGAATTACGATCTTCAGGTGT ATCCCGACAGTGAACATAGCATTGCCCATCACAACGC AAATAAAGTGATCTACGAGAGGTTATTCAAGTGGTTA GAGCGGGCATTTAACGATAGATTTTTTGTAACATTCCGT ACTTCATGCCATACTATATATCCTGCAAGGTTTCCCTT TCAGACACAATAATTGCTTTGCAATTTTACATAACCACC AATTGGCAAAAATAATCTCTTCAGTAAGTTGAATGCTT TTCAAGCCAGCACCGTGAGAAATTGCTACAGCGCGCA TTCTAACATCACTTTAAAATTCCCTCGCCGGTGCTCAC TGGAGTTTCCAACCCTTAGCTTATCAAAATCGGGTGAT AACTCTGAGTTTTTTTTTTCACTTCTATTCCTAAACCTT CGCCCAATGCTACCACCTCCAATCAACATCCCGAAAT GGATAGAAGAGAATGGACATCTCTTGCAACCTCCGGT TAATAATTACTGTCTCCACAGAGGAGGATTTACGGTA ATGATTGTAGGTGGGCCTAATG
102	Sequence of the 5'-Region used for knock out of DAP2	CACCTGGGCCTGTTGCTGCTGGTACTGCTGTTGGAAC GTTGGTATTGTTGCTGATCTAAGGCCGCCTGTTCCACA CCGTGTGTATCGAATGCTTGGGCAAAATCATCGCCTG CCGGAGGCCCCACTACCGCTTGTTCCCTCCTGCTCTTGT TTGTTTTGCTCATTGATGATATCGGCGTCAATGAATTG ATCCTCAATCGTGTGGTGGTGGTGTGCGTGATTCCCTCTT CTTCTTGAGTGCCTTATCCATATTCCTATCTTAGTGTA CCAATAATTTTGTTAAACACACGCTGTTGTTTATGAAA AGTCGTCAAAAGGTTAAAAATTCTACTTGGTGTGTGTC AGAGAAAGTAGTGCAGACCCCCAGTTTGTTGACTAGT TGAGAAGGCGGCTCACTATTGCGCGAATAGCATGAGA AATTTGCAAACATCTGGCAAAGTGGTCAATACCTGCC AACCTGCCAATCTTCGCGACGGAGGCTGTAAAGCGGG TTGGGTTCCCAAAGTGAATGGATATTACGGGCAGGAA AAACAGCCCCCTCCACACTAGTCTTTGCTACTGACATC TTCCTCTCATGTATCCCGAACACAAGTATCGGGAGTA TCAACGGAGGGTGCCCTTATGGCAGTACTCCCTGTTG

		GTGATTGTACTGCTATACGGGTCTCATTGCTTATCAG CACCATCAACTTGATACACTATAACCACAAAAATTAT CATGCACACCCAGTCAATAGTGGTATCGTTCTTAATGA GTTTGCTGATGACGATTCTCTCTTTGAATGGCACTC TGAACCTGGGAGAACTGGAGAAATGGTACCTTTTCCCC TAAATTTTCATTCCATTTCAGTGGACCGAAATAGGTCAG GAAGATGACCAGGGATATTACATTCTCTCTTCCAATTC CTCTTACATAGTAAAGTCTTTATCCGACCCAGACTTTG AATCTGTTCTATTCAACGAGTCTACAATCACTTACAAC G
103	Sequence of the 3'-Region used for knock out of DAP2	GGCAGCAAAGCCTTACGTTGATGAGAATAGACTGGCC ATTTGGGGTTGGTCTTATGGAGGTTACATGACGCTAA AGGTTTTAGAACAGGATAAAGGTGAAACATTCAAATA TGGAAATGTCTGTTGCCCTGTGACGAATTGGAAATTCT ATGATTCTATCTACACAGAAAGATACATGCACACTCC TCAGGACAATCCAAACTATTATAATTCGTCAATCCATG AGATTGATAATTTGAAGGGAGTGAAGAGGTTCTTGCT AATGCACGGAACCTGGTGACGACAATGTTCACTTCCAA AATACACTCAAAGTTCTAGATTTATTTGATTTACATGG TCTTGAAAACCTATGATATCCACGTGTTCCCTGATAGTG ATCACAGTATTAGATATCACAACGGTAATGTTATAGT GTATGATAAGCTATTCCATTGGATTAGGCGTGCATTCA AGGCTGGCAAATAAATAGGTGCAAAAATATTATTAGA CTTTTTTTTTCGTTTCGCAAGTTATTACTGTGTACCATAC CGATCCAATCCGTATTGTAATTCATGTTCTAGATCCAA AATTTGGGACTCTAATTCATGAGGTCTAGGAAGATGA TCATCTCTATAGTTTTTCAGCGGGGGGCTCGATTTGCGG TTGGTCAAAGCTAACATCAAAATGTTTGTGAGGTTCA GTGAATGGTAACTGCTGCTCTTGAATTGGTCGTCTGAC AAATTCTCTAAGTGATAGCACTTCATCTACAATCATTT GCTTCATCGTTTCTATATCGTCCACGACCTCAAACGAG AAATCGAATTTGGAAGAACAGACGGGCTCATCGTTAG GATCATGCCAAACCTTGAGATATGGATGCTCTAAAGC CTCAGTAACTGTAATTCTGTGAGTGGGATCTACCGTGA GCATTCGATCCAGTAAGTCTATCGCTTCAGGGTTGGCA CCGGGAAATAACTGGCTGAATGGGATCTTGGGCATGA ATGGCAGGGAGCGAACATAATCCTGGGCACGCTCTGA

		TCTGATAGACTGAAGTGTCTCTTCCGAAACAGTACCC AGCGTACTCAAAATCAAGTTCAATTGATCCACATAGT CTCTTCCTCTAAAAATGGGTCGGCCACCTA
104	HYG ^R resistance cassette	GATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCG GCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACA GTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTG TCGCCCCGTACATTTAGCCCATACATCCCCATGTATAAT CATTTGCATCCATACATTTTGATGGCCGCACGGCGCGA AGCAAAAATTACGGCTCCTCGCTGCGGACCTGCGAGC AGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCC CCACGCCGCGCCCCTGTAGAGAAATATAAAAGGTTAG GATTTGCCACTGAGGTTCTTCTTTTCATATACTTCCTTTT AAAATCTTGCTAGGATACAGTTCTCACATCACATCCG AACATAAACAACCATGGGTAAAAAGCCTGAACTCACC GCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCG ACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGA AGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGT GGATATGTCTGCGGGTAAATAGCTGCGCCGATGGTT TCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCG GCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGG AATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGT GCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCG AACTGCCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCAT GGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGC GGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAAT ACACTACATGGCGTGATTTTCATATGCGCGATTGCTGAT CCCCATGTGTATCACTGGCAAACGTGTATGGACGACA CCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCT GATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCAC CTCGTGACGCGGATTTCCGGCTCCAACAATGTCCTGAC GGACAATGGCCGCATAACAGCGGTCATTGACTGGAGC GAGGCGATGTTGCGGGATTCCCAATACGAGGTCGCCA ACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAG CAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGC TTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCG CATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACG GCAATTTTCGATGATGCAGCTTGGGCGCAGGGTTCGATG

		CGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGG CGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGA CCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGA CCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAATAA TCAGTACTGACAATAAAAAGATTCTTGTTTTCAAGAA CTTGTCATTTGTATAGTTTTTTTTATATTGTAGTTGTTCT ATTTTAATCAAATGTTAGCGTGATTATATTTTTTTTCG CCTCGACATCATCTGCCCAGATGCGAAGTTAAGTGCG CAGAAAGTAATATCATGCGTCAATCGTATGTGAATGC TGGTCGCTATACTGCTGTGATTTCGATACTAACGCCGC CATCCAGTGTCGAAAACGAGCT
105	Sequence of <i>PpTRP5</i> 5' integration fragment	ACGACGGCCAAATTCATGATACACACTCTGTTTCAGCT GGTTTGGACTACCCTGGAGTTGGTCCTGAATTGGCTGC CTGGAAAGCAAATGGTAGAGCCCAATTTTCCGCTGTA ACTGATGCCCAAGCATTAGAGGGATTCAAAATCCTGT CTCAATTGGAAGGGATCATTCCAGCACTAGAGTCTAG TCATGCAATCTACGGCGCATTGCAAATTGCAAAGACT ATGTCTTCGGACCAGTCCTTAGTTATTAATGTATCTGG AAGGGGTGATAAGGACGTCCAGAGTGTAGCTGAGATT TTACCTAAATTGGGACCTCAAATTGGATGGGATTTGC GTTTCAGCGAAGACATTACTAAAGAGTGA
106	Sequence of <i>PpTRP5</i> 3' integration fragment	TCGATAGCACAATATTCAACTTGACTGGGTGTTAAGA ACTAAGAGCTCTGGGAAACTTTGTATTTATTACTACCA ACACAGTCAAATTATTGGATGTGTTTTTTTTTCCAGTA CATTTCACTGAGCAGTTTGTATATACTCGGTCTTTAATC TCCATATACATGCAGATTGTAATACAGATCTGAACAG TTTGATTCTGATTGATCTTGCCACCAATATTCTATTTTT GTATCAAGTAACAGAGTCAATGATCATTGGTAACGTA ACGGTTTTTCGTGTATAGTAGTTAGAGCCCATCTTGTA CCTCATTTTCTCCCATATTAAAGTATCAGTGATTGCT GGAACGATTAACTAAGAAAAAAAAAATATCTGCACAT ACTCATCAGTCTGTAAATCTAAGTCAAACTGCTGTAT CCAATAGAAATCGGGATATACCTGGATGTTTTTTCCAC ATAAACAAACGGGAGTTCAGCTTACTTATGGTGTTGA TGCAATTCAGTATGATCCTACCAATAAAACGAACTT TGGGATTTTGGCTGTTTGAGGGATCAAAAGCTGCACC TTTACAAGATTGACGGATCGACCATTAGACCAAAGCA

		AATGGCCACCAA
107	DNA encodes GM-CSF	CCAGCTAGATCTCCATCTCCATCCACTCAACCATGGGAACA CGTTAACGCTATCCAAGAGGCTTTGAGATTGTTGAACTTGT CCAGAGACACTGCTGCTGAAATGAACGAGACTGTTGAGGT TATCTCCGAGATGTTGCGACTTGCAAGAGCCAACCTTGTTTGC AGACTAGATTGGAGTTGTACAAGCAGGGATTGAGAGGATC CTTGACTAAGTTGAAGGGACCATTGACTATGATGGCTTCCC ACTACAAGCAACACTGTCCACCAACTCCAGAAACATCCTGT GCTACTCAGATCATCACTTTCGAGTCCTT CAAAGAGAACTTGAAGGACTTCTTGTTGGTTATCCCATTG ACTGTTGGGAACCAGTTCAAGAATAATAA
108	GM-CSF	PARSPSPSTQPWEHVNAIQEALRLNLSRDTAAEMNETVEVIS EMFDLQEPTCLQTRLELYKQGLRGS�TKLKGPLTMMASHYK QHCPTPETSCATQIITFESFKENLKDFLLVIPFDCWEPVQE
109	DNA encodes CWP1-GMCSF fusion protein	ATGTTCAACCTGAAAACCTATTCTCATCTCAACACTTGC ATCGATCGCTGTTGCCGACCAAACCTTCGGTGTCTTC TAATCCGGAGTGGATCCCCATATCACTATTCGACTCTC ACTAATAGAGACGAAAAGATTGTTGCTGGAGGTGGCA ACAAAAAAGTGACCCTCACAGATGAGGGAGCTCTGAA GTATGATGGTGGTAAATGGATAGGTCTTGATGATGAT GGCTATGCGGTACAGACCGACAAACCAGTTACAGGT GGAGCACTAACGGTGGATACCTCTATTTTGACCAAGG CTTAATTGTTTGACGGAGGACTATATCGGATATGTGA AGAAACATGGTGAATGCAAAGGTGACAGCTATGGTAT GGCTTGGAAGGTACTCCCAGCCGACGATGACAAGGAT GATGACAAGGATGATGATAAAGATGATGACAAGGATT ATGACGATGACAATGACCACGGTGATGGTGATTACTA TTGCTCGATCACAGGAACCTATGCCATCAAATCCAAA GGCAGTAAGCATCAATACGAGGCCATCAAAAAAGTTG ATGCACATCCTCATGTCTTCTCTGTAGGAGGAGATCAG GGAAACGATCTGATTGTGACTTTCCAAAAGGATTGTT CGCTGGTAGATCAAGATAACAGAGGCGTATATGTTGA CCCTAATTCTGGAGAAGTCGGAAACGTTGACCCTTGG GGAGAACTCACGCCATCTGTAAATGGGATATTGACG ACGGATACCTGATCTTTAATGGTGAGTCCAATTTCAGG TCATGTCCATCTGGTAATGGATATTCATTGTCTATCAA GGATTGTGTTGGGGGAACTGACATTGGCCTTAAAGTA TGGGAGAAAGGTGGAGGTTCTTTGGTTAAGAGGGCTC CAGCTAGATCTCCATCTCCATCCACTCAACCATGGGA ACACGTTAACGCTATCCAAGAGGCTTTGAGATTGTTG AACTTGTCCAGAGACACTGCTGCTGAAATGAACGAGA CTGTTGAGGTTATCTCCGAGATGTTGCGACTTGCAAGAG CCAACCTGTTTGCAGACTAGATTGGAGTTGTACAAGC AGGGATTGAGAGGATCCTTGACTAAGTTGAAGGGACC

		ATTGACTATGATGGCTTCCCACTACAAGCAACACTGTC CACCAACTCCAGAAACATCCTGTGCTACTCAGATCAT CACTTTCGAGTCCTTCAAAGAGAACTTGAAGGACTTCT TGTTGGTTATCCCATTCGACTGTTGGGAACCAGTTCAA GAATAA
110	CWP1-GMCSF fusion protein	MFNLKTILISTLASIAVADQTFGVLLIRSGSPYHYSTLTNR DEKIVAGGGNKKVTLTDEGALKYDGGKWIGLDDDDGYA VQTDKPVTGWSTNGGYLYFDQGLIVCTEDYIGYVKKHG ECKGDSYGMWVKVLPADDDKDDDKDDDKDDDKDYDD DNDHGDGDYYCSITGTYAISKSGSKHQYEAIKKVD AHP HVFSVGGDQGNDLIVTFQKDCSLVDQDNRGVYVDPNSG EVGNVDPWGELTPSVKWDIDDGYLIFNGESNFRSCPSGN GYSLSIKDCVGGTDIGLKVWEKGGGSLVKRAPARSPSPS TQPWEHVNAIQEALRLNLSRDTAAEMNETVEVISEMFD LQEPTCLQTRLELYKQGLRGSLTKLKGPLTMMASHYKQ HCPPTPETSCATQIITFESFKENLKDFLLVIPFDCWEPVQE
111	Kex2 linker amino acid sequence	GGGSLVKR

While the present invention is described herein with reference to illustrated
embodiments, it should be understood that the invention is not limited hereto. Those having
ordinary skill in the art and access to the teachings herein will recognize additional modifications
and embodiments within the scope thereof. Therefore, the present invention is limited only by
the claims attached herein.

WHAT IS CLAIMED:

1. A method for producing a heterologous glycoprotein, comprising;
 - (a) providing a host cell that includes at least one nucleic acid molecule
 5 encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein, and wherein the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex are expressed; and
 - (b) culturing the host cell under conditions for expressing the heterologous
 10 glycoprotein to produce the heterologous glycoprotein.
2. The method of claim 1, wherein the single-subunit
 oligosaccharyltransferase is the *Leishmania sp.r* STT3A protein, STT3B protein, STT3C protein,
 STT3D protein or combinations thereof.
3. The method of claim 1, wherein the single-subunit
 15 oligosaccharyltransferase is the *Leishmania major* STT3D protein.
4. The method of claim 1, wherein at least 70% of the heterologous
 glycoproteins produced by the host cell have fully occupied *N*-glycosylation sites.
 20
5. The method of claim 1, wherein the host cell is genetically engineered to
 produce glycoproteins comprising one or more mammalian- or human-like *N*-glycans.
6. The method of claim 1, wherein the heterologous protein is erythropoietin
 25 (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and
 granulocyte-colony stimulating factor (GCSF); granulocyte macrophage-colony stimulating
 factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C;
 antithrombin III; thrombin; soluble IgE receptor α -chain; immunoglobulins such as IgG, IgG
 fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble
 30 TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase;
 urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing
 factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid
 progenitor inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle
 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating
 35 hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and
 calcium modulator and cyclophilin ligand; glucagon like protein 1; or IL-2 receptor agonist.

7. The method of claim 1, wherein the heterologous protein is an anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, or anti-CD20 antibody.

8. The method of claim 1, wherein the host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, plant cells, insect cells, and mammalian cells.

9. The method of claim 1, wherein the host cell is an *och1* mutant of *P. pastoris*.

10. A method for producing a glycoprotein composition in which at least 70% of the *N*-glycosylation sites on the glycoproteins in the composition are occupied with an *N*-glycan, comprising;

(a) providing a recombinant host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the glycoprotein, and wherein the host cell genes encoding the endogenous OTase complex are expressed; and

(b) culturing the recombinant host cell under conditions for expressing the glycoprotein to produce the compositions in which at least 70% of the *N*-glycosylation sites on the glycoproteins in the composition are occupied with the *N*-glycan.

11. The method of claim 10, wherein the single-subunit oligosaccharyltransferase is the *Leishmania* sp. STT3A protein, STT3B protein, STT3C protein, STT3D protein or combinations thereof.

12. The method of claim 10, wherein the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein.

13. The method of claim 10, wherein the heterologous protein is erythropoietin (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and granulocyte-colony stimulating factor (GCSF); granulocyte macrophage-colony stimulating factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin; soluble IgE receptor α -chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; or IL-2 receptor agonist.

14. The method of claim 10, wherein the heterologous protein is anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, or anti-CD20 antibody.

15. The method of claim 10, wherein the recombinant host cell is a yeast or filamentous fungus host cell.

16. A host cell, comprising:

- (a) a first nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase oligosaccharyltransferase; and
- (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the host cell expresses the genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex.

17. The method of claim 16, wherein the single-subunit oligosaccharyltransferase is the *Leishmania* sp. STT3A protein, STT3B protein, STT3C protein, or STT3D protein.

18. The method of claim 16, wherein the single-subunit oligosaccharyltransferase is the *Leishmania major* STT3D protein.

19. The host cell of claim 16, wherein the host cell is genetically engineered to produce glycoproteins comprising one or more mammalian- or human-like *N*-glycans.

20. The host cell of claim 16, wherein the heterologous glycoprotein is
 5 erythropoietin (EPO); cytokines such as interferon α , interferon β , interferon γ , and interferon ω ; and granulocyte-colony stimulating factor (GCSF); granulocyte macrophage-colony stimulating factor (GM-CSF); coagulation factors such as factor VIII, factor IX, and human protein C; antithrombin III; thrombin; soluble IgE receptor α -chain; immunoglobulins such as IgG, IgG fragments, IgG fusions, and IgM; immunoadhesions and other Fc fusion proteins such as soluble
 10 TNF receptor-Fc fusion proteins; RAGE-Fc fusion proteins; interleukins; urokinase; chymase; urea trypsin inhibitor; IGF-binding protein; epidermal growth factor; growth hormone-releasing factor; annexin V fusion protein; angiostatin; vascular endothelial growth factor-2; myeloid progenitor inhibitory factor-1; osteoprotegerin; α -1-antitrypsin; α -feto proteins; DNase II; kringle 3 of human plasminogen; glucocerebrosidase; TNF binding protein 1; follicle stimulating
 15 hormone; cytotoxic T lymphocyte associated antigen 4 – Ig; transmembrane activator and calcium modulator and cyclophilin ligand; glucagon like protein 1; or IL-2 receptor agonist.

21. The host cell of claim 16, wherein the heterologous glycoprotein is an
 anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-
 20 VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, or anti-CD20 antibody.

22. The host cell of claim 16, wherein the host cell is selected from the group
 25 consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus*
 30 *niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, plant cells, insect cells, and mammalian cells.

23. A glycoprotein composition comprising:
 35 a plurality of antibodies wherein at least 70% of the antibody molecules in the composition have both *N*-glycosylation sites occupied and about 50 to 70 mole% of the *N*-

glycans are G0, 15-25 mole% of the *N*-glycans are G1, and about 5 to 15 mole% of the *N*-glycans comprise a Man₅GlcNAc₂ core structure and a pharmaceutically acceptable carrier.

24. The composition of claim 23, wherein the antibodies comprise an antibody selected from the group consisting of anti-Her2 antibody, anti-RSV (respiratory syncytial virus) antibody, anti-TNF α antibody, anti-VEGF antibody, anti-CD3 receptor antibody, anti-CD41 7E3 antibody, anti-CD25 antibody, anti-CD52 antibody, anti-CD33 antibody, anti-IgE antibody, anti-CD11a antibody, anti-EGF receptor antibody, and anti-CD20 antibody.

25. A recombinant *Pichia pastoris* host cell that has a genotype that is the same as the genotype of recombinant *Pichia pastoris* strain YGLY14401.

26. Use of the recombinant *Pichia pastoris* to produce an antibody composition comprising anti-RSV antibodies that have the *N*- and *O*-glycosylation pattern

Man5	About 9.5 %
G0	About 59.9 %
G0F	0 %
G1	About 20 %
G1F	0 %
G2	About 2.8 %
G2F	0 %
A2	0 %
Hybrid	About 7.8 %
O-glycans occupancy (mol/mol)	About 3.0
Mannose (single mannose)	About 96 %
Mannobiose (two mannose residues)	About 4 %

and a pharmaceutically acceptable carrier.

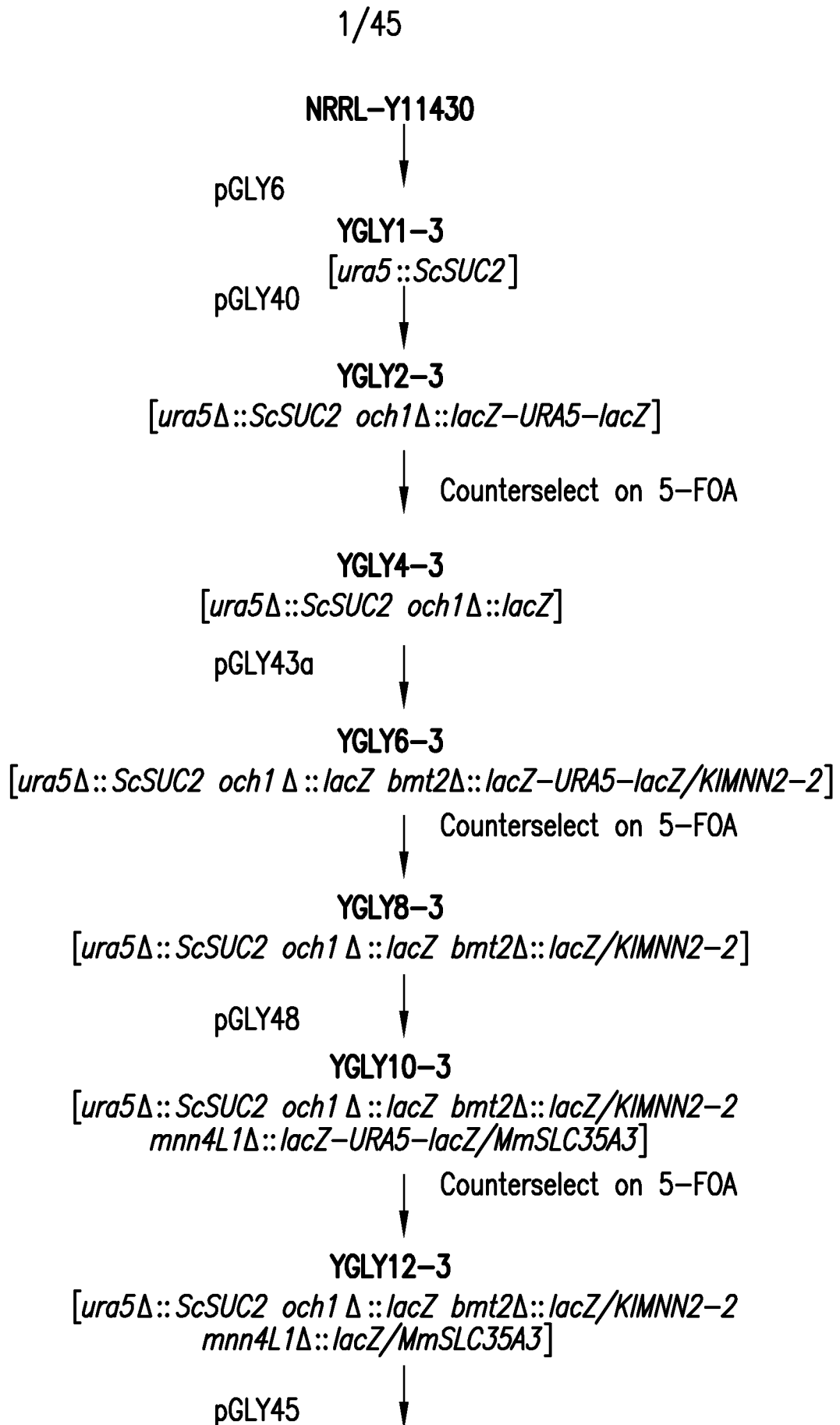


FIG. 1A

2/45

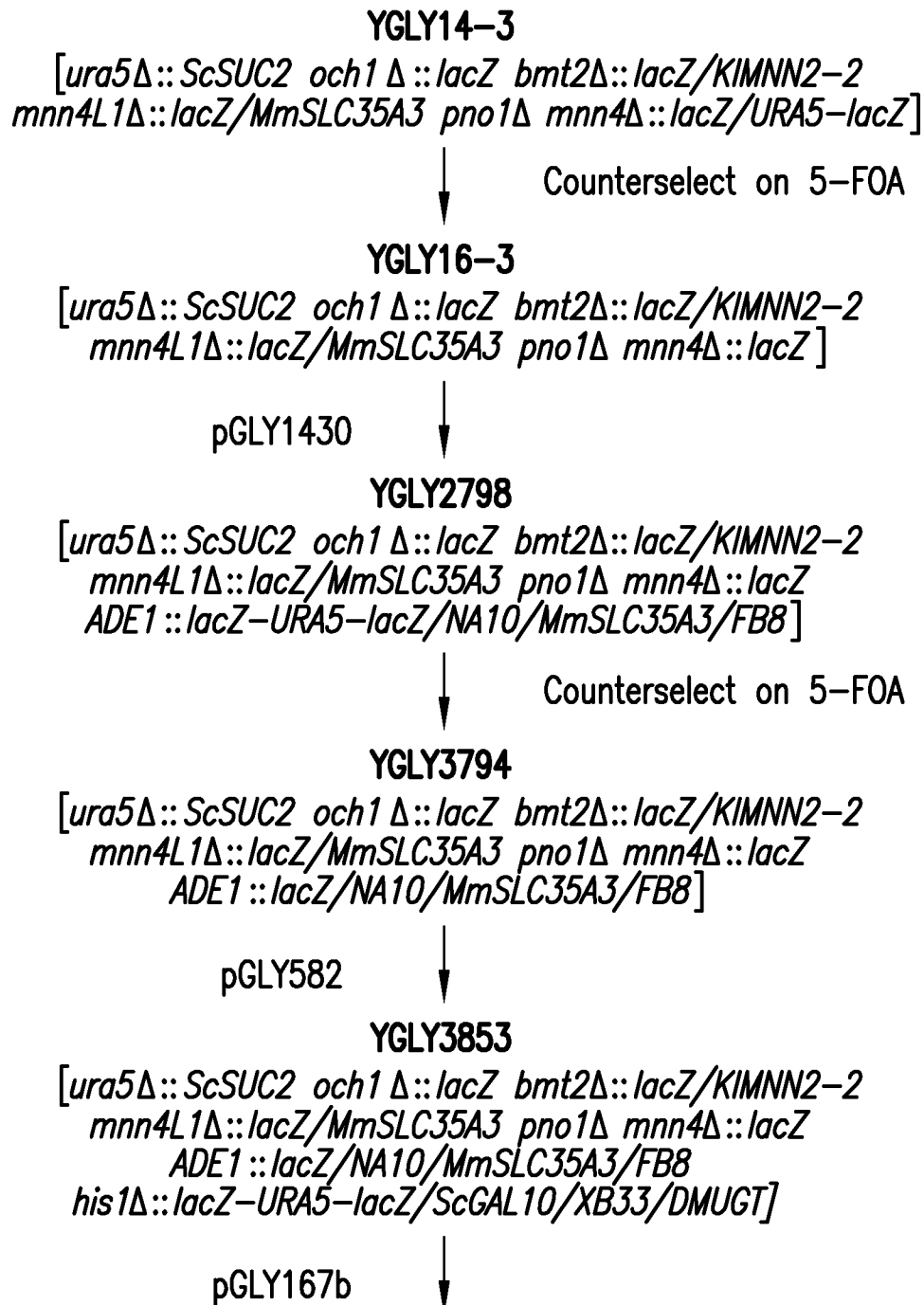


FIG. 1 B

3/45

YGLY4754

[*ura5*Δ::*ScSUC2 och1* Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1 ::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ-URA5-lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*]



Counterselect on 5-FOA

YGLY4799

[*ura5*Δ::*ScSUC2 och1* Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1 ::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*]

pGLY3411

**YGLY6903**

[*ura5*Δ::*ScSUC2 och1* Δ::*lacZ bmt2*Δ::*lacZ/KIMNN2-2*
*mnn4L1*Δ::*lacZ/MmSLC35A3 pno1*Δ *mnn4*Δ::*lacZ*
ADE1 ::*lacZ/NA10/MmSLC35A3/FB8*
*his1*Δ::*lacZ/ScGAL10/XB33/DmUGT*
*arg1*Δ::*HIS1/KD53/TC54*
*bmt4*Δ::*lacZ-URA5-lacZ*]



Counterselect on 5-FOA

FIG. 1C

4/45

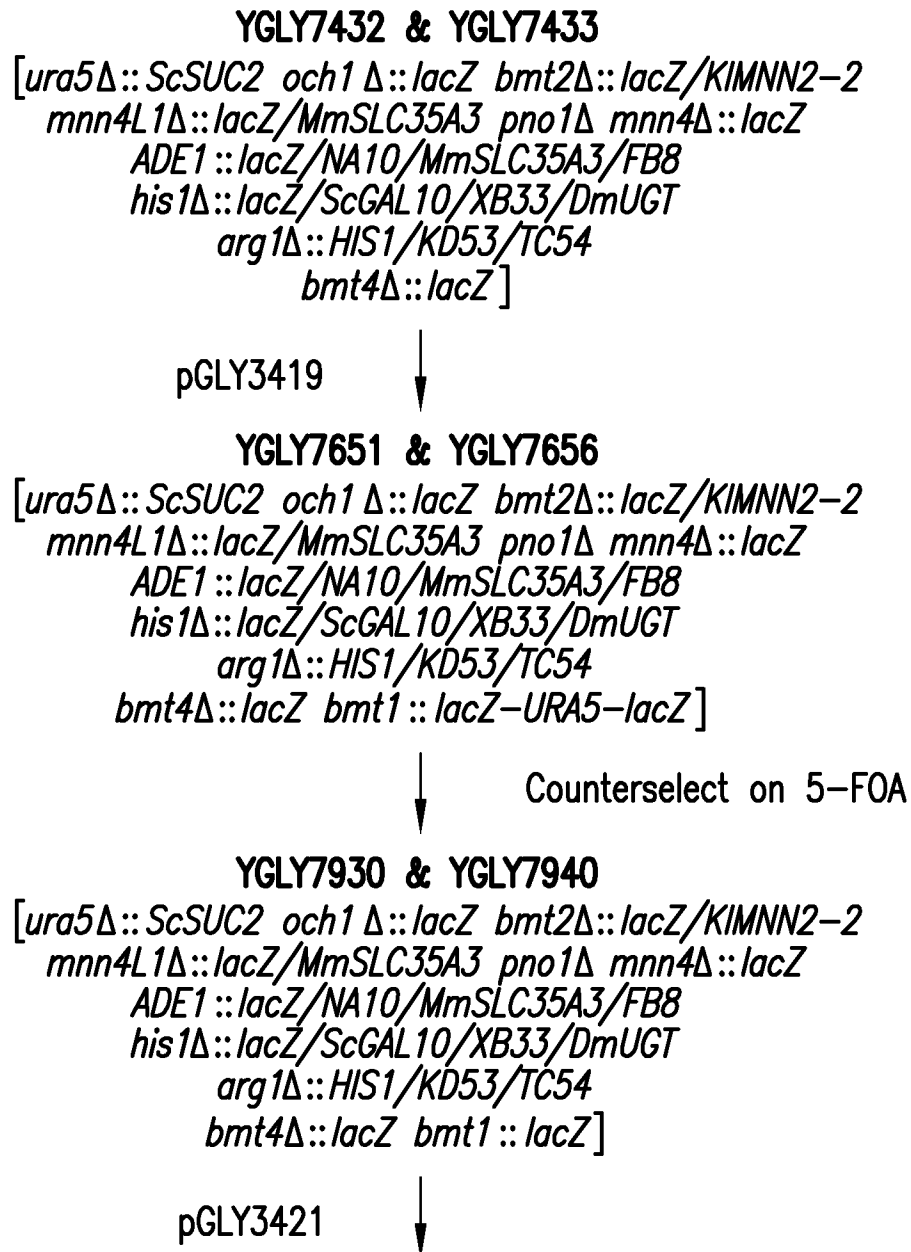


FIG. 1D

5/45

YGLY7961 & YGLY7965

[*ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ URA5-lacZ]

pGLY3673

**YGLY8316 & YGLY8323**

[*ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ URA5-lacZ
PRO1::ARG1/AOX1p-TrMDS1]

FIG. 1E

6/45

YGLY8316

[ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ URA5-lacZ]
PRO1::ARG1/AOX1p-TrMDS1]

pGLY6833

**YGLY13992**

[ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ-URA5-lacZ
PRO1::ARG1/AOX1p-TrMDS1]
TRP2::Sh ble/AOX1p-anti-Her2-Ab

FIG. 1F

7/45

YGLY8323

[ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ-URA5-lacZ
PRO1::ARG1/AOX1p-TrMDS1]

pGLY6564

**YGLY14401**

[ura5Δ::ScSUC2 och1 Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1::lacZ bmt3::lacZ URA5-lacZ
PRO1::ARG1/AOX1p-TrMDS1]
TRP2::Sh ble/AOX1p-anti-RSV-Ab

FIG. 1 G

Glossary:

ScSUC2	<i>S. cerevisiae</i> Invertase
OCH1	Alpha-1,6-mannosyltransferase
KIMNN2-2:	<i>K. lactis</i> UDP-GlcNAc transporter
BMT1:	Beta-mannose-transfer {beta-mannose elimination}
BMT2:	Beta-mannose-transfer {beta-mannose elimination}
BMT3:	Beta-mannose-transfer {beta-mannose elimination}
BMT4:	Beta-mannose-transfer {beta-mannose elimination}
MNN4L1:	MNN4-like 1 (charge elimination)
MmSLC35A3	Mouse homologue of UDP-GlcNAc Transporter
PN01:	Phosphomannosylation of N-glycans (charge elimination)
MNN4:	Mannosyltransferase (charge elimination)
ScGAL10	UDP-glucose 4-epimerase
XB33	Truncated HsGalT1 fused to ScKRE2 leader
DmUGT	UDP-Galactose transporter
KD53	Truncated DmMNSII fused to ScMNN2 leader
TC54	Truncated RnGNTII fused to ScMNN2 leader
NA10	Truncated HsGNTI fused to PpSEC12 leader
FB8:	Truncated MmMNS1A fused to ScSEC12 leader
TrMDS1:	Secreted <i>T. reesei</i> MNS1
Sh ble:	Zeocin resistance marker

FIG. 1H

9/45

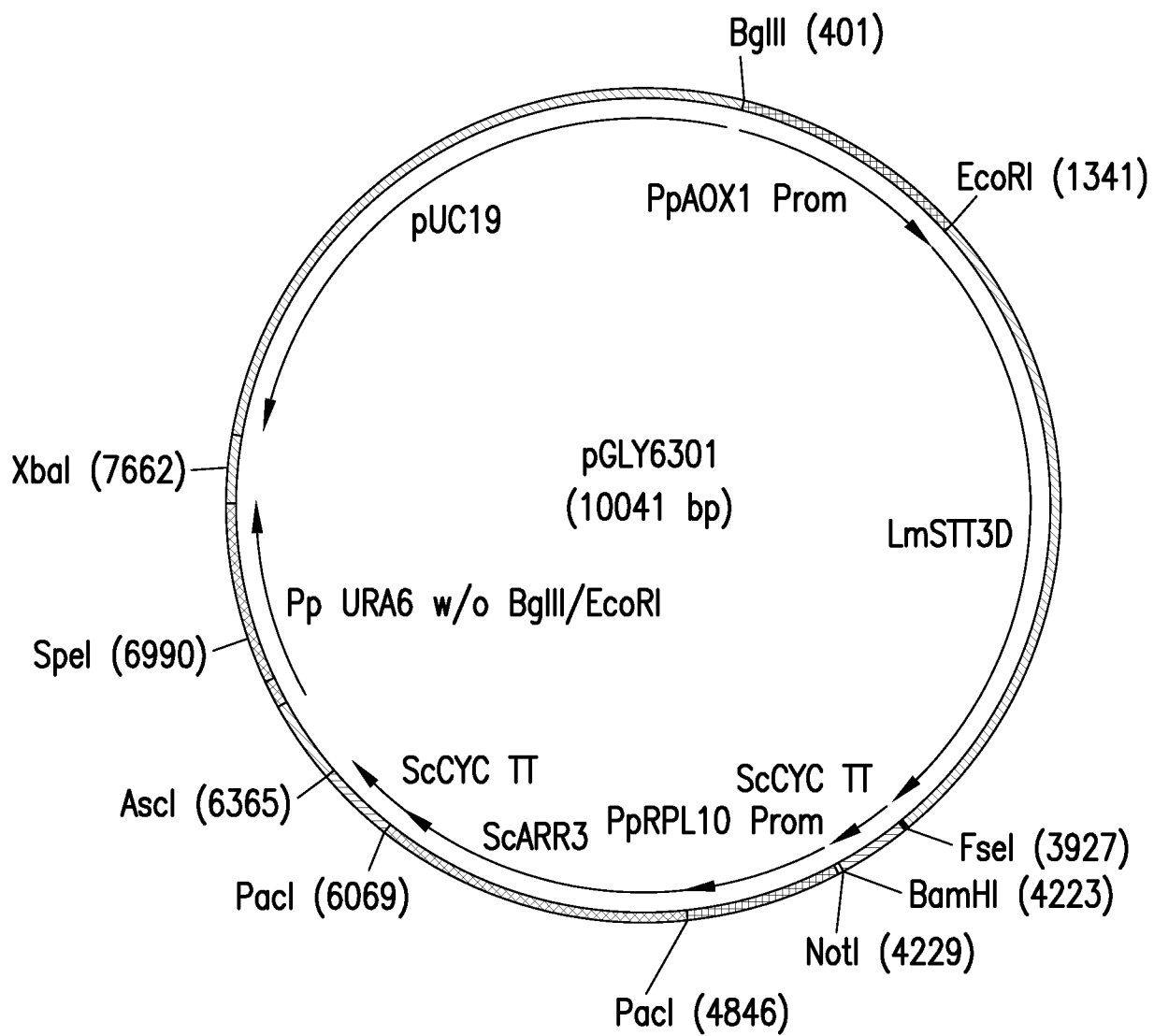


FIG.2

10/45

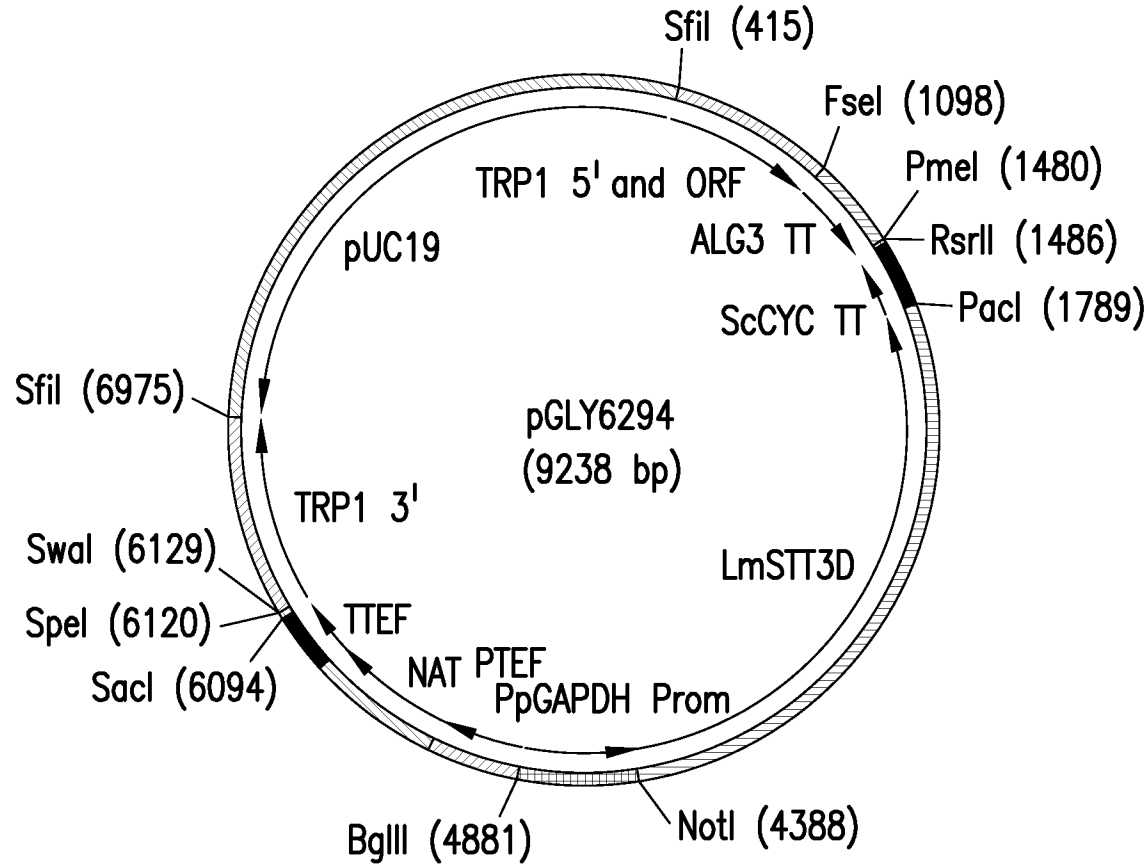


FIG.3

11/45

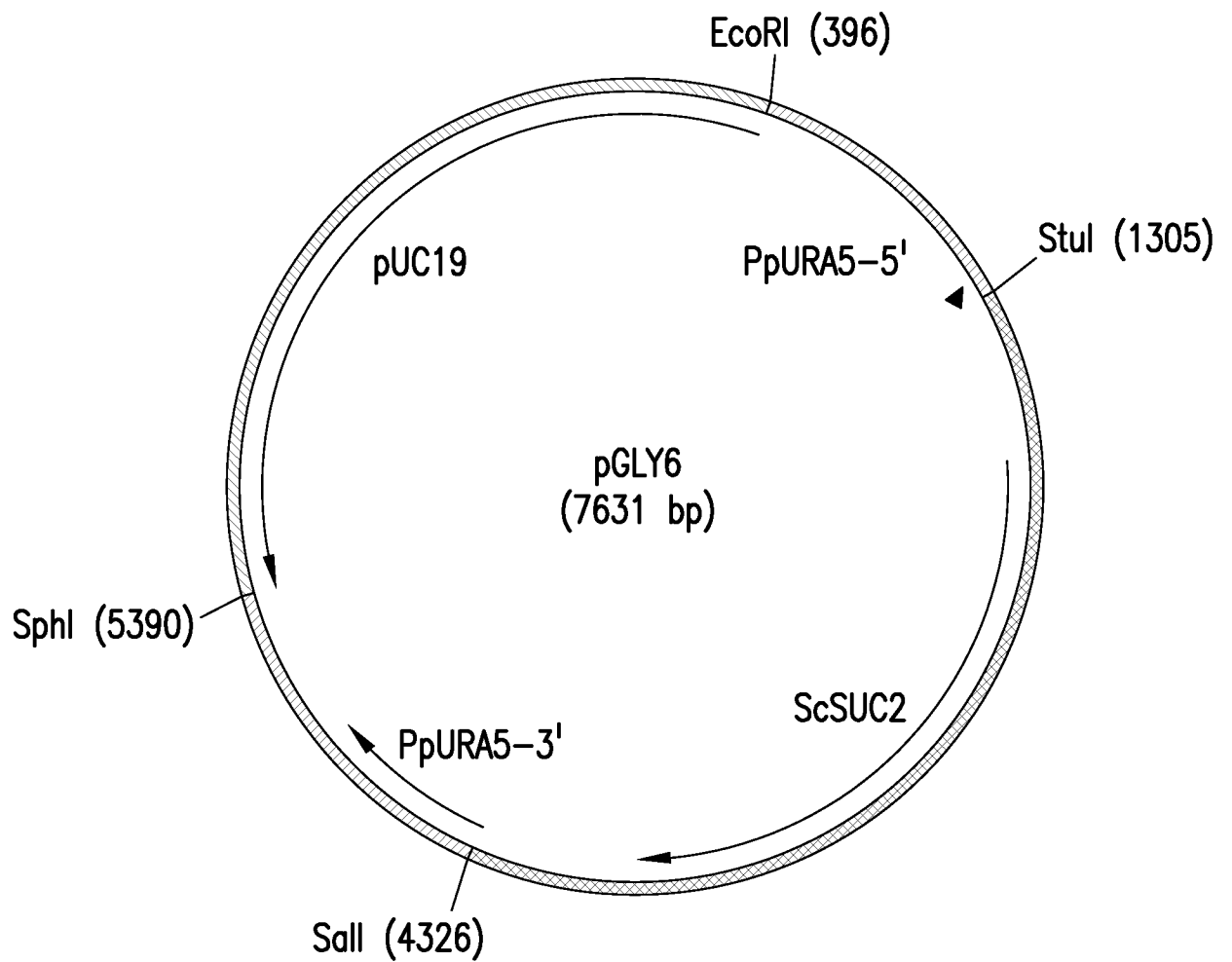


FIG.4

12/45

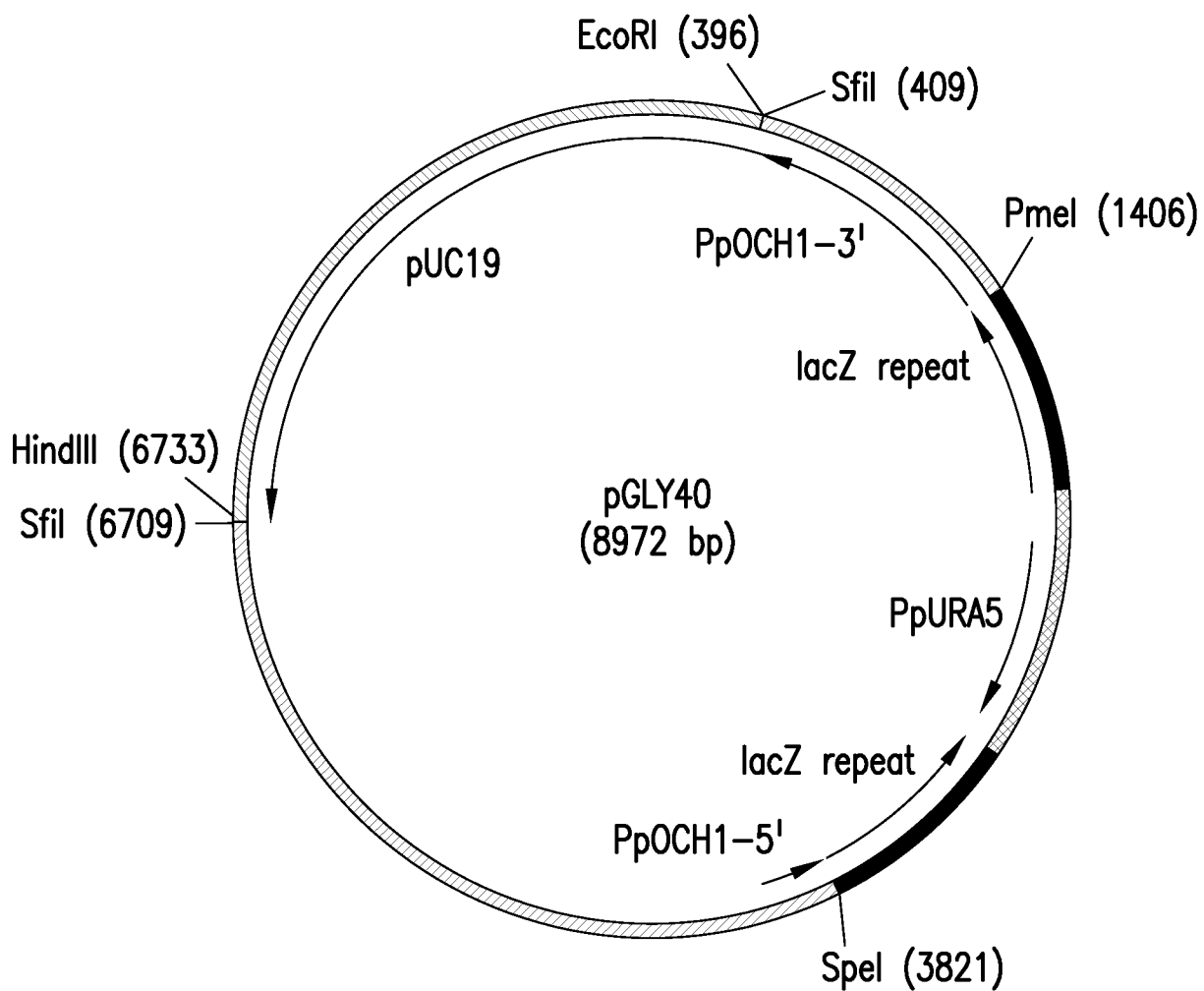


FIG.5

13/45

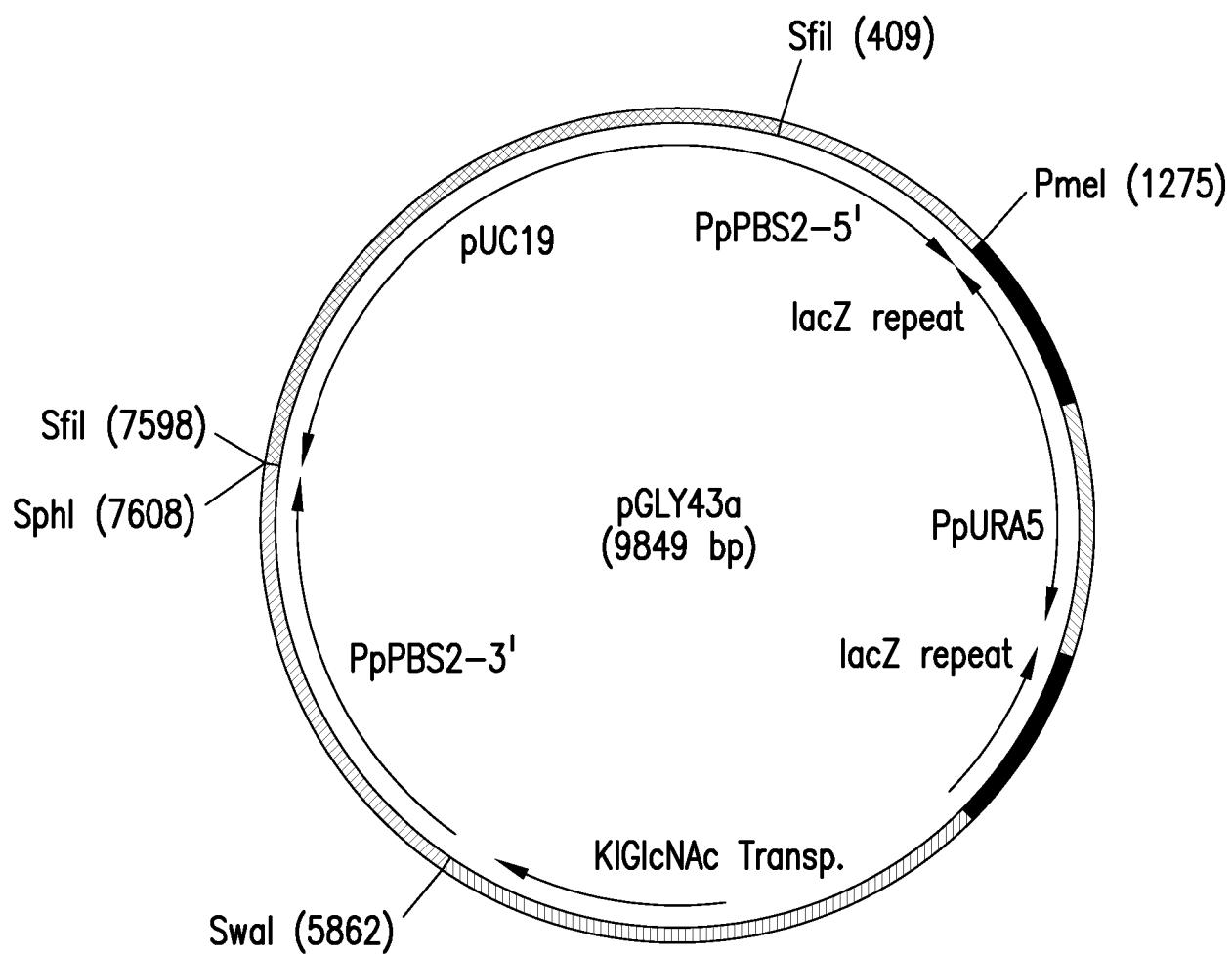


FIG.6

14/45

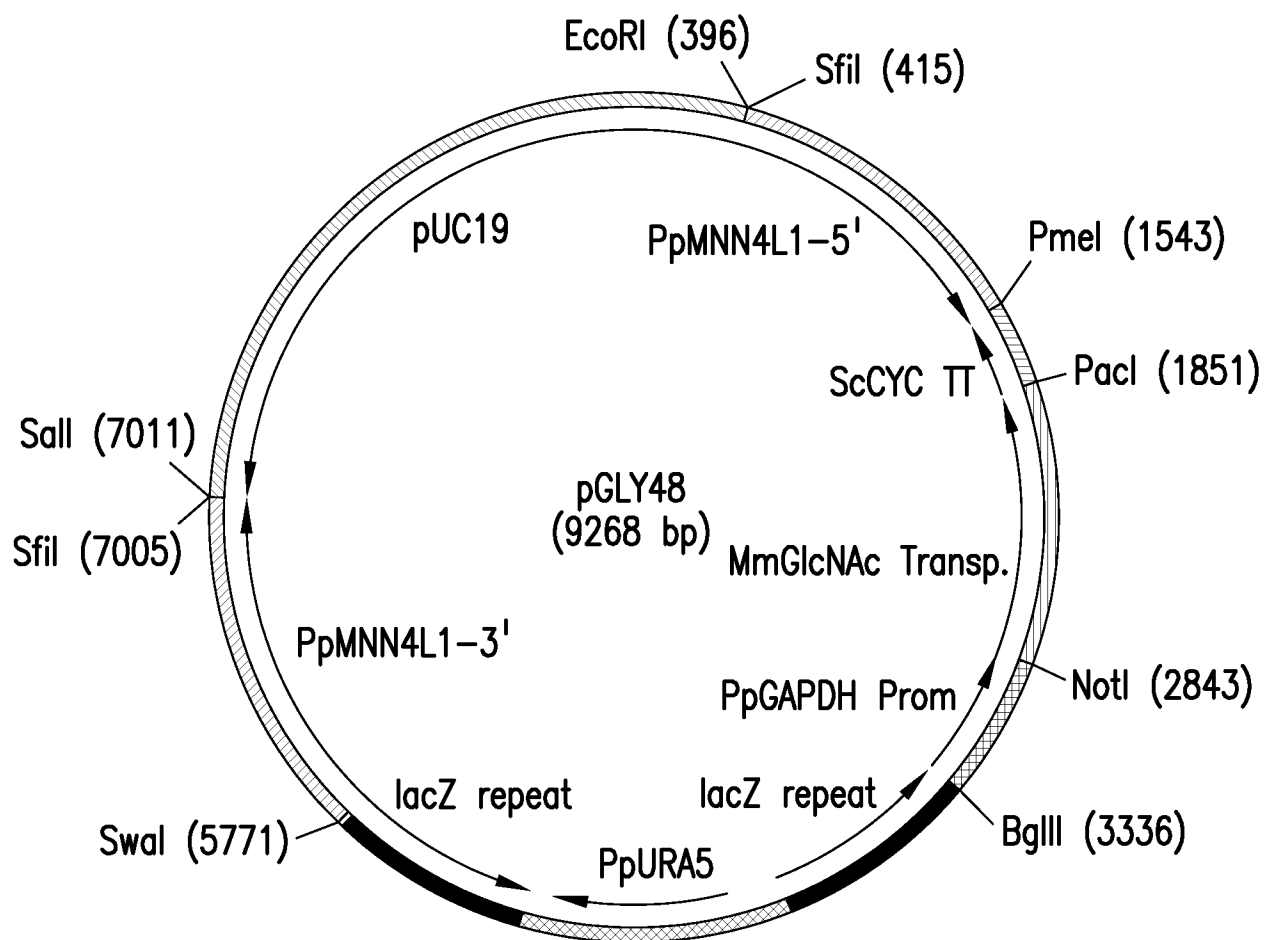


FIG.7

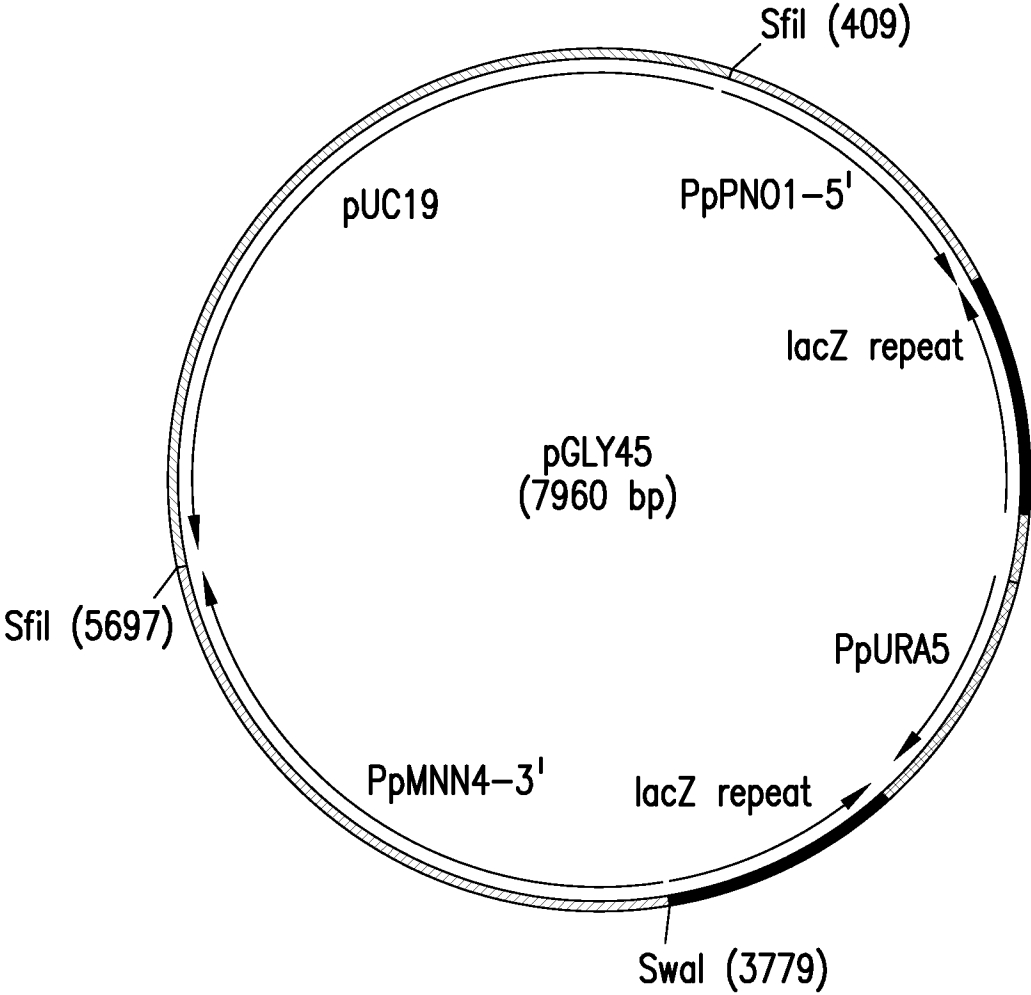


FIG.8

16/45

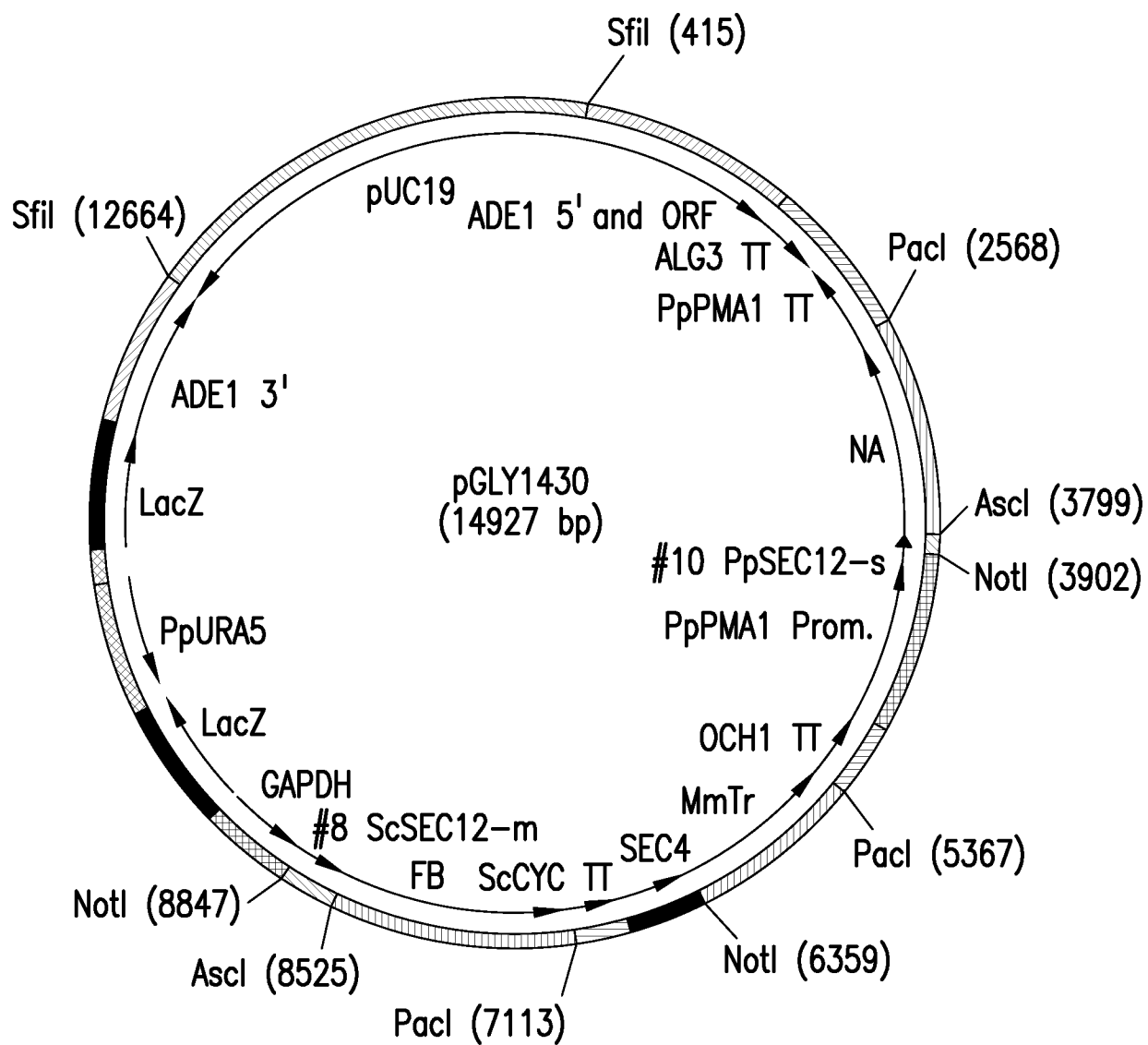


FIG.9

17/45

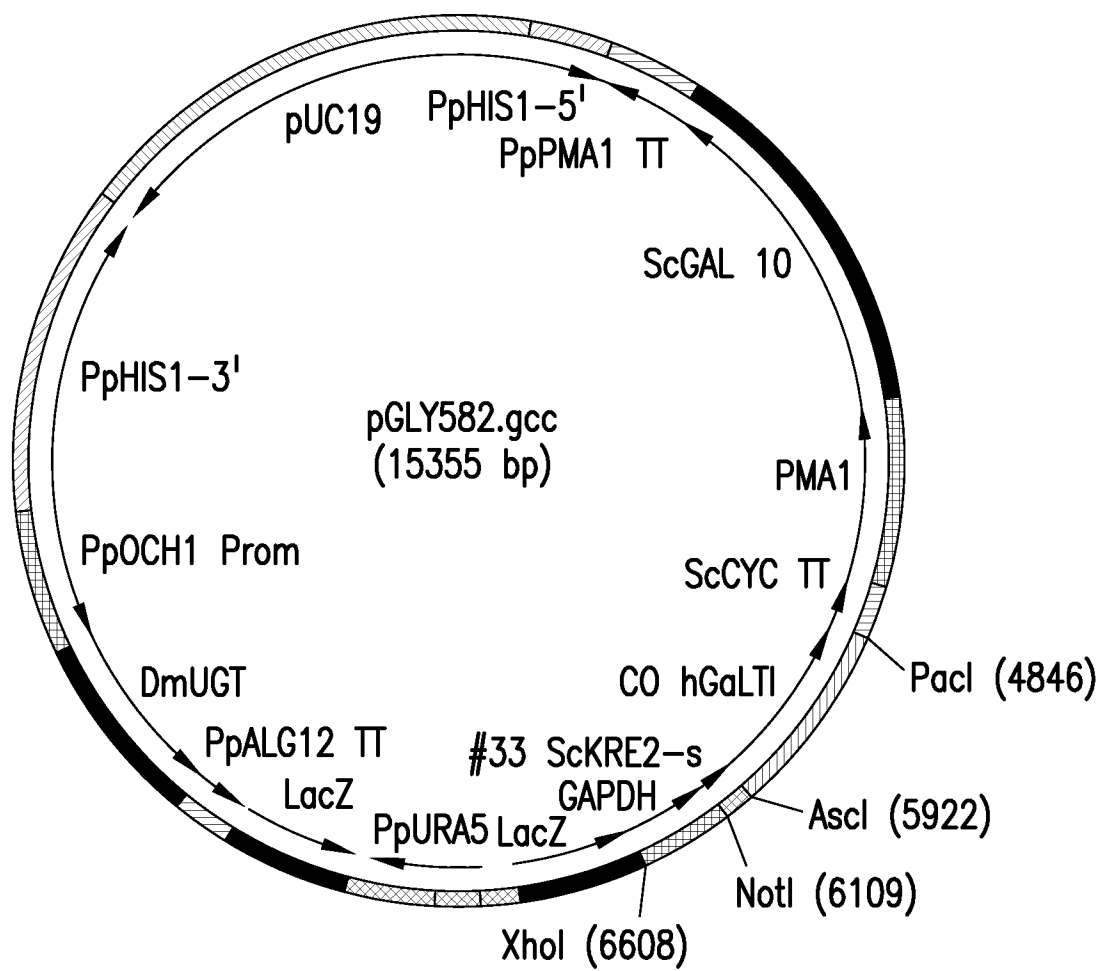


FIG.10

18/45

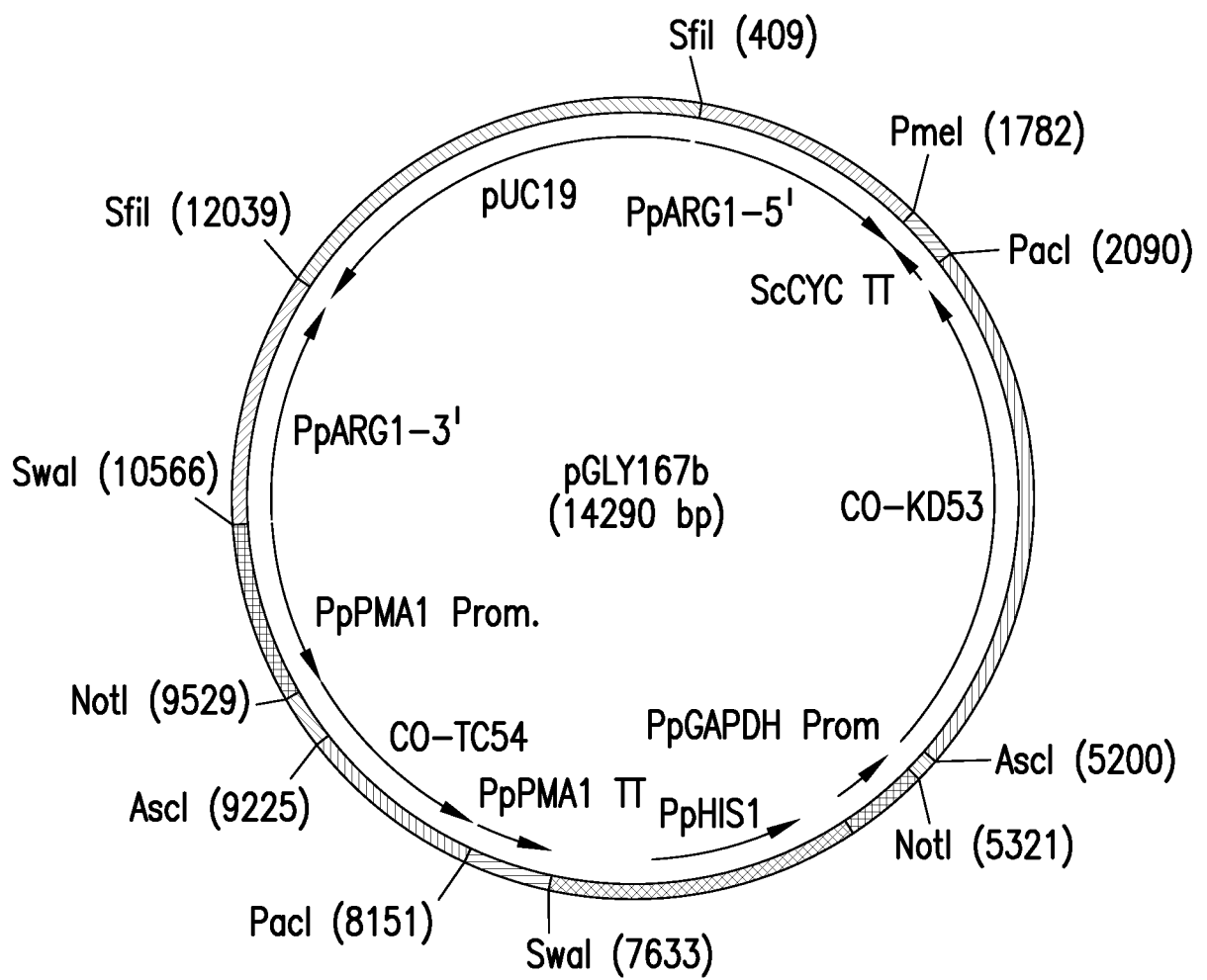


FIG. 11

19/45

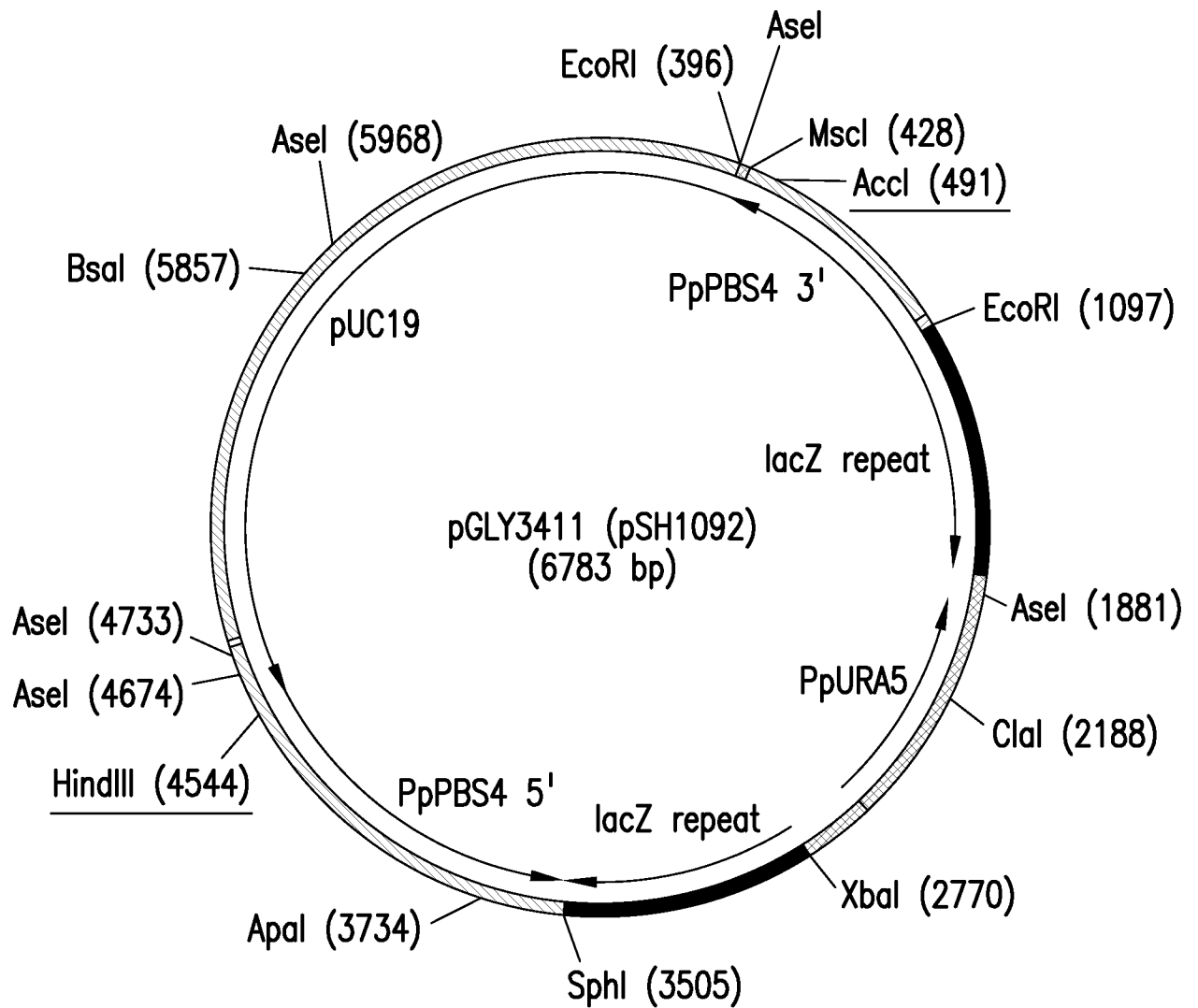


FIG.12

20/45

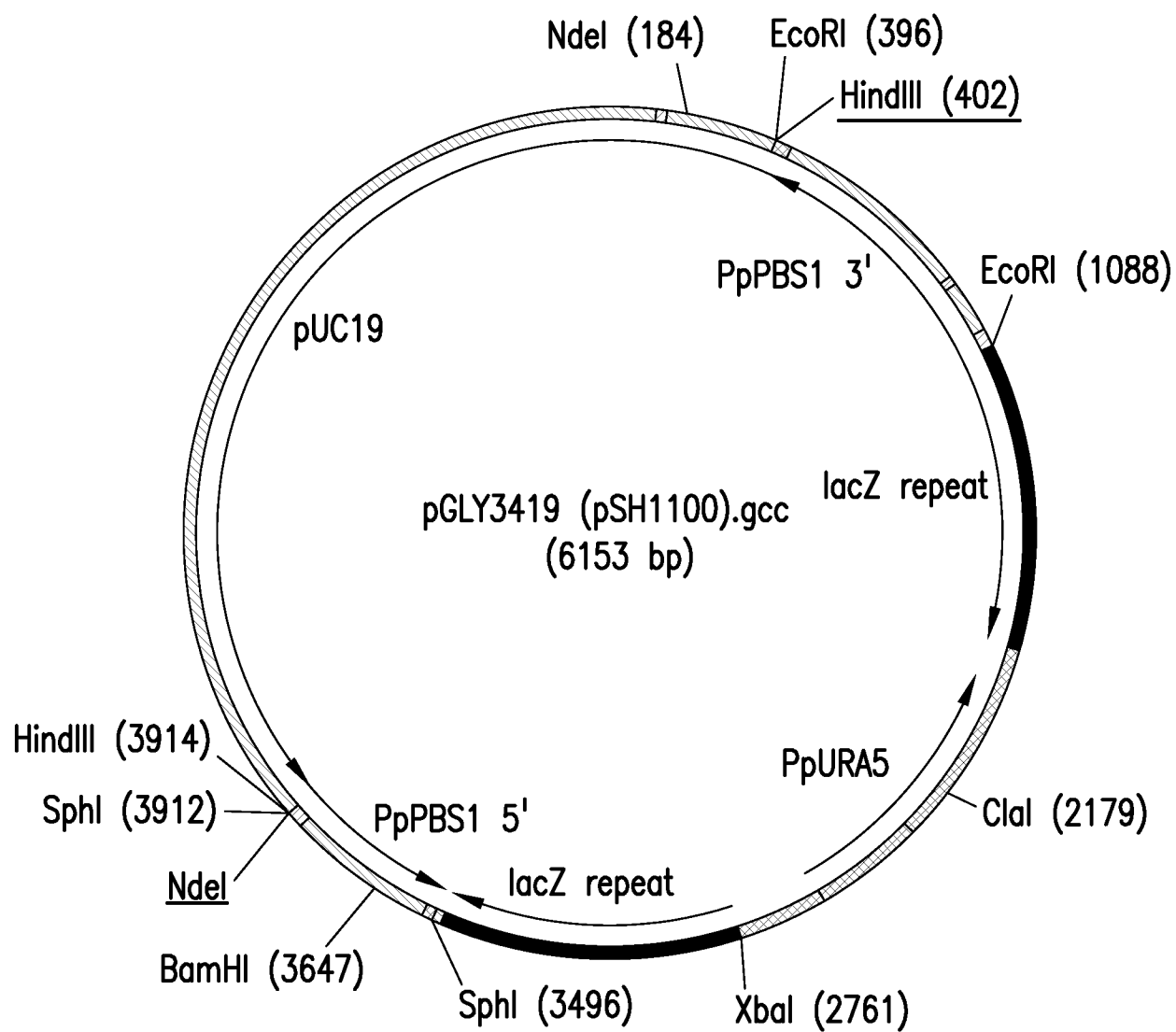


FIG.13

21/45

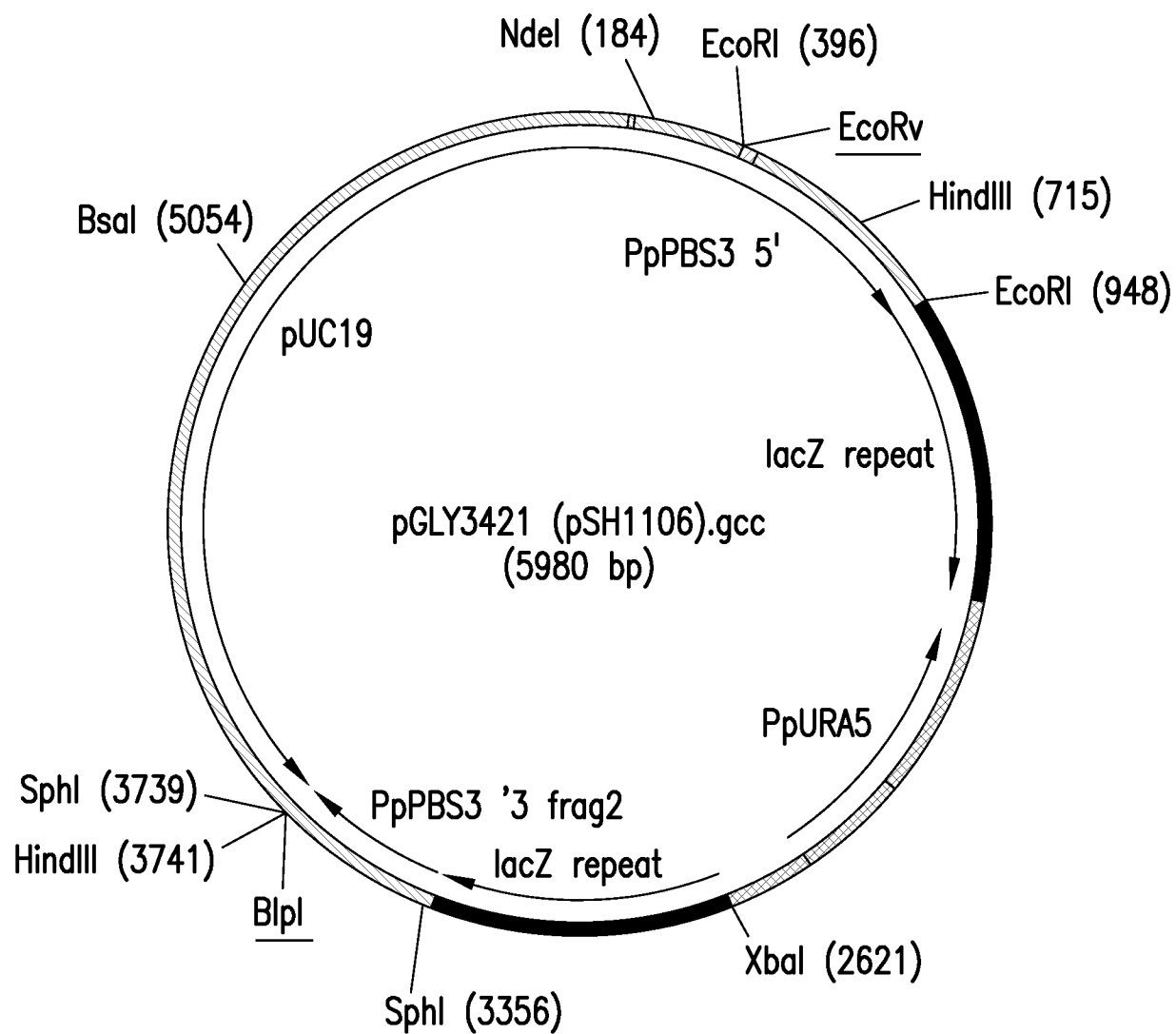


FIG.14

22/45

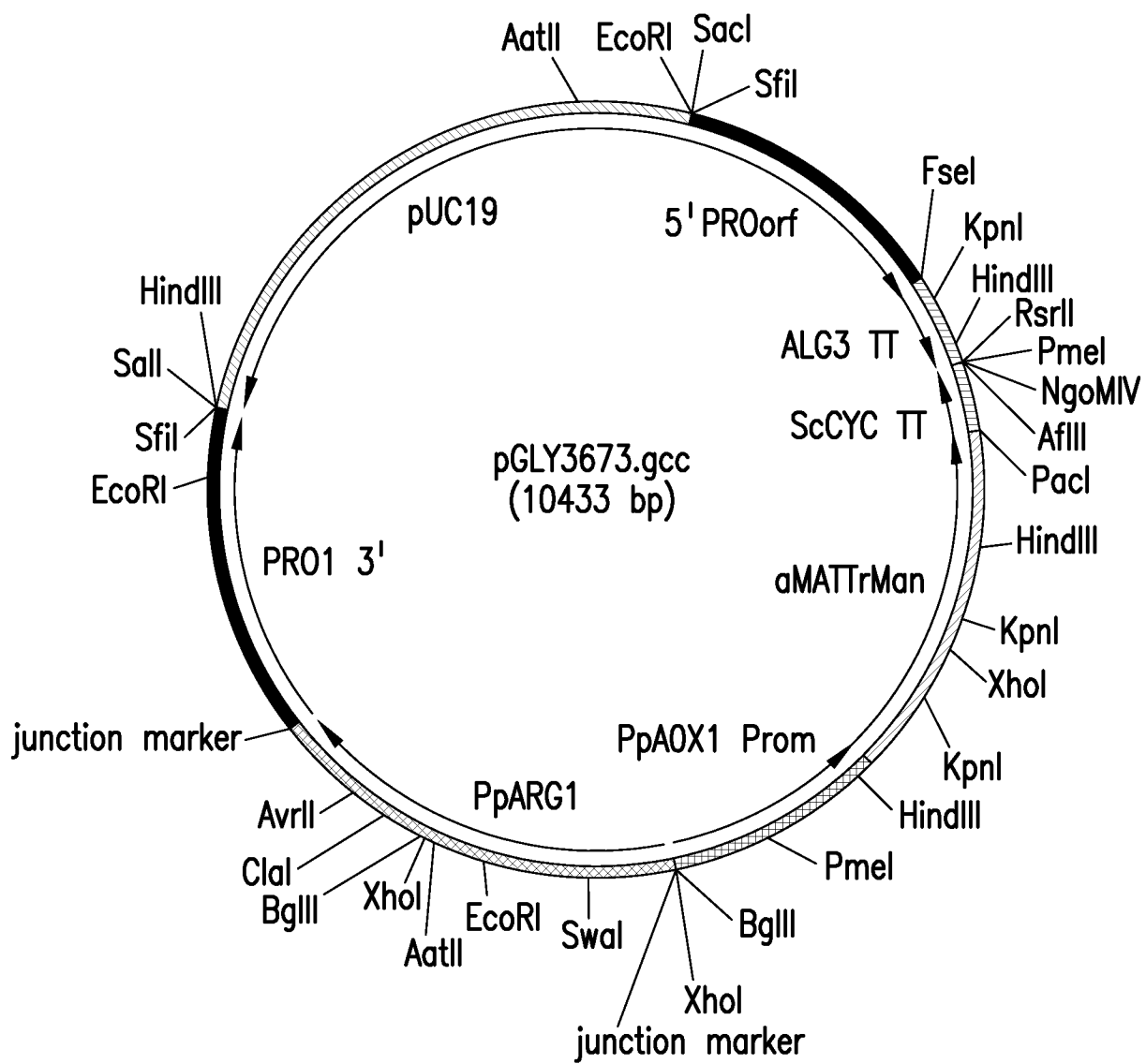


FIG.15

23/45

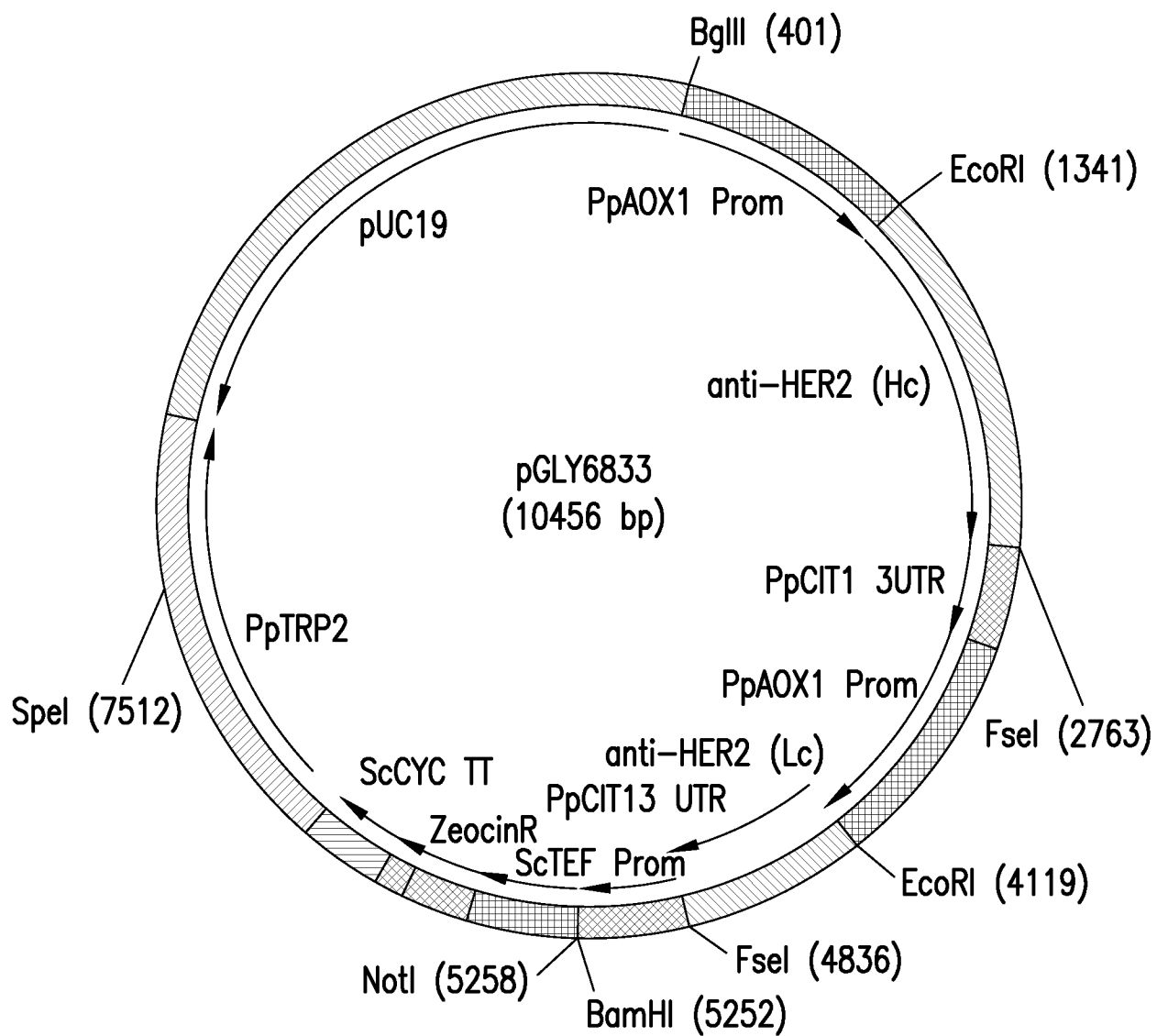


FIG.16

24/45

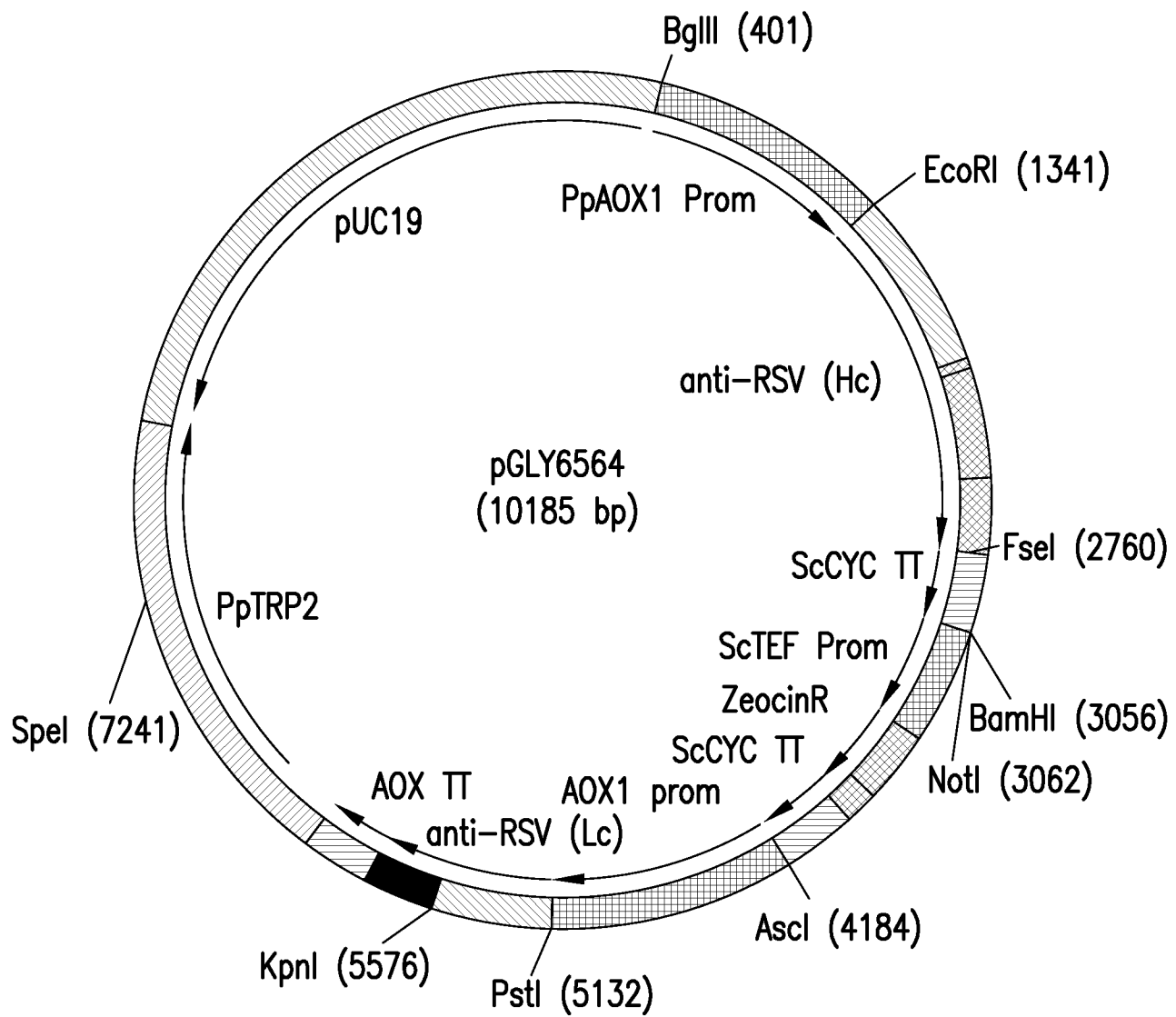


FIG.17

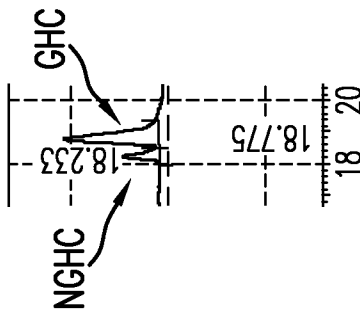
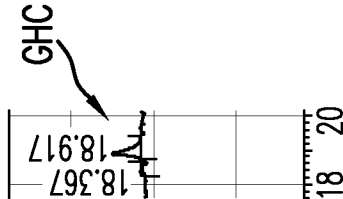
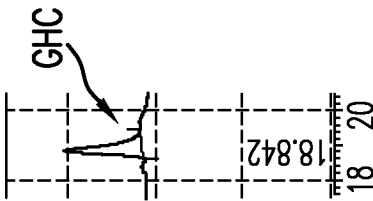
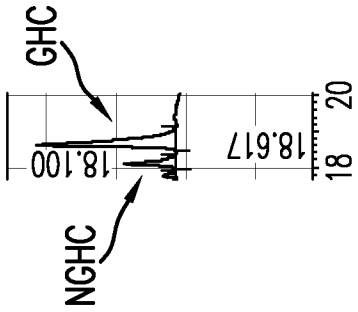
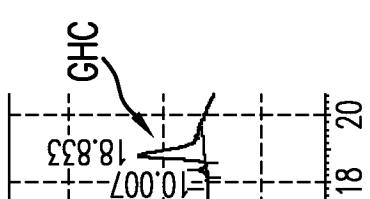
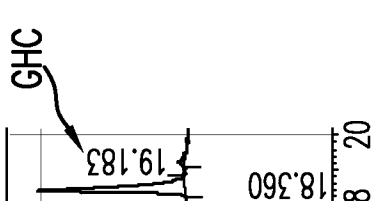
N-Glycan occupancy (CE)				
Ant-HER2	yGLY13992 (control)	yGLY17368 (GAPDH prom-driven LmSTT3D)	yGLY17351 (AOX1 prom-driven LmSTT3D)	
				
	80%	94%	99%	
		yGLY17354 (GAPDH prom-driven LmSTT3D)	yGLY17319 (AOX1 prom-driven LmSTT3D)	
Anti-RSV	yGLY14401 (control)			
	82%	94%	99%	

FIG.18

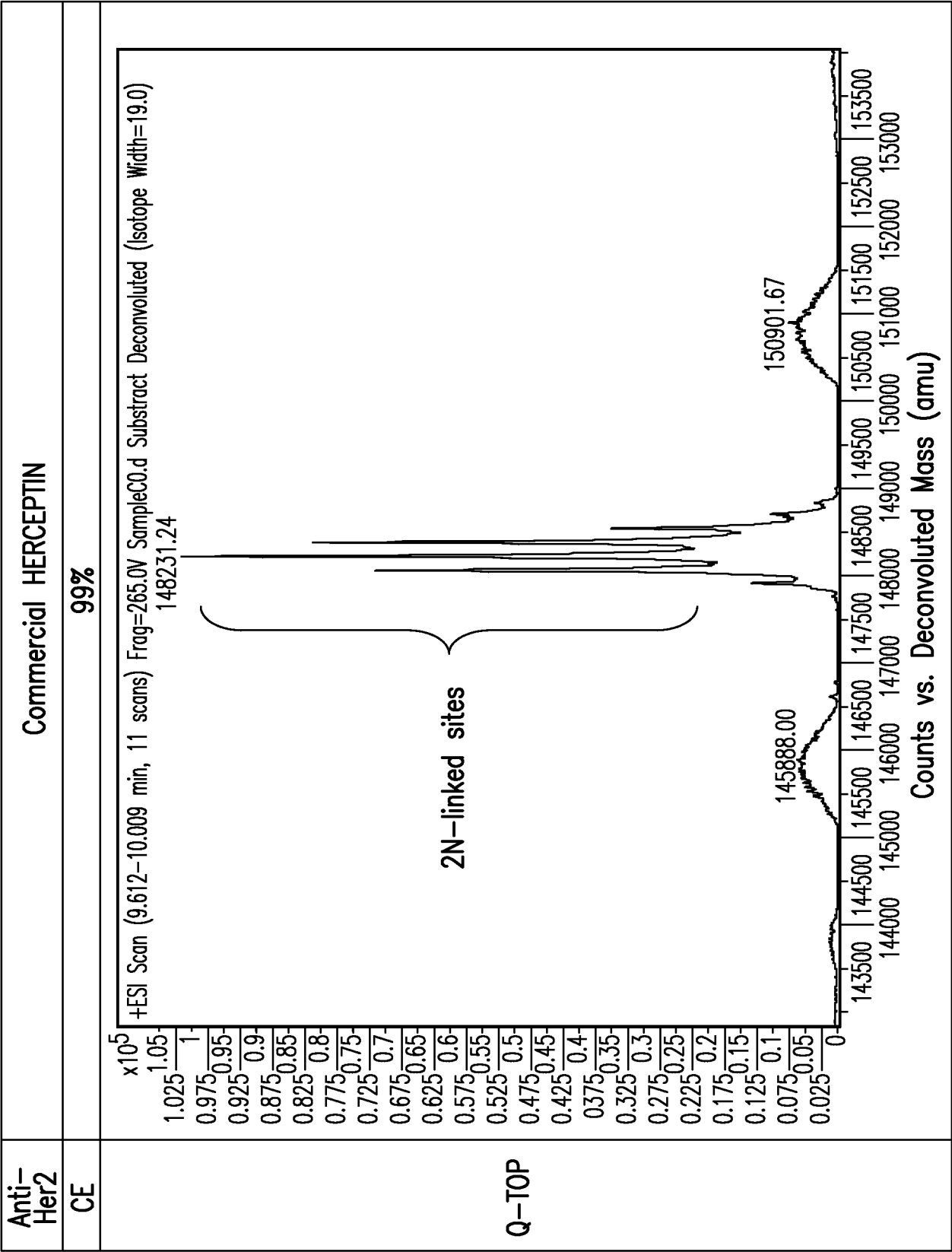


FIG.19A

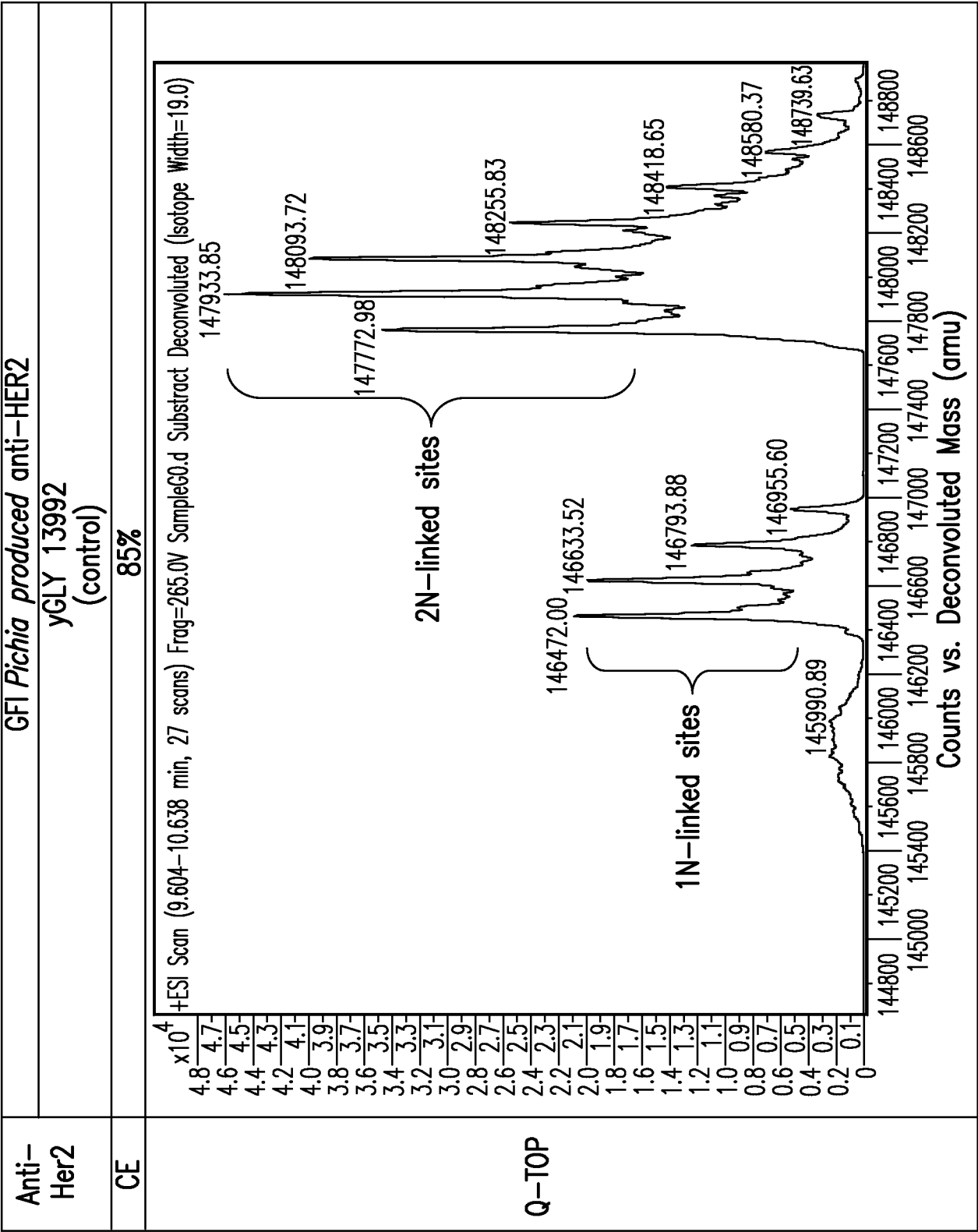


FIG.19B

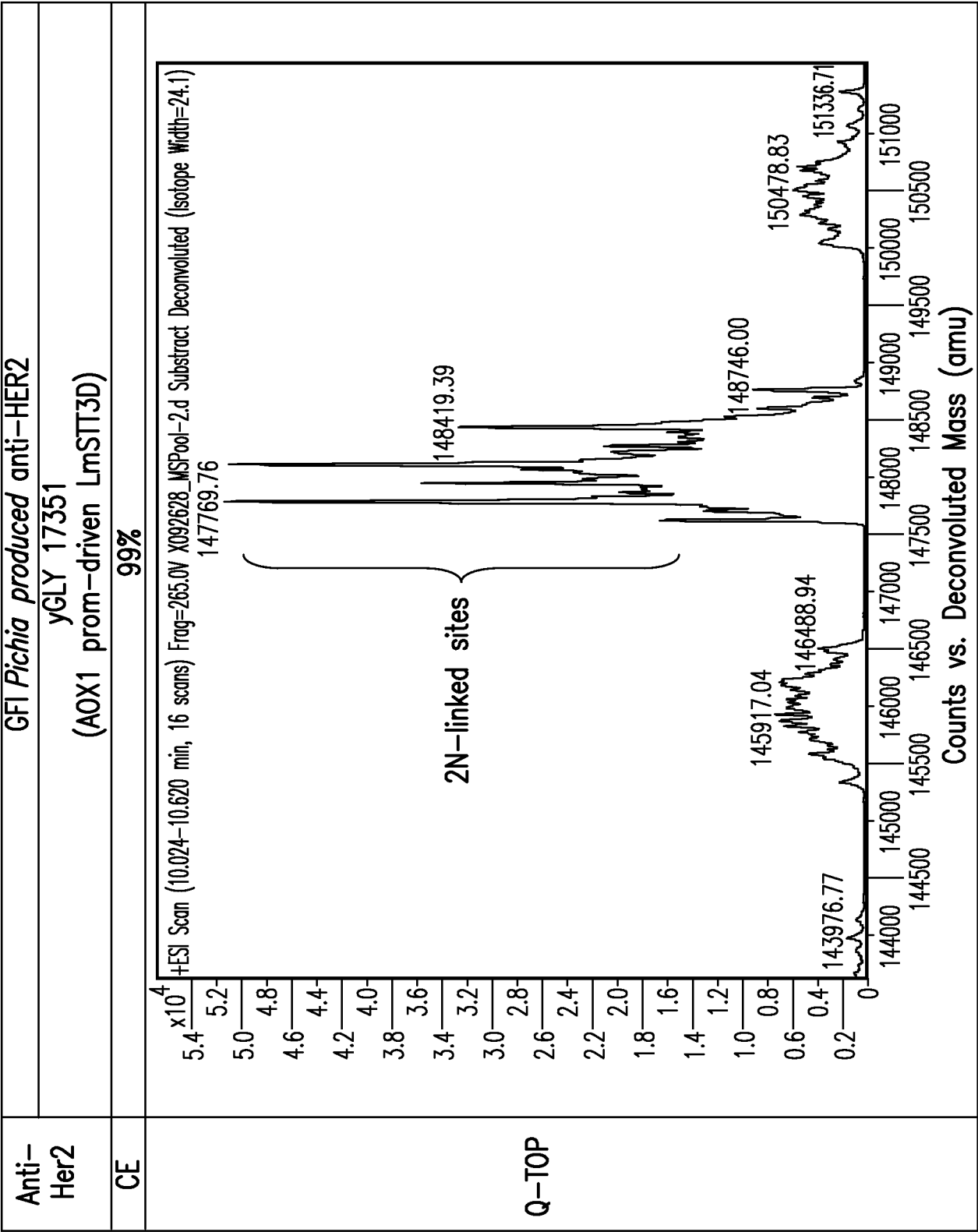
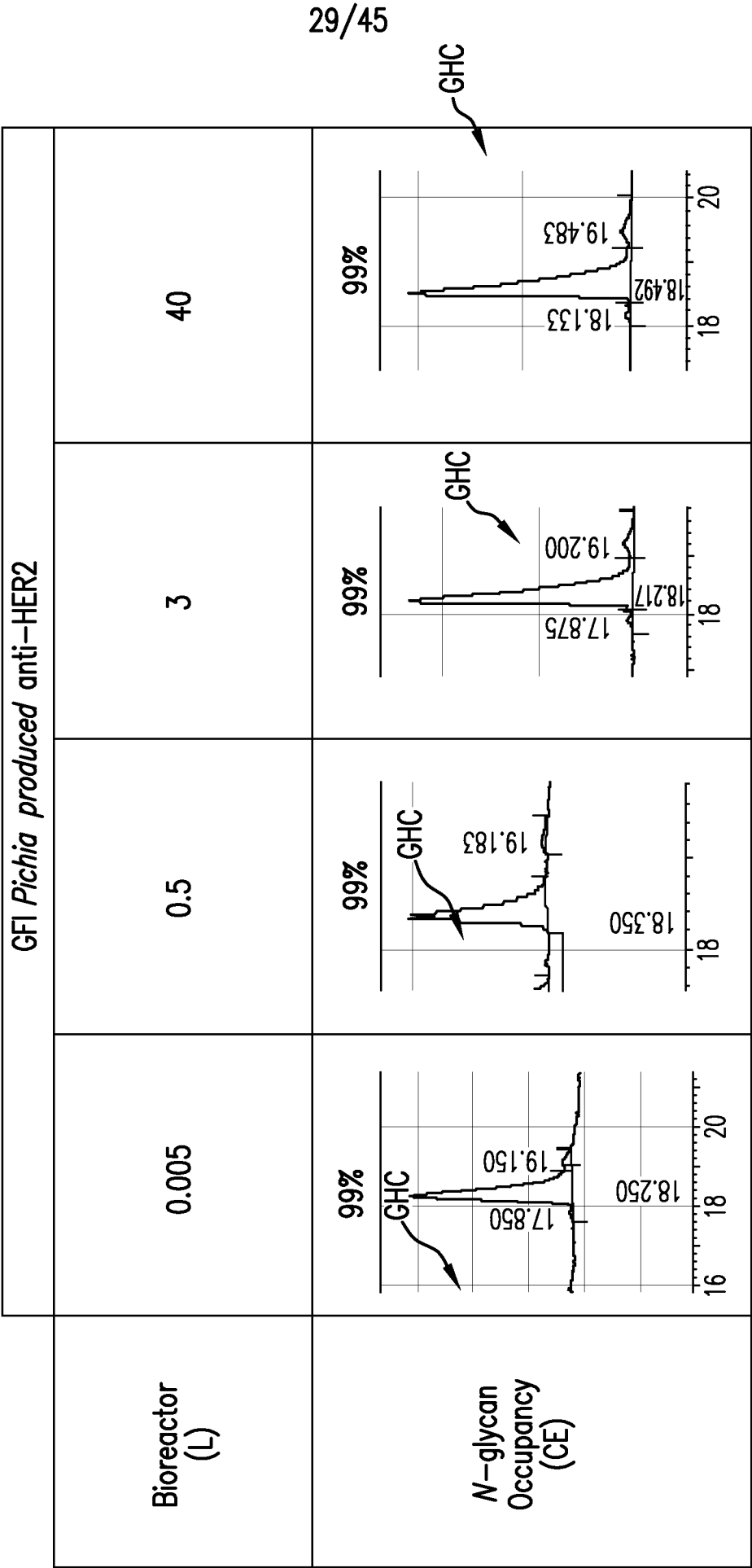


FIG.19C



29/45

FIG.20

30/45

GLN HERCEPTIN (Lot: F090742)

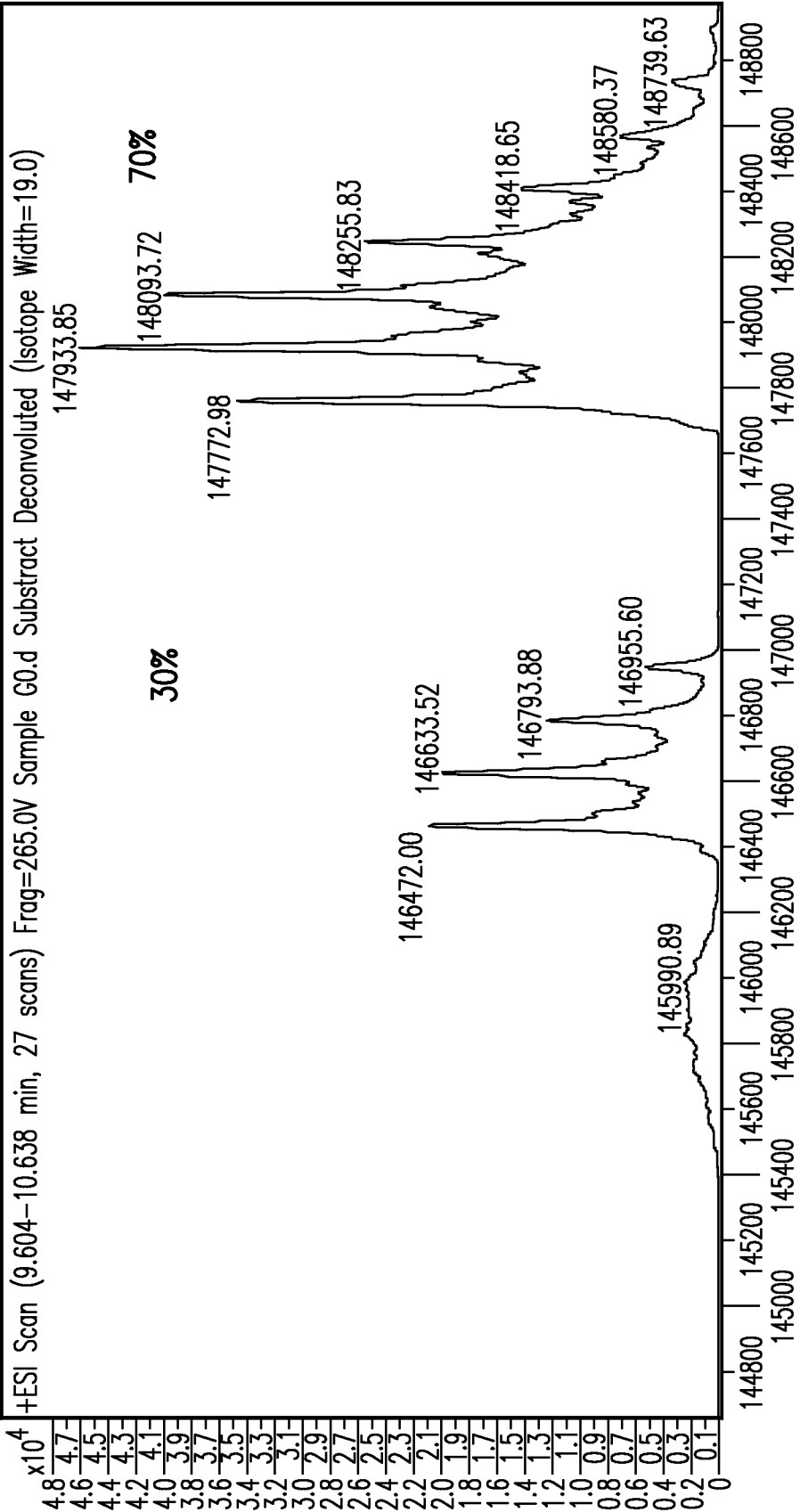
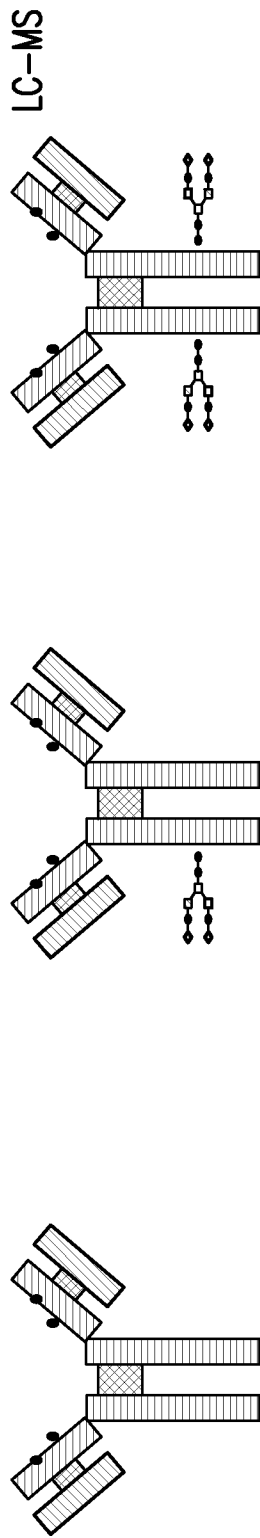


FIG.21A

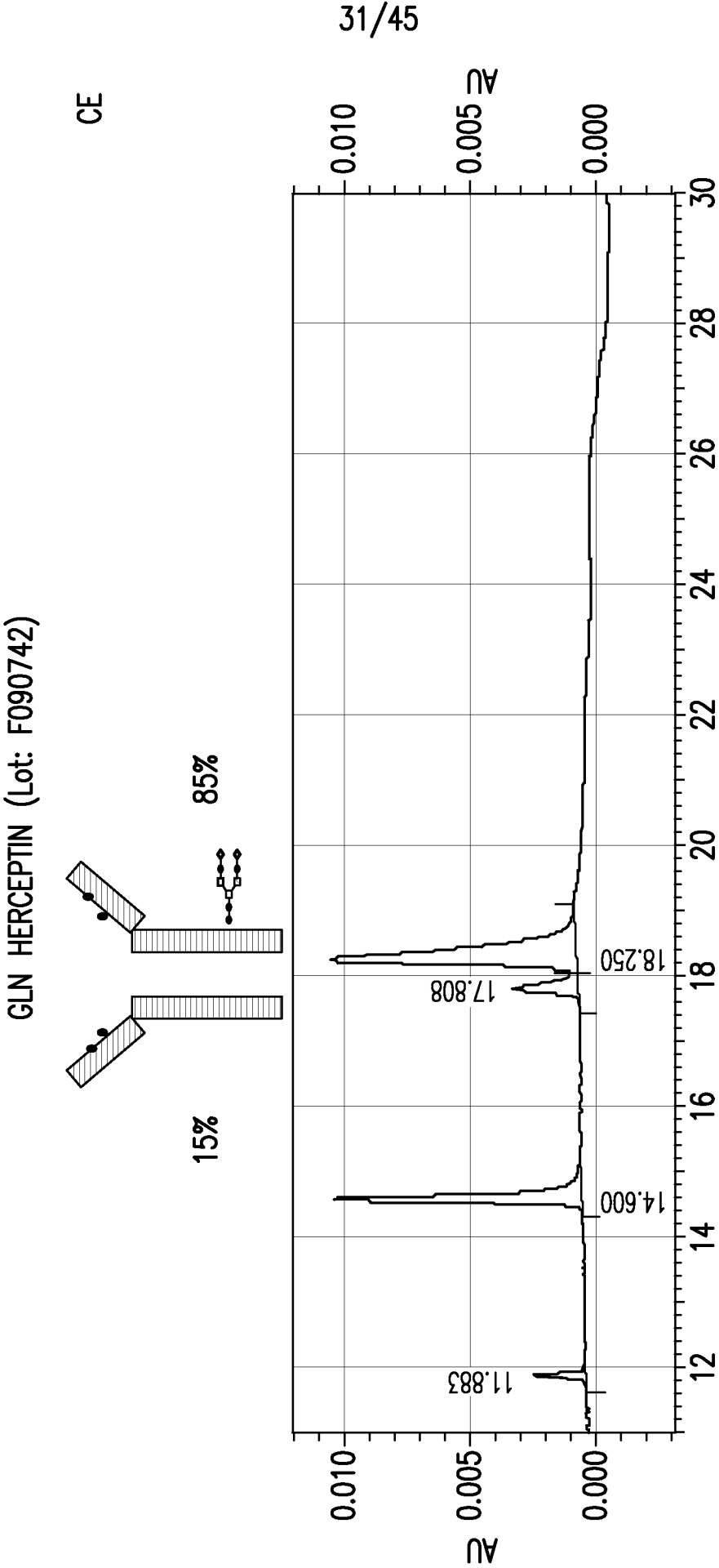


FIG.21B

GLN HERCEPTIN (Lot: F090742) + PNGase F

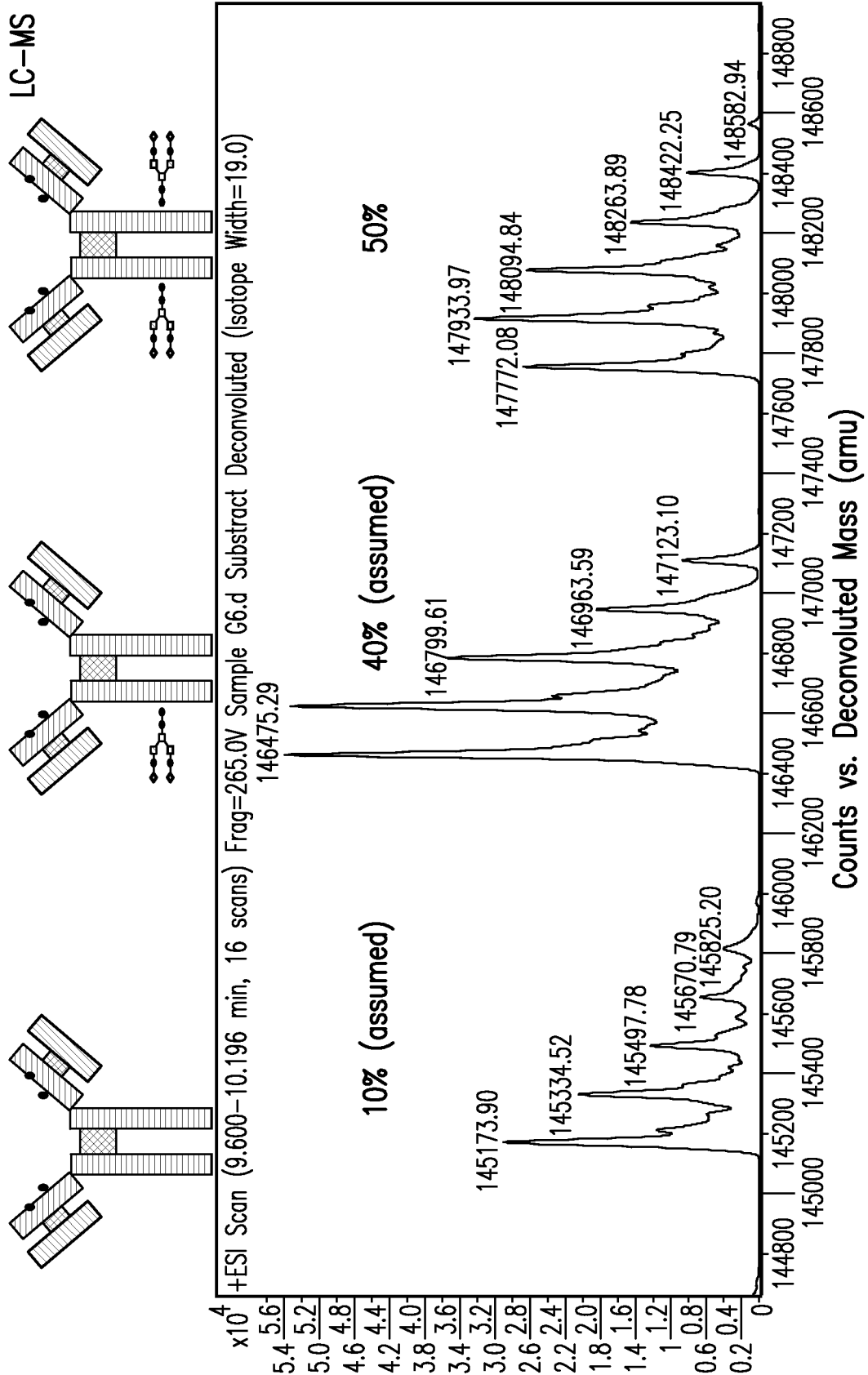


FIG.22A

GLN HERCEPTIN (Lot: F090742) + PNGase F

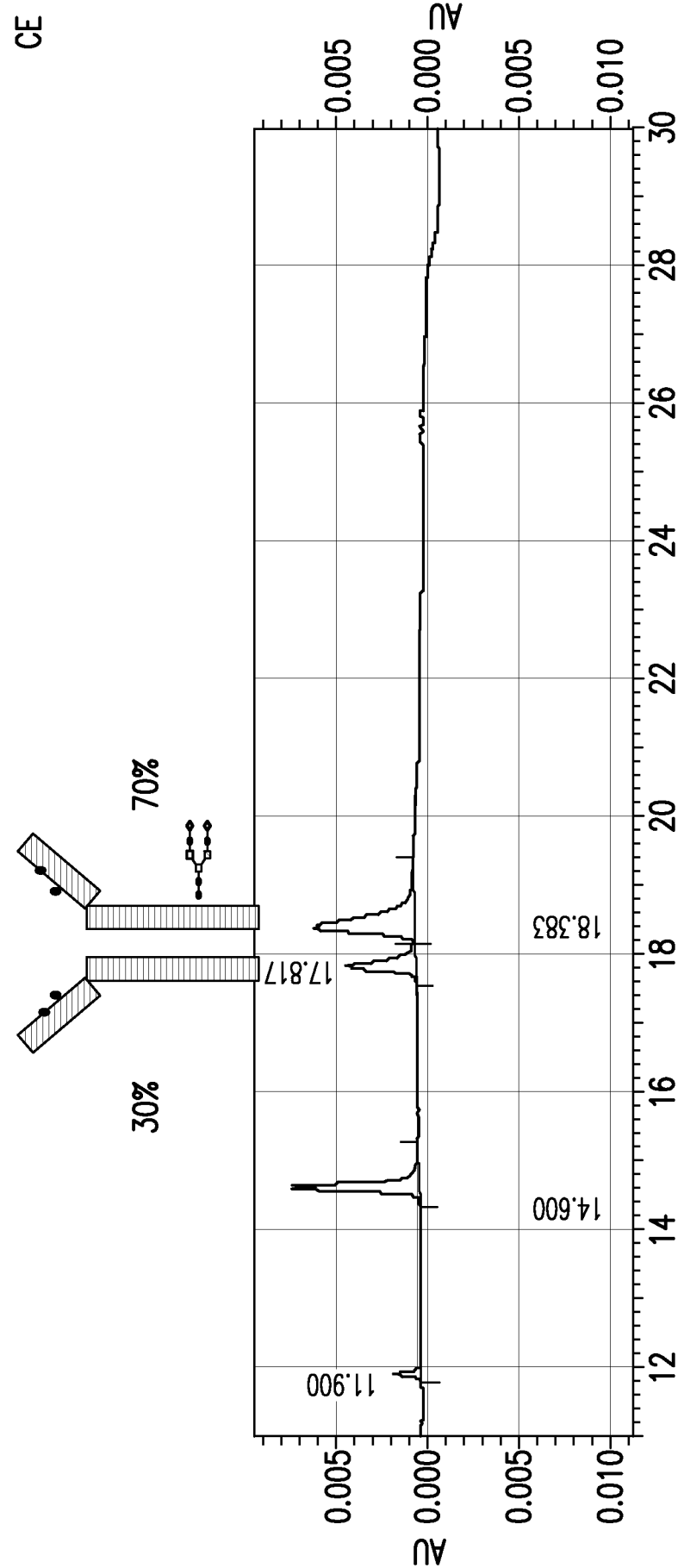


FIG.22B

34/45

YGLY7961

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ-URA5-lacZ]

pGLY2456

**YGLY8146**

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ-URA5-lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]



Counterselect on 5-FOA

YGLY9296

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

pGLY5048

**FIG.23A**

35/45

YGLY9469

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1]

pGLY5019 ↓

YGLY9797

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2Δ::Nat^R]

pGLY5085 ↓

FIG.23B

36/45

YGLY12900

[ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2Δ::Nat^R
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33]

pGLY7240 ↓

YGLY15660

[ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2Δ::Nat^R
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33
TRP2::Sh ble/AOX1p/PpCWP1-GMCSF]

pGLY6301 ↓

FIG.23C

37/45

YGLY16349

[*ura5Δ::ScSUC2 och1Δ::lacZ bmt2Δ::lacZ/KIMNN2-2*
mnn4L1Δ::lacZ/MmSLC35A3 pno1Δ mnn4Δ::lacZ
ADE1::lacZ/NA10/MmSLC35A3/FB8
his1Δ::lacZ/ScGAL10/XB33/DmUGT
arg1Δ::HIS1/KD53/TC54
bmt4Δ::lacZ bmt1Δ::lacZ bmt3Δ::lacZ
TRP2::ARG1/MmCST/HsGNE/HsCSS/HsSPS/MmST6-33
ste13Δ::lacZ-URA5-lacZ/TrMDS1 dap2Δ::Nat^R
TRP5::Hyg^RMmCST/HsGNE/HsCSS/HsSPS/MmST6-33
TRP2::Sh ble/AOX1p/PpCWP1-GMCSF
URA6::ScARR3/AOX1p/LmSTT3D]

FIG.23D

38/45

Glossary:

ScSUC2	<i>S. cerevisiae</i> Invertase
OCH1	Alpha-1,6-mannosyltransferase
K1MNN2-2:	<i>K. lactis</i> UDP-GlcNAc transporter
BMT1:	Beta-mannose-transfer (beta-mannose elimination)
BMT2:	Beta-mannose-transfer (beta-mannose elimination)
BMT3:	Beta-mannose-transfer (beta-mannose elimination)
BMT4:	Beta-mannose-transfer (beta-mannose elimination)
MNN4L1:	MNN4-like 1 (charge elimination)
MmSLC35A3	Mouse homologue of UDP-GlcNAc transporter
PN01:	Phosphomannosylation of N-linked oligosaccharides (charge elimination)
MNN4:	Mannosyltransferase (charge elimination)
ScGAL10	UDP-glucose 4-epimerase
XB33	Truncated HsGalT1 fused to ScKRE2 leader
DmUGT	UDP-Galactose transporter
KD53	Truncated DmMNSII fused to ScMNN2 leader
TC54	Truncated RnGNTII fused to ScMNN2 leader
NA10	Truncated HsGNTI fused to PpSEC12 leader
FB8:	Truncated MmMNS1A fused to ScSEC12 leader
MmCST	Mouse CMP-sialic acid transporter
HsGNE	Human UDP-GlcNAc 2-epimerase/N-acetylmannosamine kinase
HsCSS	Human CMP-sialic acid synthase
HsSPS	Human N-acetylneuraminate-9-phosphate synthase
MmST6-33	Truncated Mouse alpha-2,6-sialyl transferase fused to ScKRE2 leader
TrMDS1:	Secreted <i>T. reesei</i> MNS1
STE13	Golgi dipeptidyl aminopeptidase
DAP2	Vacuolar dipeptidyl aminopeptidase
NatR	Nourseothricin resistance marker
HygR	Hygromycin resistance marker

FIG.23E

39/45

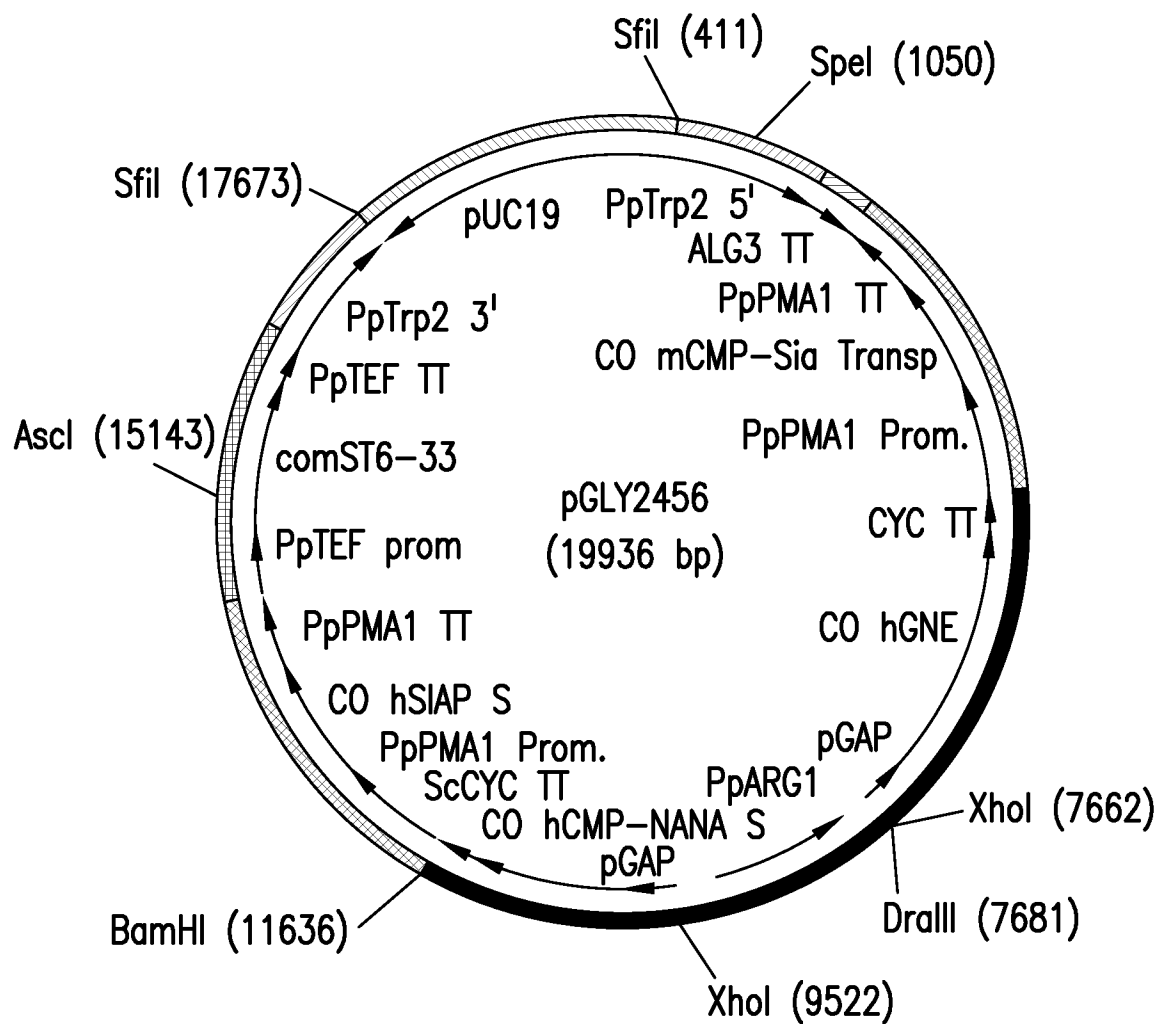


FIG.24

40/45

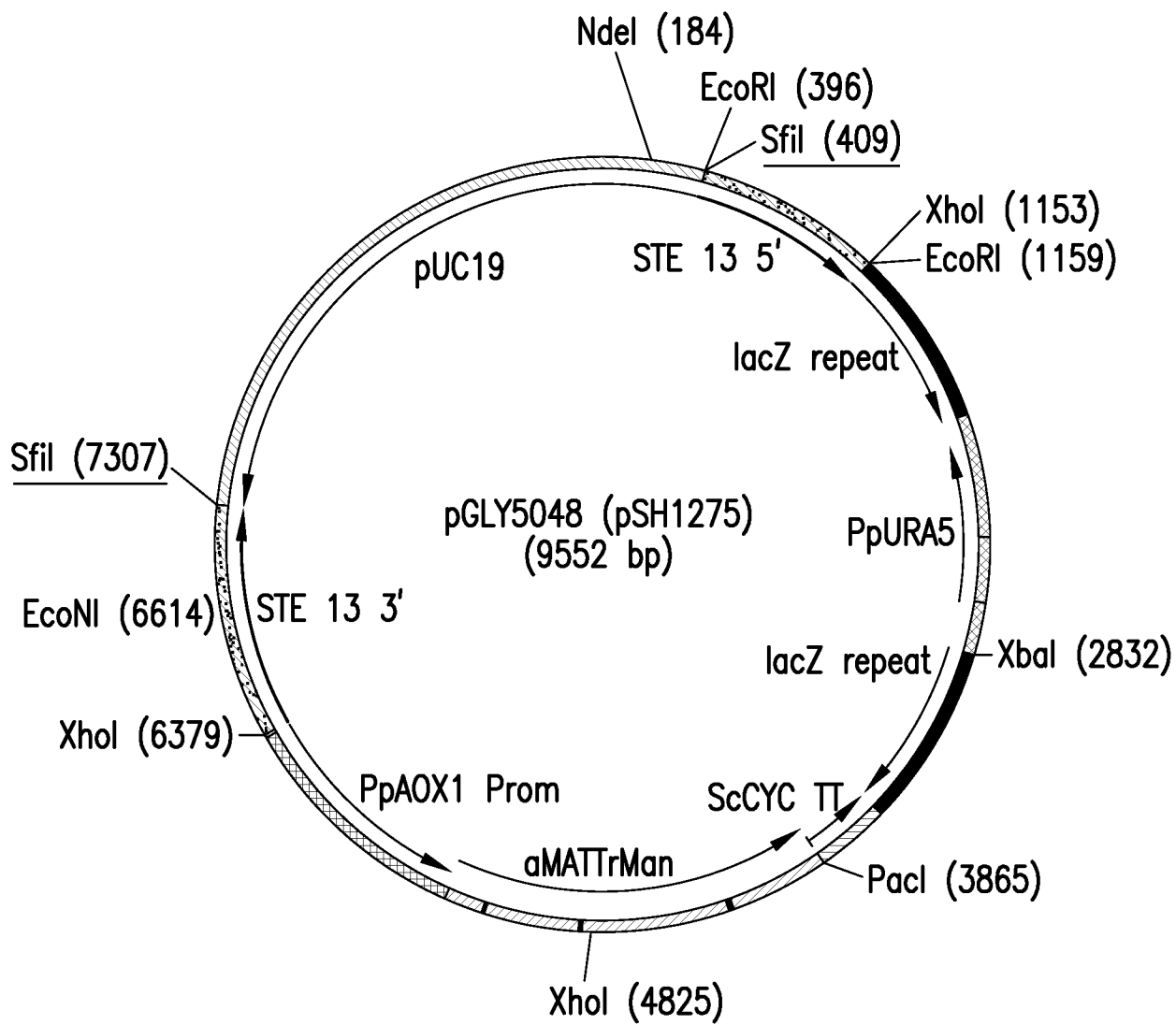


FIG.25

41/45

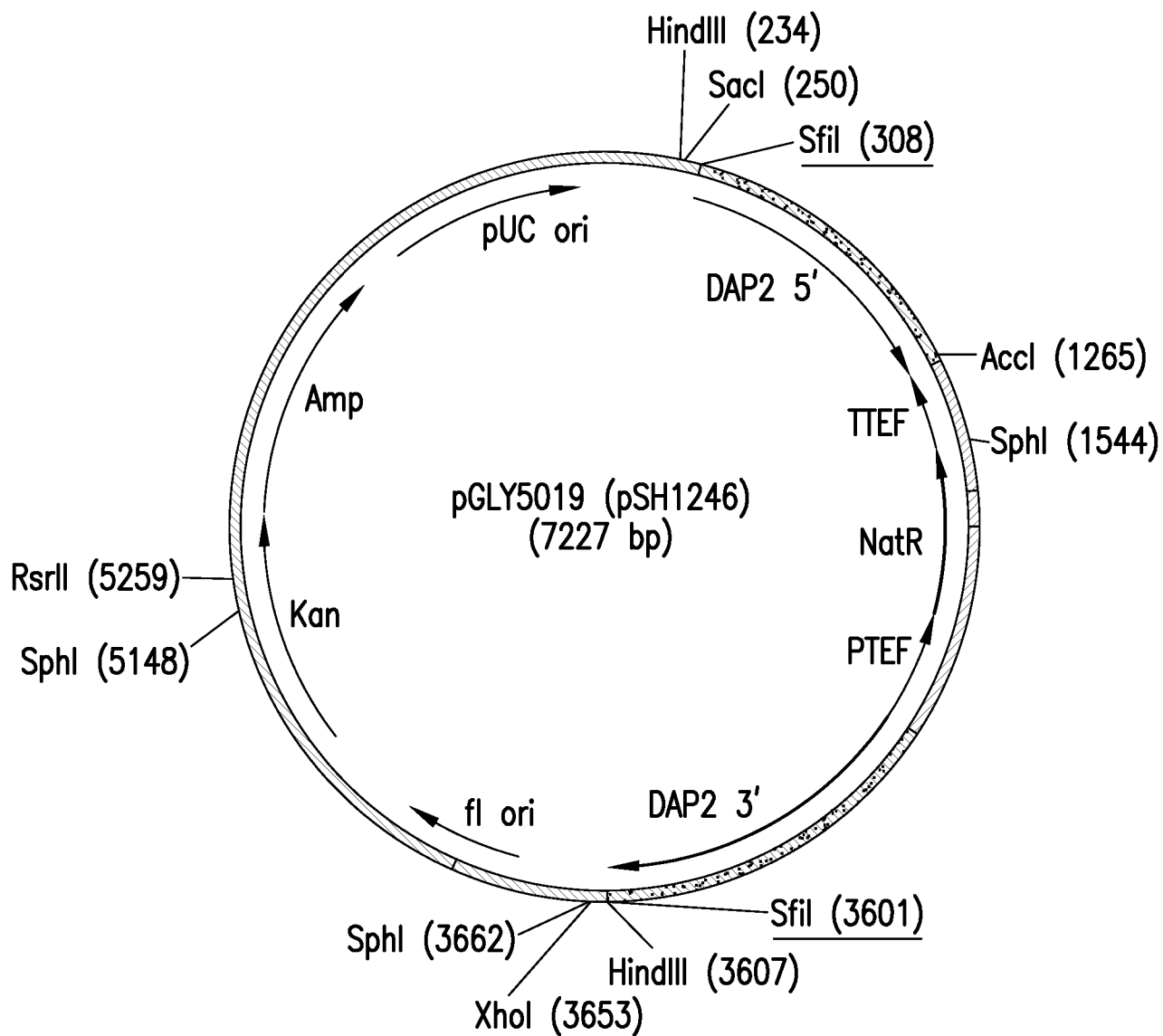


FIG.26

42/45

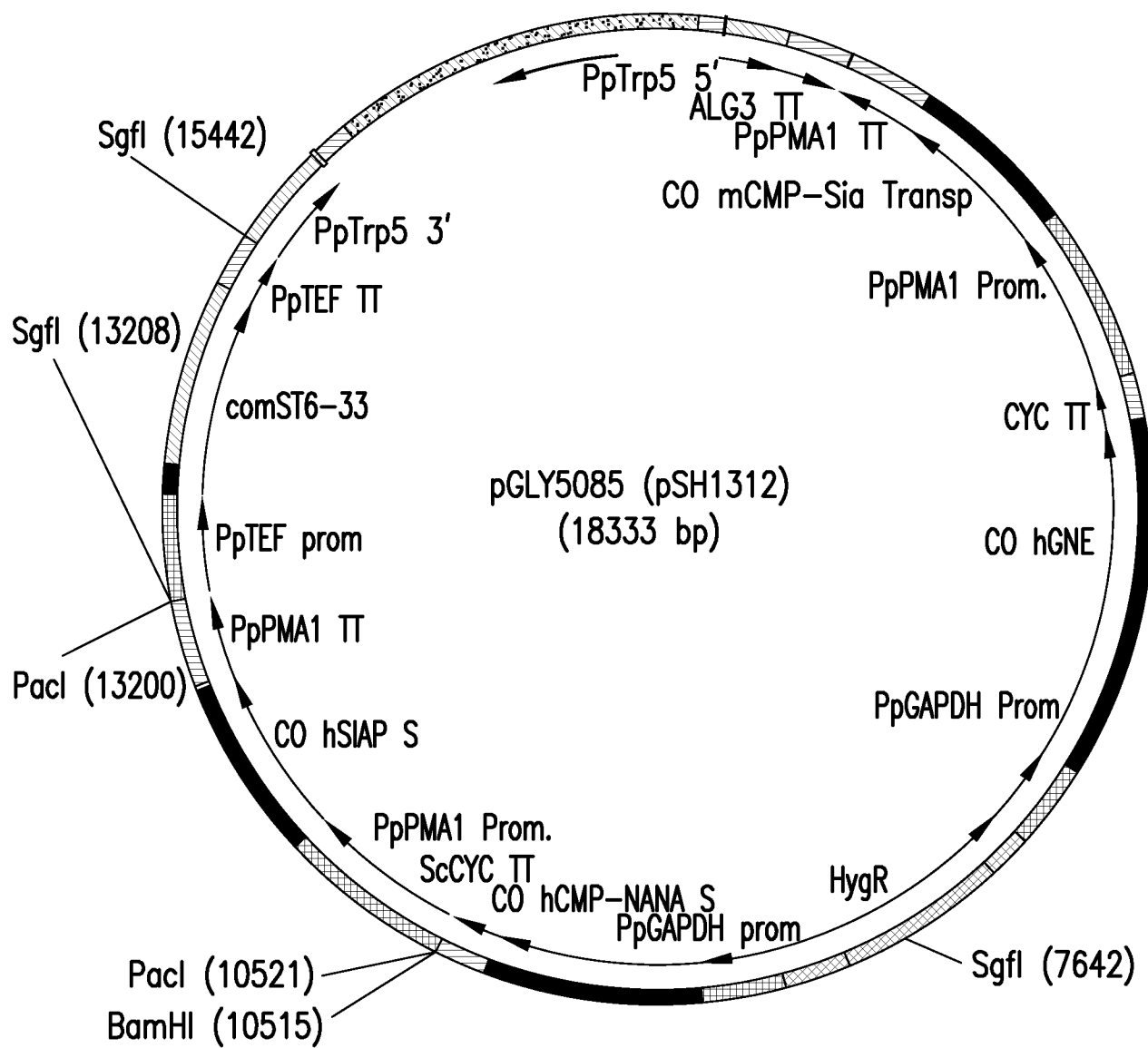


FIG.27

43/45

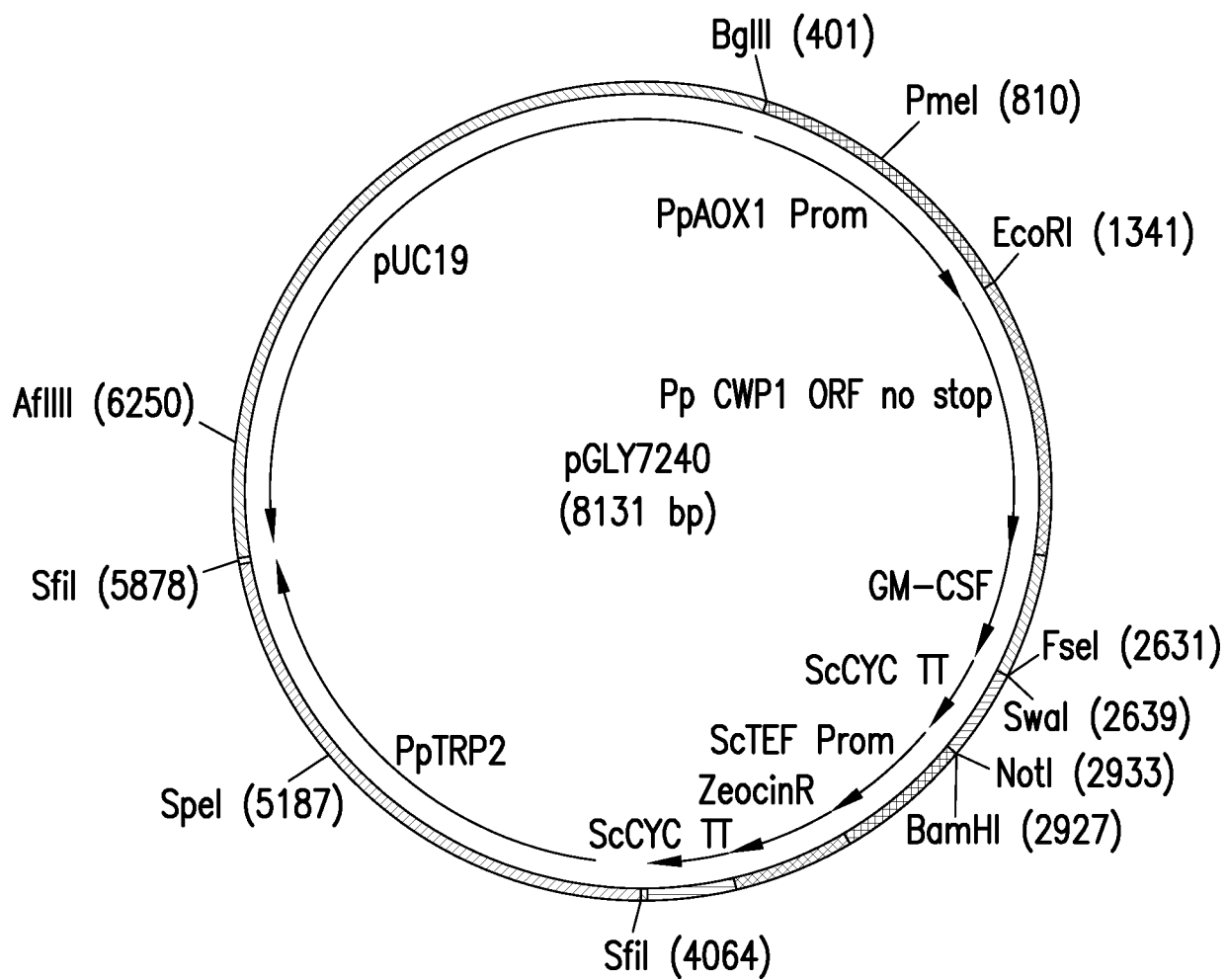


FIG.28

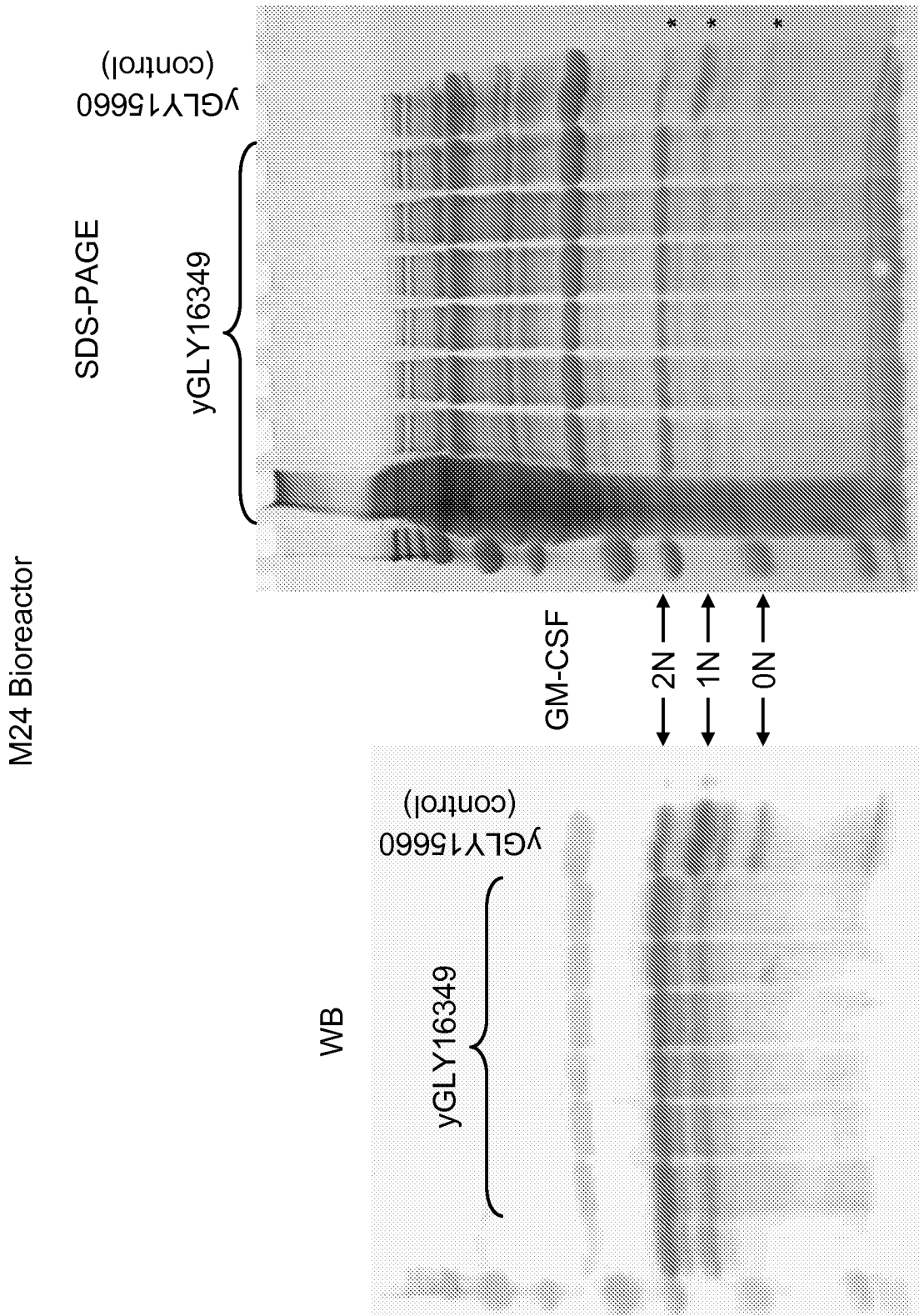


FIG.29

45/45

yGLY15560 (control)

+ESI Scan (14.615–17.642 min, 189 scans)

Frag=265.0V GM-CSF.d Subtract Deconvoluted (Isotope Width=8.9)

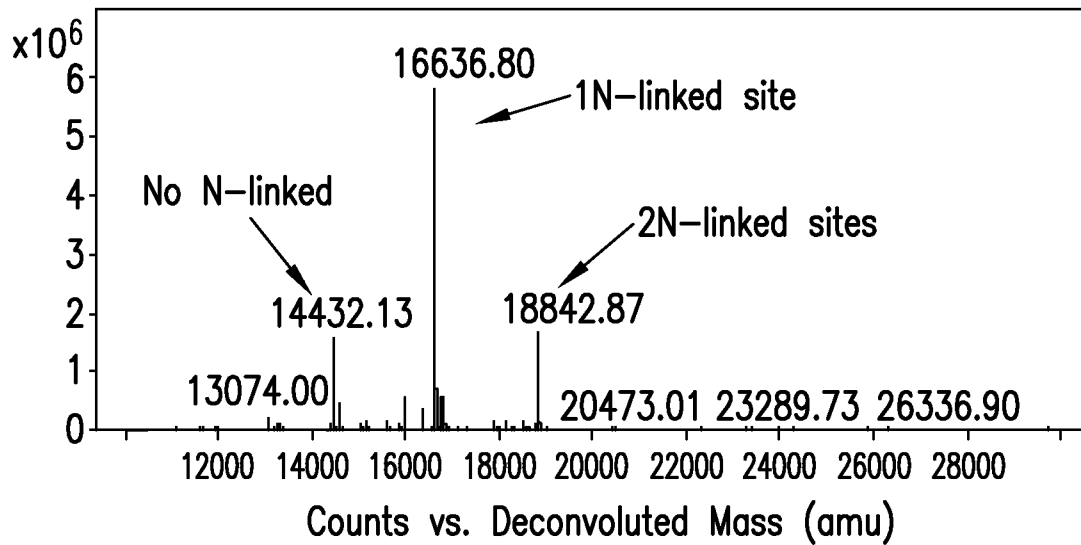


FIG.30A

yGLY16349 (LmSTT3D)

+ESI Scan:1 (15.118–16.891 min, 111 scans)

Frag=265.0V SF081209_Capto_MMC.d Subtract Deconvoluted (l...

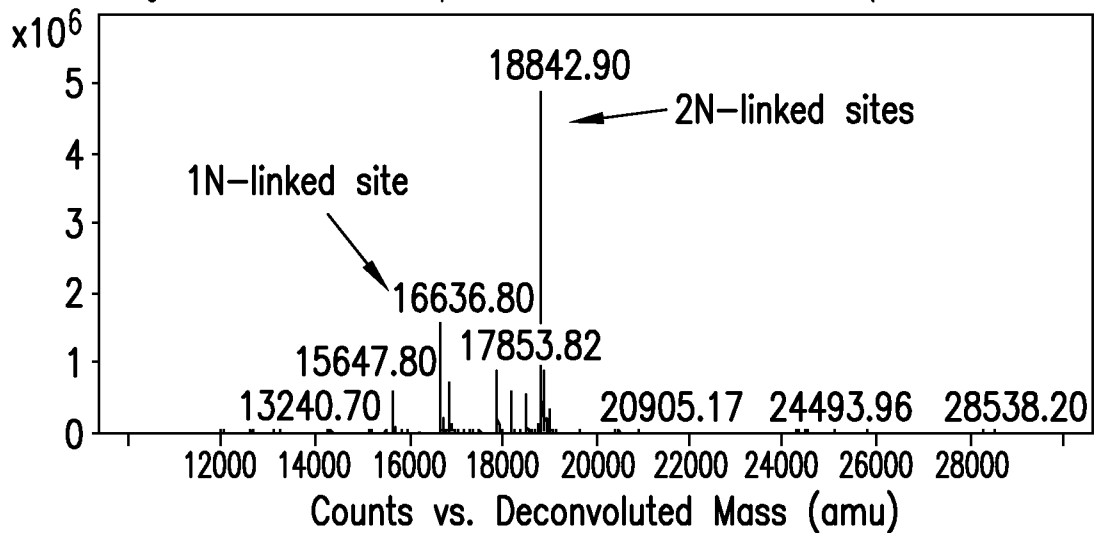


FIG.30B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/25878

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 1/00; C12N 15/74 (2011.01)

USPC - 435/255.5; 435/471

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

USPC: 435/255.5; 435/471

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

USPC: 435/69.1 (keywords below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST (PGPB,USPT,USOC,EPAB,JPAB); Google; PubMed: oligosaccharyltransferase, single-subunit, STT3A, STT3B, STT3C, STT3D, LmSTT3D, glycoprotein, heterologous, glycosylated, n-glycan, erythropoietin, epo, pichia pastoris, och1

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	NASAB et al. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. Mol. Biol. Cell 2008, 19(9):3758-3768; abstract; pg 3759, para 8; pg 3762, para 2; pg 3763, para 1; Table 1	1-6, 8-9
Y	US 2006/0252672 A1 (BETENBAUGH et al.) 9 November 2006 (09.11.2006) para [0008], [0011], [0013], [0016]	1-6, 8-9
Y	US 2009/0069232 A1 (CALLEWAERT et al.) 12 March 2009 (12.03.2009) para [0003], [0113], [0135]	9
A	HESE et al. The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. Glycobiology 2009, 19(2):160-171	1-6, 8-9
A	MAITA et al. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. J. Biol. Chem. 12 February 2010, 285(7):4941-4950	1-6, 8-9
A	IZQUIERDO et al. Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. EMBO J. 2009, 28(17):2650-2661	1-6, 8-9

☐ Further documents are listed in the continuation of Box C.


* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 June 2011 (14.06.2011)

Date of mailing of the international search report

27 JUN 2011

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/25878

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

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

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

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

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I+: claims 1-9, 26, drawn to a method for producing a heterologous glycoprotein by (a) providing a host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein, and wherein the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex are expressed; and (b) culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein. The first invention is restricted to the erythropoietin. Should an additional fee(s) be paid, Applicant is invited to elect an additional heterologous protein(s) to be searched. The exact claims searched will depend on Applicant's election.
[NOTE: Claims 7 and 26 were excluded from the first invention and not searched as they are directed to nonelected subject matter.]

----- (Continued on Extra Sheet) -----

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

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

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

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-6, 8-9, restricted to erythropoietin

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

Continuation of Box III - observations where unity of invention is lacking:

Group II, claims 10-15, 23-24, drawn to a method for producing a glycoprotein composition in which at least 70% of the N-glycosylation sites on the glycoproteins in the composition are occupied with an N-glycan, comprising (b) culturing the recombinant host cell under conditions for expressing the glycoprotein to produce the compositions in which at least 70% of the N-glycosylation sites on the glycoproteins in the composition are occupied with the N-glycan.

Group III+, claims 16-22, drawn to a host cell, comprising: (a) a first nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the host cell expresses the genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex. The first invention is restricted to the erythropoietin. Should an additional fee(s) be paid, Applicant is invited to elect an additional heterologous protein(s) to be searched. The exact claims searched will depend on Applicant's election.

Group IV, claim 25, drawn to a recombinant *Pichia pastoris* host cell that has a genotype that is the same as the genotype of recombinant *Pichia pastoris* strain YGLY 1440 1.

The inventions listed as Groups I+ through IV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions of Groups I+ through III+ do not include the inventive concept of a recombinant *Pichia pastoris* host cell strain YGL Y 14401, as required by Group IV.

The inventions of Group I+ do not include the inventive concept of a a method for producing a glycoprotein composition in which at least 70% of the N-glycosylation sites on the glycoproteins in the composition are occupied with an N-glycan, comprising (b) culturing the recombinant host cell under conditions for expressing the glycoprotein to produce the compositions in which at least 70% of the N-glycosylation sites on the glycoproteins in the composition are occupied with the N-glycan, as required by Group III.

The inventions of Groups I+ and II share the technical feature of a method for producing a heterologous glycoprotein by (a) providing a host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase and a nucleic acid molecule encoding the heterologous glycoprotein, and wherein the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex are expressed; and (b) culturing the host cell under conditions for expressing the heterologous glycoprotein to produce the heterologous glycoprotein. However, this shared technical feature does not represent a contribution over prior art as being obvious over an article titled "All in one: *Leishmania* major STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*" by NASAB et al. (Mol. Biol. Cell 2008, 19(9):3758-3768) (hereinafter "NASAB") in view of US 2006/0252672 A1 to BETENBAUGH et al. (hereinafter "Betenbaugh") as follows:

Nasab teaches a method for producing a glycoprotein (pg 3762, para 2), comprising:

(a) providing a host cell that includes at least one nucleic acid molecule encoding a heterologous single-subunit oligosaccharyltransferase (yeast cell comprising a nucleic acid molecule encoding *Leishmania* major STT3D - pg 3759, para 8; 3762, para 2) and a glycoprotein (CPY - pg 3762, para 2); and

(b) culturing the host cell under conditions for expressing the glycoprotein to produce the glycoprotein (pg 3762, para 2). Nasab does not teach that the glycoprotein is a heterologous glycoprotein encoded by a heterologous nucleic acid molecule. Betenbaugh teaches a method for producing a heterologous glycoprotein comprising providing a host cell that includes at least one nucleic acid molecule encoding an oligosaccharyltransferase (para [0011], [0016]) and a nucleic acid molecule encoding a heterologous glycoprotein (para 0011), [0016]). It would have been obvious to one of skill in the art that any known glycoprotein could be inserted heterologously, as taught by Betenbaugh, into the host cells taught by Nasab, in order to produce it in its glycosylated form, because heterologous protein expression via introduction of heterologous nucleic acid molecules was routine in the art.

As to the limitation that the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex are expressed, Nasab teaches that all of the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex except for the endogenous *stt3* gene are expressed (delta *stt3* mutant strain - pg 3759, para 8; pg 3762, para 2). However, it would have been obvious that expression of all of the endogenous OTase complex host cell genes would not interfere with expression of the heterologous OTase, and that expression of all of said endogenous host cell genes would be preferable if it was desired to perform the disclosed method in any other types of host cells, in which the endogenous genes would be expressed unless they were actively knocked out, which would be unnecessarily time-consuming and labor-intensive. As said method would have been obvious to one of ordinary skill in the art at the time of the invention, this cannot be considered a special technical feature that would otherwise unify the groups.

The inventions of Group III+ share the technical feature of a host cell, comprising: (a) a first nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase oligosaccharyltransferase; and (b) a second nucleic acid molecule encoding a heterologous glycoprotein; and the host cell expresses the genes encoding the proteins comprising the endogenous oligosaccharyltransferase (OTase) complex. However, this shared technical feature does not represent a contribution over Nasab and Betenbaugh as follows:

Nasab teaches a host cell, comprising: (a) a first nucleic acid molecule encoding at least one heterologous single-subunit oligosaccharyltransferase oligosaccharyltransferase (pg 3759, para 8; 3762, para 2); and (b) a glycoprotein (pg 3762, para 2).

Nasab does not teach that the glycoprotein is a heterologous glycoprotein encoded by a heterologous nucleic acid molecule. Betenbaugh teaches a host cell for producing a heterologous glycoprotein comprising a first nucleic acid molecule encoding an oligosaccharyltransferase (para [0011], [0016]) and a nucleic acid molecule encoding a heterologous glycoprotein (para 0011), [0016]). It would have been obvious to one of skill in the art that any known glycoprotein could be inserted heterologously, as taught by Betenbaugh, into the host cells taught by Nasab, in order to produce it in its glycosylated form, because heterologous protein expression via introduction of heterologous nucleic acid molecules was routine in the art.

(Continued on Extra Sheet)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 11/25878

Continuation of Box III - observations where unity of invention is lacking (first extra sheet):

As to the limitation that the host cell expresses the genes encoding the proteins comprising the endogenous oligosaccharyltransferase OTase complex, Nasab teaches that all of the endogenous host cell genes encoding the proteins comprising the endogenous OTase complex except for the endogenous stt3 gene are expressed (delta stt3 mutant strain - pg 3759, para 8; pg 3762, para 2). However, it would have been obvious that expression of all of the endogenous OTase complex host cell genes would not interfere with expression of the heterologous OTase, and that expression of all of said endogenous host cell genes would be preferable if it was desired to perform the disclosed method in any other types of host cells, in which the endogenous genes would be expressed unless they were actively knocked out, which would be unnecessarily time-consuming and labor-intensive. As said host cell would have been obvious to one of ordinary skill in the art at the time of the invention, this cannot be considered a special technical feature that would otherwise unify the groups.

Another special technical feature of the inventions listed as Group I+ and III+ is the specific heterologous protein recited therein. As no significant structural similarities can readily be ascertained among the proteins, the inventions do not share a special technical feature. Without a shared special technical feature, the inventions lack unity with one another.

Groups I+ through IV therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.