Described herein are methods and genetically engineered fungal cells useful for producing target molecules containing mammalian-like complex N-glycans or containing intermediates in a mammalian glycosylation pathway.

A.

\[
\text{Man}_5\text{GlcNAc}_2 \quad \text{Man}_3\text{GlcNAc}_2
\]

B.

[Diagram of genome with various genes and regulatory elements]
FIGURE 1

A.

\[
\begin{align*}
\text{Man}_5\text{GlcNAc}_2 & \quad \text{Man}_3\text{GlcNAc}_2
\end{align*}
\]

B.

\[
P_{\text{Lac}} \rightarrow \text{loxP} \rightarrow \text{YIURA3}
\]

\[
p_{\text{YLOCH1PUT TOPC1}} \quad 5848 \text{ bp}
\]

\[
\text{ZeoCMR} \quad \text{KanamycinR} \quad \text{ccdB lethal gene ORF} \quad \text{OCH1 T fragment} \quad \text{loxP} \quad \text{loxR}
\]
FIGURE 2

- Dextran
- RNaseB
- polD ImuGA WT strain
- polD ImuGA Δuch1 = G013
- polD ImuGA Δuch1 - cured = G014
FIGURE 4

pold Inga Δovt1 - cured = G014

G014 + TEF1ManHDEL(Rd) = G016

G014 + Hpd4ManHDEL(Rd) = G018

G018 - cured = G036
FIGURE 6

Dextran

RNase B

pold imaga AochI - cured = G014

G014 + Hp4dManHDEL(Tg-oxyT) = G053

G014 + Hp4dManHDEL(Tg-leu2) = G046
FIGURE 9

Dextran

RNaseB

G018 - cured = G036

G036 + TELJKre2-GnT1 (Rd) = G040

G040 + in vitro α-1,2-mannosidase

G040 + in vitro Nε-de-N-acetylhexosaminidase
FIGURE 12

Dextran

RNaseB

G036 + TEF/Kroc2-GnT1 (Rd) = G040

G040 + TEF/Mnn2-ManII (Rd) = G043

G043 + in vitro α:1,2-mannosidase

G043 + in vitro hexosaminidase
GTGCTGCAGCGACAGCGCTGGAATCTAGCATCGTGTATCAACCCAGGCCGCGACACCATCTCTCAACCGAGCAAGCTGCTGAACGTGGGCTTCCAGGAGGCCTGGAAGGA
CTACGACTACACCTGTTTCTCAGATCCGGAGTGAACTGCTATCCCATAGAAGAACCAAAAGCGCCTACCACTTGTTCTCCAGGCCCGAACACATCTCTCTGTCG
CGGCTCTCTCTGCCCTACGTCAGTACTTGGGCGGCCGTTTTCTGCCCCTGCTAAGCAGCAGGTTCTGACCACTCAACGCTTTCCCAACACACTACTGGGCT
CGGGCGAGAGCAGACGACATTTTCAACCGAAGCTGGTCCGGAGGCATTGCTATCTCTCTCTGGACCGCAACGCCCGTGCTGGGCGATGTCGAATGAATCGG
CAACCCCGAGGATTTGACCGAATTGTCACTAAAGGAACCAATGCTGTCTGACGCGCTGTAATCTCTCAGTTACCAGGTTGGCTGACCTGCGACAGCGATA
CCCTCTGTACACCCAGATCCACCGTGGACATCGGCACACCCCTCTTTAG (SEQ ID NO:10)
FIGURE 15

G036 + TEF/Kre2-GalT1 (Rd) = G040

G040 + TEF/Mnn2-Gal10-GalT1 (Rd) = G044

G044 + α-1,2-mannosidase

G044 + galactosidase

G044 + galactosidase and hexosaminidase

G040 + TEF/Mnn2-Gal10-GalT1 (Rd) = G044

G044 + α-1,2-mannosidase

G044 + galactosidase

G044 + galactosidase and hexosaminidase
FIGURE 16

pYLAig3 PUT-ALG6

8386 bp
FIGURE 17

Dextran

RNaseB

G018 - cured = G036

[Medium 1]

G036 + 5alg3-Hpd4dLG6 = G039

G039 + α-1,2-mannosidase

[Medium 2]

G036 + 5alg3-Hpd4dLG6 = G039
FIGURE 19

RNaseB

Dextran

strain G039

G039 + TEF1Kre2-GnTH(HygR) = G047

G047 + in vitro α-1,2-mannosidase

G047 + in vitro hexosaminidase

G045 + TEF1Kre2-GnTI = G048

G048 + in vitro α-1,2-mannosidase

G048 + in vitro hexosaminidase
FIGURE 24

Dextran

RNaseB

G045 + TEFKre2-GnTI = G048

G045 + TEFKre2-GnTI + TEFKre2-GnTI1 = G050

G050 + /α-1,2-mannosidase

G050 + /α-hexosaminidase

G039 + TEFKre2-GnTI (Hyg) = G047

G047 + TEFKre1-GnTI = G051

G051 + /α-1,2-mannosidase

G051 + /α-hexosaminidase
FIGURE 28

G045 + Hpd/krz2-GntT (Tg-ade2) = G057

RNaseB

G057 + Hpd/Gls2α/β (Tg-ura2) = G061

G061 + α-1,2-mannosidase

G061 + Jack bean α-mannosidase

G057 + Hpd/Gls2α/β (Rd) = G060

G060 + α-1,2-mannosidase

G060 + Jack bean α-mannosidase
FIGURE 30B

MKLSTILFTACATLAAALPSITPSEAALQKRGGDIQMTQSPSSASVGVDRVTITCR
ASQDVNTAVAQYQQKPGKAPKLLYSAFLYGVSFSGRSGSDFTLTLSSLSQPEDFA
TYYCQYHTTPFGQGTKVEKRVTAAAPSVFIFFPSDEQLKSCTASVCLNNFYPREA
KVKWVQDNALQGGSQVESVFQDNSKDSYSLSTLTSKADYEHKVACEVTHQGLSSPVTKSFNRGEC
FIGURE 32

RNaseB

Man3-platform (G045)

G045 + Hp4dKre2-GnTI (Tg) = G057

G057 + Hp4dAnGls2(α+β) (Tg) = G061

G061 + Hp4dMnn2-GnTII (Tg) = G071
(μg/l active herceptin)

<table>
<thead>
<tr>
<th>time-points</th>
<th>2</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>14</th>
<th>16</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>hours</td>
<td>0 hr</td>
<td>25 hr</td>
<td>39 hr</td>
<td>63 hr</td>
<td>87 hr</td>
<td>120 hr</td>
<td>134.5 hr</td>
<td>148 hr</td>
</tr>
<tr>
<td>titer in μg/l</td>
<td>224</td>
<td>1996</td>
<td>2880</td>
<td>3627</td>
<td>4642</td>
<td>5427</td>
<td>8950</td>
<td>11730</td>
</tr>
</tbody>
</table>
FIGURE 34A
FIGURE 34B
YEAST STRAINS PRODUCING MAMMALIAN-LIKE COMPLEX N-GLYCANS

CROSS-REFERENCE TO RELATED

[0001] This application claims priority to U.S. Application Ser. No. 61/262,828, filed on Nov. 19, 2009. The disclosure of the prior application is incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The invention relates to methods and materials for producing glycoproteins in fungal cells, and more particularly, to genetically engineering fungal cells to produce proteins containing mammalian-like complex N-glycans or proteins containing intermediates within a mammalian glycosylation pathway.

BACKGROUND

[0003] High performance expression systems are required to produce most biopharmaceuticals (e.g., recombinant proteins) currently under development. The biological activity of many of these biopharmaceuticals is dependent on their post-translational modification (e.g., phosphorylation or glycosylation). A yeast-based expression system combines the ease of genetic manipulation and fermentation of a microbial organism with the capability to secrete and to modify proteins. However, recombinant glycoproteins produced in yeast cells exhibit mainly heterogeneous high-mannose and hybrid-mannose glycan structures, which can be detrimental to protein function, downstream processing, and subsequent therapeutic use, particularly where glycosylation plays a biologically significant role.

SUMMARY

[0004] The methods and genetically engineered fungal cells described herein can be used to produce target molecules (e.g., target proteins) that contain mammalian-like N-glycans or contain intermediates within the mammalian (e.g., human) glycosylation pathway. Target molecules isolated from such engineered cells can be used for biopharmaceutical applications including antibody production, cytokine production, and for treatment of metabolic disorders such as lysosomal storage disorders.

[0005] In one aspect, this document features a method of producing a fungal cell (e.g., Yarrowia lipolytica or Arxula adeninivorans) capable of producing proteins comprising GlcNacMan, GlcNAc2, N-glycans. The method includes providing a fungal cell genetically engineered to produce proteins comprising Man, GlcNAc2, N-glycans; and introducing into the cell a nucleic acid encoding a GlcNAc-transferase I, wherein the nucleic acid includes a nucleotide targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment (e.g., Golgi apparatus), wherein expression of the GlcNAc-transferase I in the fungal cell produces proteins including GlcNAcMan, GlcNAc2, N-glycans. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to include the GlcNAcMan, GlcNAc2, N-glycans. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon-β, GM-CSF, Interferon γ, or erythropoietin.

[0006] The fungal cell genetically engineered to produce proteins containing Man, GlcNAc2, N-glycans can be deficient in OCH1 activity and include a nucleic acid encoding an α-1,2-mannosidase, wherein the nucleic acid encoding the α-1,2-mannosidase includes a nucleotide sequence encoding a targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum. The targeting sequence can be an HDEL sequence.

[0007] The method further can include introducing into a cell a nucleic acid encoding a mannosidase II, wherein the nucleic acid encoding the mannosidase II includes a nucleotide sequence encoding a targeting sequence to target the encoded mannosidase II to the Golgi apparatus, wherein expression of the mannosidase II in the fungal cell produces proteins containing GlcNAcMan, GlcNAc2, N-glycans.

[0008] The method further can include introducing into a cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing Ga1GlcNAcMan, GlcNAc2, N-glycans.

[0009] In another aspect, this document features a method of producing a target protein containing GlcNAcMan, GlcNAc2, N-glycans. The method includes providing a fungal cell (e.g., Yarrowia lipolytica or Arxula adeninivorans) genetically engineered to include a nucleic acid encoding a GlcNAc-transferase I, an α-1,2-mannosidase, and a mannosidase II, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein the fungal cell is deficient in OCH1 activity; and introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein containing GlcNAcMan, GlcNAc2, N-glycans. The nucleic acid encoding the α-1,2-mannosidase can include an endoplasmic reticulum targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum. For example, the targeting sequence can be an HDEL sequence. The nucleic acid encoding the GlcNAc-transferase I and the mannosidase II can include a Golgi targeting sequence, or Golgi targeting sequences, to target the encoded GlcNAc-transferase I and mannosidase II to the Golgi apparatus. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon-β, GM-CSF, Interferon γ, or erythropoietin.

[0010] In some embodiments, the method further can include introducing into the cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the
galactosyltransferase in the fungal cell produces the target protein modified to contain GalGlcNAcMan$_3$GlcNAc$_2$ N-glycans. The target protein modified to contain GalGlcNAcMan$_4$GlcNAc$_2$ N-glycans can be isolated from the fungal cell.

[0011] This document also features a method of making a fungal cell (e.g., *Yarrowia lipolytica* or *Arcula adeninivorans*) capable of producing proteins containing GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The method includes providing a fungal cell genetically engineered to produce proteins containing Man$_3$GlcNAc N-glycans; introducing into the cell a nucleic acid encoding a GlcNAc-transferase I, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment (e.g., Golgi apparatus), wherein expression of the GlcNAc-transferase I in the fungal cell produces proteins containing GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon-β, GM-CSF, Interferon γ, or erythropoietin.

[0012] The fungal cell genetically engineered to produce proteins containing Man$_3$GlcNAc$_2$ N-glycans can be deficient in ALG3 activity, and include a nucleic acid encoding an α-1,2-mannosidase, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum. Such a fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding α-1,3-glucosyltransferase (e.g., ALG6).

[0013] The method further can include introducing into the cell a nucleic acid encoding a GlcNAc-transferase II, wherein the nucleic acid encoding the GlcNAc-transferase II includes a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of the GlcNAc-transferase II in the fungal cell produces proteins containing GlcNAcMan$_3$GlcNAc$_2$ N-glycans.

[0014] The method further can include introducing into the cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing GalGlcNAcMan$_4$GlcNAc$_2$ or Gal$_2$GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The galactosyltransferase can be a fusion of a UDP-Glc-4-epimerase and catalytic domain of a β-1,4-galactosyltransferase I. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GalGlcNAcMan$_4$GlcNAc$_2$ or Gal$_2$GlcNAcMan$_3$GlcNAc$_2$ N-glycans.

[0015] The method further can include introducing into the cell a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of the α and β subunits of the Glucosidase II in the fungal cell produces proteins including GalGlcNAcMan$_4$GlcNAc$_2$ or Gal$_2$GlcNAcMan$_3$GlcNAc$_2$ N-glycans.

[0016] This document also features a method of producing a target protein containing Gal$_2$GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The method includes providing a fungal cell genetically engineered to be deficient in ALG3 activity and including a nucleic acid encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase include a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment (e.g., the Golgi apparatus); and introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein containing Gal$_2$GlcNAcMan$_3$GlcNAc$_2$ N-glycans.

[0017] In another aspect, this document features an isolated fungal cell genetically engineered to produce proteins containing GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The fungal cell can be deficient in OCH1 activity and include a nucleic acid encoding an α-1,2-mannosidase, a GlcNAc-transferase I, and a mannosidase II, wherein the nucleic acid encoding the α-1,2-mannosidase, the GlcNAc-transferase I, and the mannosidase II includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the α-1,2-mannosidase, the GlcNAc-transferase I, and the mannosidase II in the fungal cell produces proteins containing GlcNAcMan$_3$GlcNAc$_2$ N-glycans. The fungal cell further can include a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GlcNAcMan$_3$GlcNAc$_2$ N-glycans.
proteins containing GlcNAc₂Man₉GlcNAc₂ N-glycans. The genetically engineered fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding an α-1,3-glucoyltransferase. A genetically engineered fungal cell also can include a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GlcNAc₂Man₉GlcNAc₂ N-glycans. A fungal cell further can include a nucleic acid encoding the α and β subunits of a α-glucosidase I, wherein expression of the α and β subunits of the α-glucosidase I in the fungal cell produces the protein containing GlcNAc₂Man₉GlcNAc₂ N-glycans. The fungal cell further can include a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans.

[0021] This document also features a substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins containing Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans. The cells are genetically engineered to be deficient in ALG3 activity and include a nucleic acid encoding a GalGlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase include a nucleotide sequence encoding a targeting sequence, or nucleotides sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase in the cell produces proteins containing Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans. The genetically engineered fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding an α-1,3-glucoyltransferase (e.g., ALG6). The cells further can include a nucleic acid encoding the α and β subunits of the α-glucosidase II, wherein expression of the α and β subunits of the α-glucosidase II in the fungal cell produces the target protein containing Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans.

[0022] In another aspect, this document features a substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins containing Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans, wherein the cells are genetically engineered to be deficient in OCH1 activity and include a nucleic acid encoding an α-1,2-mannosidase, a GlcNAc-transferase I, a mannossidase II, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the α-1,2-mannosidase, the GlcNAc-transferase I, the mannossidase II, the GlcNAc-transferase II, and the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the α-1,2-mannosidase, GlcNAc-transferase I, mannossidase II, GlcNAc-transferase II, and galactosyltransferase in the cells produces proteins comprising Gal₉GlcNAc₂Man₉GlcNAc₂ N-glycans.

[0023] This document also features a composition that includes a glycoprotein, wherein at least 50% (e.g., at least 70% or at least 85%) of the N-glycans on the glycoprotein are GlcNAc₂Man₉GlcNAc₂ N-glycans.

[0024] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the exemplary methods and materials are described below. All publications, patent applications, patents, Genbank Accession Nos. and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present application, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

[0025] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

**DESCRIPTION OF DRAWINGS**

[0026] FIG. 1A is a representation of Man₉GlcNAc₂ and Man₀GlcNAc₂ structures.

[0027] FIG. 1B is a schematic diagram of plasmid pYIOCH1 PUT TOPO.

[0028] FIG. 2 is a series of electroforeograms depicting N-glycan analysis of secreted proteins obtained from p01d Inuga *Yarrowia lipolytica* wild-type cells or Δoch1 p01d Inuga Yarrowia lipolytica cells. The main N-glycan upon OCH1 inactivation becomes Man₀GlcNAc₂. Analysis was performed using DNA sequencer-assisted, fluorophore-assisted carbohydrate electrophoresis (DSA-FACE). “M5,” “M6,” “M8,” and “M9,” refer to the number of mannose residues conjugated to the base N-acetylglucosamine structure. The Y-axis represents the relative fluorescence units as an indication of the amount of each N-glycan structure. The X-axis represents the relative mobility of each N-glycan structure through a capillary. The top electroforeogram is an analysis of dextran for use as a mobility standard.

[0029] FIG. 3 is a schematic diagram of plasmids pYLHUXDL2preManHDEL and pYLHUXDL2preManHDEL.

[0030] FIG. 4 is a series of electroforeograms depicting the N-glycan profile after introduction of a ManHDEL (~HDEL-tagged α-1,2-mannosidase) expression cassette (either under TEF1 or Hp4d promoter control) into strain G014. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 3. The major N-glycan upon mannosidase expression is Man₀GlcNAc₂. Curing of the URA3 marker from one of these strains (G018, see Table 2) does not change the N-glycan profile.

[0031] FIG. 5 is a schematic of the construction strategy for plasmids JME926 pPRLen2-ADE2ex-Hp4dManHDEL(Y1) and OXYP289 pPTAxp1-LEU2ex-Hp4dManHDEL(Y1). See FIG. 23 for the construction of vector pYLTmAX6GnTII.

[0032] FIG. 6 is a series of electroforeograms depicting the N-glycan profile after introduction of a ManHDEL (~HDEL-tagged α-1,2-mannosidase) expression cassette (under Hp4d promoter control) into strain G014 by targeted integration (1g) in either the LEU2 or the AXP1 locus. Man₀GlcNAc₂ becomes the main N-glycan.

[0033] FIG. 7 is a depiction of the amino acid sequence (SEQ ID NO:3) and Yarrowia codon optimized nucleotide sequence (SEQ ID NO:4) of the fusion protein between the 100 N-terminal amino acids of Kre2p and the catalytic domain of human GlcNAc-transferase I. In bold: Kre2p part
of fusion protein; in normal font: GnTI I part of fusion protein; underlined: start and stop codons.

[0034] FIG. 8 is a schematic diagram of the construction strategy for plasmids pYLTmAXhGnTI and pYLTmHp4 mA.XhGnTI.

[0035] FIG. 9 is a series of electrophoreograms depicting the N-glycan profile after introduction of the GnTI I activity into strain G036 by transformation with a vector expressing GnTI I. "Rd" stands for "random integration" via the zeta sequences present on the vectors shown in FIG. 8. The major N-glycan upon expression of the GnTI I activity is GlcNAcManα1,2Manα1,2GlcNAc 2. In vitro treatment with α-1,2-mannosidase does not change the profile significantly, indicating that only small amounts of high-mannose N-glycans other than Manα1,2Manα1,2GlcNAc 2 are present. In vitro hexosaminidase treatment results in a shift from GlcNAcManα1,2Manα1,2GlcNAc 2 towards Manα1,2GlcNAc 2.

[0036] FIG. 10 is a depiction of the amino acid sequence (SEQ ID NO:7) and Yarrowia codon optimized nucleotide sequence (SEQ ID NO:8) of the fusion protein between the 36 N-terminal amino acids of Mnp2p and the catalytic domain of Drosophila melanogaster mannosidase II. In bold: Mnp2p part of fusion protein; in normal font: Man II part of fusion protein; underlined: start and stop codons.

[0037] FIG. 11 is a schematic depiction of the construction strategy for plasmids pYLTmAXhManII and pYLTmAXhManII (EU22ex).

[0038] FIG. 12 is a series of electrophoreograms depicting the N-glycan profile after introduction of the Man II activity into strain G040 by transformation with a Man II-expressing vector. "Rd" stands for "random integration" via the zeta sequences present on the vectors shown in FIG. 11. Upon expression of the Man II activity a new peak appears with higher electrophoretic mobility, as well as a 'shoulder' peak running at almost the same position as Manα1,2GlcNAc 2. In vitro hexosaminidase treatment results in a shift for ward for these peaks (next to the observed shift from GlcNAcManα1,2Manα1,2GlcNAc 2 towards Manα1,2GlcNAc 2), indicating the presence of terminal GlcNAc and thus identifying the peaks as GlcNAcManα1,2Manα1,2GlcNAc 2 and GlcNAcManα1,2Manα1,2GlcNAc 2. In vitro treatment with α-1,2-mannosidase does not change the profile significantly, indicating that only small amounts of high-mannose N-glycans other than Manα1,2Manα1,2GlcNAc 2 are present.

[0039] FIG. 13 is the amino acid sequence (SEQ ID NO:9) and Yarrowia codon optimized nucleotide sequence (SEQ ID NO:10) of the fusion protein between the 46 N-terminal amino acids of Mnp2p, the Schizosaccharomyces pombe UDP-Glc-4-epimerase-like protein and the catalytic domain of human β-1,4-galactosyltransferase I. The Mnp2p part of the fusion protein is from 1-46, linker sequences are from 47-49 and 405-408, epitope sequences of the fusion protein are from 50-404, and the Man II part of the fusion protein is from 409-763 of SEQ ID NO:9. The Mnp2p part is from nucleotides 1-138, linker sequences are from nucleotides 139-147 and 1213-1224, epitope sequences are from nucleotides 148-1212, and the Man II part is from 1225-2289 of SEQ ID NO:10. Start and stop codons are underlined.

[0040] FIG. 14 is a schematic depiction of the construction strategy for plasmids pYLTmAXsSpGal10hGnTI and pYLTmAXsSpGal10hGnTI (ADE2ex).

[0041] FIG. 15 is a series of electrophoreograms depicting the N-glycan profile after introduction of the Gal10-GnTI activity into strain G044. The resulting transformant G044 was cultivated in 2 different media. "Rd" stands for "random integration" via the zeta sequences present on the vectors shown in FIG. 14. Upon expression of the Gal10-GnTI activity a new peak appears running at a position between Manα1,2GlcNAc 2 and Manα1,2GlcNAc 2. In vitro galactosidase treatment results in a shift forward for this peak and an equal increase of GlcNAcManα1,2GlcNAc 2 (the latter being confirmed as representing this N-glycan by the double treatment with galactosidase and hexosaminidase). This indicates the presence of terminal galactose and thus identifying the new peak of the G044 profile as GalGlcNAcManα1,2GlcNAc 2. In vitro treatment with α-1,2-mannosidase indicates the presence of a large amount of high-mannose N-glycans (especially Manα1,2GlcNAc 2) that were not yet trimmed to Manα1,2GlcNAc 2.

[0042] FIG. 16 is a schematic depiction of plasmid pYLaig3PUT-ALG6.

[0043] FIG. 17 is a series of electrophoreograms depicting the N-glycan profile after introduction of pYLaig3PUT-ALG6 into strain G036. Overexpression of ALG6 results in a significant amount of glycosylated peaks (GlcManα1,2GlcNAc 2 and GlcManα1,2GlcNAc 2), indicating that the GlcManα1,2GlcNAc 2 structure that was transferred to the nascent protein is not completely trimmed towards Manα1,2GlcNAc 2 by glucosidase II. Depending on the growth medium, the generated Manα1,2GlcNAc 2 is partially (still some Manα1,2GlcNAc 2 and Manα1,2GlcNAc 2) or almost completely trimmed towards Manα1,2GlcNAc 2 by the action of the ER-localized HDEL-tagged F. reesei α-1,2-mannosidase. The Manα1,2GlcNAc 2 and Manα1,2GlcNAc 2 peaks are identified as such, by their sensitivity towards α-1,2-mannosidase. Because of the capping glucose, GlcManα1,2GlcNAc 2 and GlcManα1,2GlcNAc 2 are insensitive towards this treatment. Jack bean mannosidase is partially capable of removing the free α-1,6-linked mannose while it also converts Manα1,2GlcNAc 2 into Manα1,2GlcNAc 2.

[0044] FIG. 18 is a schematic depiction of the construction strategy for plasmid pYLTmAXhGnTI (Hyg ex).

[0045] FIG. 19 is a series of electrophoreograms depicting the N-glycan profiles after introduction of the GnTI I activity into either the non-cured (G039) or cured (G045) version of the Δalg3-Δhp4dΔALG6 strain by transformation with a GnTI I-expressing vector. The generation of GlcNAcManα1,2GlcNAc 2 was proven via a hexosaminidase digest. The new peak completely shifts back towards Manα1,2GlcNAc 2. In strain G048 conversion towards GlcNAcManα1,2GlcNAc 2 was not complete since some Manα1,2GlcNAc 2 could still be observed. This strain also has some remnant Manα1,2GlcNAc 2 as shown by the α-1,2-mannosidase digest.

[0046] FIG. 20 is a schematic depiction of the construction strategy for plasmid JME925 pPTAdc2-URA3ex-Hp4dGnTI.

[0047] FIG. 21 is a series of electrophoreograms depicting N-glycan profiles after introduction of the GnTI I activity into the cured version of the Δalg3-Δhp4dΔALG6 strain (=G045); integration of an Hp4d-driven expression construct into the ADE2 locus (Tg ade2). In this cultivation the amount of glycosylated N-glycans was high and conversion of Manα1,2GlcNAc 2 to Manα1,2GlcNAc 2 was not complete. A new peak running next to Manα1,2GlcNAc 2 was observed in transformant G057 and could be designated as GlcNAcManα1,2GlcNAc 2 based on the result if the hexosaminidase digest: the new peak completely shifts back towards Manα1,2GlcNAc 2.
[0048] FIG. 22 is the amino acid sequence (SEQ ID NO:17) and Yarrowia codon optimized nucleotide sequence (SEQ ID NO:18) of the fusion protein between the 36 N-terminal amino acids of Mann2 and the catalytic domain of rat GlcNAc-transferase II. In bold: Mann2 part of fusion protein; in normal font: GntII part of fusion protein; underlined: start and stop codons.

[0049] FIG. 23 is a schematic depiction of the construction strategy for plasmids pYLTmxARntII and pYLMxARntII (ADE2 ex).

[0050] FIG. 24 is a series of electroforograms depicting N-glycan profiles after introduction of the GntII activity into a strain synthesizing GlcNAcMan,GlcNAc2. The resulting strains were either obtained via double transformation of G045 with the GntII and GntII expression constructs or via transformation of G047 with the GntII expression construct. In both cases, the peak representing GlcNAcMan,GlcNAc2 almost completely disappeared and a new peak, about one glucose unit larger, appeared. Hexosaminidase treatment indicates the presence of two terminal GlcNAc residues onto the new N-glycan; the peak shifts about two glucose units to the left and thus represents GlcNAcMan,GlcNAc2, α-1,2-mannosidase treatment does not result into major differences, indicating that there are only limited amounts of Man1, α-GlcNAc2 present.

[0051] FIG. 25 is a schematic diagram of plasmids pYLTUXdL2preAnGlcIβ and pYLm2Etx1ElRepI2preAnGlcIβ for expression of the glucoamidase II activity.

[0052] FIG. 26 is a schematic of the construction strategy for plasmids JME923 pPTun3-LEU2ex-TefL2preAnGlcIβ [a1][a1].

[0053] FIG. 27 is a schematic of the construction strategy for plasmids JME923 pPTun3-LEU2ex-Hp4L2preAnGlcIβ [a1][a1] and Zeta1-LEU2ex-Hp4L2preAnGlcIβ [a1][a1].

[0054] FIG. 28 is a series of electroforograms depicting N-glycan profiles after introduction of the glucoamidase II activity into a strain synthesizing GlcNAcMan,3GlcNAc2. The resulting strains were either obtained via random (G060) or targeted (G061) integration of a dual expression construct for the gus2α and gus2β subunit. In both cases, a reduction of glucosylated peaks is observed. α-1,2-mannosidase treatment indicates that not all of the generated Man1,GlcNAc2 was converted towards Man1,GlcNAc2 by the heterologous HDEL-tagged α-1,2-mannosidase. Because of the capping glucoses, GlcMan1,GlcNAc2 and GlcMan1,GlcNAc2 are insensitive towards this treatment. Jack Bean mannosidase is partially capable of removing the free α-1,6-linked mannose on both the remaining glucosylated N-glycans and GlcNAcMan1,GlcNAc2. Furthermore, this treatment converts Man1,GlcNAc2 into Man1,GlcNAc2. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 27. “Tg-ade2” and “Tg-tup3” stands for targeted integration in the ADE2 resp. URA3 locus.

[0055] FIG. 29 is a series of electroforograms depicting the N-glycan profile of the secretome of strains G070 and G071, which were generated via the introduction of GlcNAc-transferase II into strain G061. The N-glycans were treated with either α-1,2-mannosidase (removing all terminal α-1,2-linked mannose residues) or hexosaminidase (removes terminal β-1,2-linked GlcNAc residues) to allow identification of the peaks in the G070 and G071 native profiles. The glucose-containing N-glycans are not sensitive to either of the two enzymes. The α-1,2-mannosidase treatment results in the trimming of Man5,GlcNAc2 and Man4,GlcNAc2 towards Man1,GlcNAc2. The hexosaminidase treatment removes the β-1,2-linked terminal GlcNAc residues that have been added by GlcNAc-transferase I and II to generate Man3,GlcNAc2.

[0056] FIG. 30A is the nucleotide sequence of the synthetic preproIp2-light chain (LC) (SEQ ID NO:32).

[0057] FIG. 30B is the amino acid sequence of the synthetic preproIp2-LC (SEQ ID NO:33).

[0058] FIG. 31A is the nucleotide sequence of the synthetic preproIp2-heavy chain (HC) (SEQ ID NO:34).

[0059] FIG. 31B is the amino acid sequence of the synthetic preproIp2-HC (SEQ ID NO:35).

[0060] FIG. 32 is a series of electroforograms depicting the N-glycan profile analysis of SuperT’glyceral shake-flask cultivations of glyco-engineered strains G045, G057, G061 and G071 that were transformed with pYLHp4L2preproHerHIC/LC (GUT2ex)-ori2. See Table 2 for a description of strains G045, G057, G061 and G071.

[0061] FIG. 33 is a graph of the results from a functional ELISA at different time-points in the G096 fed-batch fermentation.

[0062] FIGS. 34A and B are series of electroforograms depicting the N-glycan profile analysis of the secretome at different time-points within the G096 fed-batch fermentation.

DETAILED DESCRIPTION

[0063] As described herein, in vivo synthesis of mammalian-like complex N-glycans on yeast-secreted glycoproteins can be based on either a Man5,GlcNAc2 or Man4,GlcNAc2 base structure (see FIG. 1A. “Man” refers to mannose, and “GlcNAc” refers to N-glucosamine). To produce the Man5,GlcNAc2 base structure, yeast cells can be engineered such that α-1,2-mannosidase activity is increased in an intracellular compartment and Outer CFain elongation (OCH1) activity is decreased. To produce the Man3,GlcNAc2 base structure, activity of Asparagine Linked Glycosylation 3 (ALG3) and, in some embodiments, OCH1 is decreased, activity of α-1,2-mannosidase and, in some embodiments, activity of the 1,3-glucosyltransferase is increased. The N-glycan profile of proteins produced in yeast cells can be altered by further engineering the yeast cells to contain one or more of the following activities: GlcNAc transferase I (GntI) activity, mannosidase II activity, GlcNAc transferase II (GntII) activity, glucosidase II activity, and galactosyltransferase (GalT) activity. For example, expressing GntI in a yeast cell producing Man5,GlcNAc2 or Man4,GlcNAc2 N-glycans results in the transfer of a GlcNAc moiety to the Man5,GlcNAc2 or Man4,GlcNAc2 N-glycans such that GlcNAcMan5,GlcNAc2 or GlcNAMan4,GlcNAc2 N-glycans, respectively, are produced. In cells producing GlcNAcMan5,GlcNAc2 N-glycans, expressing a mannosidase II results in two mannose residues being removed from GlcNAcMan5,GlcNAc2 N-glycans to produce GlcNAcMan5,GlcNAc2 N-glycans. In cells producing GlcNAcMan4,GlcNAc2 N-glycans, expressing GntII results in the transfer of another GlcNAc moiety to GlcNAcMan4,GlcNAc2 N-glycans to produce GlcNAcMan4,GlcNAc2 N-glycans. Expressing GalT in cells producing GlcNAcMan5,GlcNAc2 or GlcNAcMan4,GlcNAc2 N-glycans results in the transfer of galactose to the GlcNAcMan5,GlcNAc2 or GlcNAcMan4,GlcNAc2 N-glycans to produce
GalGlcNAcMan3GlcNAc2 or GalαGlcNAcMan3GlcNAc2 N-glycans. In some embodiments, glucosidase H (e.g., by expressing α and β subunits) can be expressed to increase production of the Manα3GlcNAc2 base structure.

Target Molecules

[0064] Target molecules, as used herein, refer to any molecules that undergo N-glycosylation in a genetically engineered cell (e.g., a fungal cell such as Yarrowia lipolytica, Aralula adeninivorans, or other related species dimorphic yeast cell; a plant cell, or an animal cell). In some embodiments, the target molecules are capable of being trafficked through one or more steps of the Yarrowia lipolytica or Aralula adeninivorans (or other related species dimorphic yeast) secretory pathway, resulting in their N-glycosylation by the host cell machinery. The target molecules can be endogenous or exogenous.

[0065] Suitable target proteins include pathogen proteins (e.g., tetanus toxoid; diptheria toxoid; viral surface proteins (e.g., cytomegalovirus (CMV) glycoproteins B, H and G; human immunodeficiency virus 1 (HIV-1) envelope glycoproteins; Rous sarcoma virus (RSV) envelope glycoproteins; herpes simplex virus (HSV) envelope glycoproteins; Epstein Barr virus (EBV) envelope glycoproteins; varicella-zoster virus (VZV) envelope glycoproteins; human papilloma virus (HPV) envelope glycoproteins; Influenza virus glycoproteins; and Hepatitis B family surface antigens), lysosomal proteins (e.g., glucocerebrosidase, ceramide-β, or galactocerebroside), insulin, glucagon, growth factors, cytokines, chemokines, a protein binding to an Fc receptor, antibodies or fragments thereof; or fusions of any of the proteins to antibodies or fragments of antibodies (e.g., protein-Fc). Growth factors include, e.g., vascular endothelial growth factor (VEGF), insulin-like growth factor (IGF), bone morphogenetic protein (BMP), Granulocyte-colony stimulating factor (G-CSF), Granulocyte-macrophage colony stimulating factor (GM-CSF), Nerve growth factor (NGF); a Neurotrophin, Platelet-derived growth factor (PDGF), Erythropoietin (EPO), Thrombopoietin (TPO), Myostatin (GDF-8), Growth Differentiation factor-9 (GDF9), basic fibroblast growth factor (bFGF or FGF2), Epidermal growth factor (EGF), Hepatocyte growth factor (HGF). Cytokines include, e.g., interleukins (e.g., IL-1 to IL-33 such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, or IL-15) and interferons (e.g., interferon β or interferon γ). Chemokines include, e.g., CCL-1 (C) to CCL-39, MCP-1, MIP-1α, MIP-1β, RANTES, C10, MIP-2, MARC, MCP-3, MCP-2, MIP-2, CCL18, MIP-1γ, Eotaxin, MCP-5, MCP-4, NCC-1, CCR10, HCC-1, Leukotactin-1, LEC, NCC-4, TARC, PARC, or Eotaxin-2. Also included are tumor glycoproteins (e.g., tumor-associated antigens), for example, carcinoembryonic antigen (CEA), human mucins, HER-2/neu, and prostate-specific antigen (PSA) [Henderson and Finn, Advances in Immunology, 62, pp. 217-56 (1996)]. In one embodiment, the target protein is an anti-HER2/neu antibody. In some embodiments, the target protein can be associated with a lysosomal storage disorder, which target proteins include, e.g., alpha-L-iduronidase, beta-D-glucosidase, beta-glucoisidase, beta-hexosaminidase, beta-D-mannosidase, alpha-L-fucosidase, arylsulfatase B, arylsulfatase A, alpha-N-acetyl-galactosaminidase, aspartoacylglucosaminidase, iduronate-2-sulfatase, alpha-glucosaminidase-N-acetyltransferase, beta-D-glucuronidase, hyaluronidase, alpha-L-mannosidase, alpha-neuraminidase, phosphotransferase, acid lipase, acid ceramidase, sphingomyelinase, thioesterase, cathepsin K, and lipoprotein lipase.

[0066] Target proteins also can be fusion proteins. Fusions proteins include, e.g., a fusion of (i) any protein described herein or fragment thereof with (ii) an antibody or fragment thereof. As used herein, the term “antibody fragment” refers to (a) an antigen-binding fragment or (b) an Fc part of the antibody that can interact with an Fc receptor. An antigen binding fragment can be, for example, a Fab, F(ab')2, Fv, and single chain Fv (scFv) fragment. An scFv fragment is a single polypeptide chain that includes both the heavy and light chain variable regions of the antibody from which the scFv is derived. In addition, diabodies [Poljak (1994) Structure 2(12):1121-1123; Hudson et al. (1999) J. Immunol. Methods 231(1-2):177-189] and intrabodies [Huston et al. (2001) Hum. Antibodies 10(3-4):127-142; Wheeler et al. (2003) Mol. Ther. 8(3):355-366; Stocks (2004) Drug Discov. Today 9(22): 960-966] can be used in the methods of the invention.

[0067] Target proteins can also be joined to one or more of a polymer, a carrier, an adjuvant, or an immunotoxin, or a detectable (e.g., fluorescent, luminescent, or radioactive) moiety. For example, a target protein can be joined to polyethylene glycol, which can be used to increase the molecular weight of small proteins and/or increase circulation residence time.

Genetically Engineered Cells

[0069] Genetically engineered cells described herein can be used to produce target molecules that contain mammalian-like N-glycans or target molecules that contain intermediates within the mammalian glycosylation pathway. For example, as described herein, nucleic acids encoding one or more enzymes can be introduced into a fungal cell such that the cell produces the desired N-glycan (e.g., GlcNAcMan3GlcNAc2, GlcNAcMan3GlcNAc2, GlcNAcMan3GlcNAc2, GlcNAcMan3GlcNAc2, GalGlcNAcMan3GlcNAc2, or GalαGlcNAcMan3GlcNAc2 N-glycans). Thus, in any of the embodiments described herein, a fungal cell may contain a nucleic acid encoding one enzyme, or a nucleic acid may encode multiple enzymes. Each such nucleic acid also can contain a targeting sequence as discussed below. In addition, a nucleic acid encoding a target molecule also can be introduced into the fungal cell such that the target molecule is produced and modified to contain the desired N-glycan (e.g., GlcNAcMan3GlcNAc2, GlcNAcMan3GlcNAc2, GlcNAcMan3GlcNAc2, GalGlcNAcMan3GlcNAc2, or GalαGlcNAcMan3GlcNAc2 N-glycans).

[0070] The terms “nucleic acid” and “polynucleotide” are used interchangeably herein, and refer to both RNA and DNA, including cDNA, genomic DNA, synthetic DNA, and DNA (or RNA) containing nucleic acid analogs. Nucleic acids can have any three-dimensional structure. A nucleic acid can be double-stranded or single-stranded (i.e., a sense strand or a antisense strand). Non-limiting examples of nucleic acids include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, siRNA, micro-RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers, as well as nucleic acid analogs. “Polypeptide” and “protein” are used inter-
changeably herein and mean any peptide-linked chain of amino acids, regardless of length or post-translational modification.

[0071] An “isolated nucleic acid” refers to a nucleic acid that is separated from other nucleic acid molecules that are present in a naturally-occurring genome, including nucleic acids that normally flank one or both sides of the nucleic acid in a naturally-occurring genome (e.g., a yeast genome). The term “isolated” as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence, since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

[0072] An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., any paramyxovirus, retrovirus, lentivirus, adenovirus, or herpes virus), or into the genome DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not considered an isolated nucleic acid.

[0073] The term “exogenous” as used herein with reference to nucleic acid and a particular host cell refers to any nucleic acid that does not occur in (and cannot be obtained from) that particular cell as found in nature. Thus, a non-naturally-occurring nucleic acid is considered to be exogenous to a host cell once introduced into the host cell. It is important to note that non-naturally-occurring nucleic acids can contain nucleic acid subsequences or fragments of nucleic acid sequences that are found in nature provided that the nucleic acid as a whole does not exist in nature. For example, a nucleic acid molecule containing a genomic DNA sequence within an expression vector is non-naturally-occurring nucleic acid, and thus is exogenous to a host cell once introduced into the host cell, since that nucleic acid molecule as a whole (genomic DNA plus vector DNA) does not exist in nature. Thus, any vector, autonomously replicating plasmid, or virus (e.g., retrovirus, adenovirus, or herpes virus) that as a whole does not exist in nature is considered to be non-naturally-occurring nucleic acid. It follows that genomic DNA fragments produced by PCR or restriction endonuclease treatment as well as cDNAs are considered to be non-naturally-occurring nucleic acid since they exist as separate molecules not found in nature. It also follows that any nucleic acid containing a promoter sequence and polypeptide-encoding sequence (e.g., cDNA or genomic DNA) in an arrangement not found in nature is non-naturally-occurring nucleic acid. A nucleic acid that is naturally-occurring can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of yeast x is an exogenous nucleic acid with respect to a cell of yeast y once that chromosome is introduced into a cell of yeast.

[0074] Cells suitable for genetic engineering include, e.g., fungal cells (e.g., Yarrowia lipolytica or any other related dimorphic yeast cells described herein), plant cells, or animal cells. The cells can be primary cells, immortalized cells, or transformed cells. The cells can be those in an animal, e.g., a non-human mammal. Such cells, prior to the genetic engineering specified herein, can be obtained from a variety of commercial sources and research resource facilities, such as, for example, the American Type Culture Collection (Rockville, Md.).

[0075] Genetic engineering of a cell can include genetic modifications such as: (i) deletion of an endogenous gene encoding a protein having N-glycosylation activity; (ii) introduction of a recombinant nucleic acid encoding a mutant form of a protein (e.g., endogenous or exogenous protein) having N-glycosylation activity (i.e., expressing a mutant protein having an N-glycosylation activity); (iii) introduction or expression of an RNA molecule that interferes with the functional expression of a protein having the N-glycosylation activity; (iv) introduction of a recombinant nucleic acid encoding a wild-type (e.g., endogenous or exogenous) protein having N-glycosylation activity (i.e., expressing a protein having an N-glycosylation activity); or (v) altering the promoter or enhancer elements of one or more endogenous genes encoding proteins having N-glycosylation activity to thus alter the expression of their encoded proteins. RNA molecules include, e.g., small-interfering RNA (siRNA), short hairpin RNA (shRNA), anti-sense RNA, or micro RNA (miRNA). It is understood that item (ii) includes, e.g., replacement of an endogenous gene with a gene encoding a protein having greater N-glycosylation activity relative to the endogenous gene so replaced. Genetic engineering also includes altering an endogenous gene encoding a protein having an N-glycosylation activity to produce a protein having additions (e.g., a heterologous sequence), deletions, or substitutions (e.g., mutations such as point mutations; conservative or non-conservative mutations). Mutations can be introduced specifically (e.g., site-directed mutagenesis or homologous recombination) or can be introduced randomly (for example, cells can be chemically mutagenized as described in, e.g., Newman and Ferro-Novick (1987) J. Cell Biol. 105(4):1587.

[0076] The genetic modifications described herein can result in one or more of (i) an increase in one or more N-glycosylation activities in the genetically modified cell, (ii) a decrease in one or more N-glycosylation activities in the genetically modified cell, (iii) a change in the localization or intracellular distribution of one or more N-glycosylation activities in the genetically modified cell, or (iv) a change in the ratio of one or more N-glycosylation activities in the genetically modified cell. It is understood that an increase in the amount of an N-glycosylation activity can be due to overexpression of one or more genes having N-glycosylation activity, an increase in copy number of an endogenous gene (e.g., gene duplication), or an alteration in the promoter or enhancer of an endogenous gene that stimulates an increase in expression of the protein encoded by the gene. A decrease in one or more N-glycosylation activities can be due to overexpression of a mutant form (e.g., a dominant negative form) of one or more proteins having N-glycosylation altering activities, introduction or expression of one or more interfering RNA molecules that reduce the expression of one or...
more proteins having an N-glycosylation activity, or deletion of one or more endogenous genes that encode a protein having N-glycosylation activity.

[0077] Methods of deleting or disrupting one or more endogenous genes are described in the accompanying Examples. For example, to disrupt a gene by homologous recombination, a "gene replacement" vector can be constructed in such a way to include a selectable marker gene. The selectable marker gene can be operably linked, at both 5' and 3' ends, to portions of the gene of sufficient length to mediate homologous recombination. The selectable marker can be one of any number of genes which either complement host cell auxotrophy or provide antibiotic resistance, including URA3, LEU2 and HIS3 genes. Other suitable selectable markers include the CAT gene, which confers chloramphenicol resistance to yeast cells, or the lacZ gene, which results in blue colonies due to the expression of β-galactosidase. Linearized DNA fragments of the gene replacement vector are then introduced into the cells using methods well known in the art (see below). Integration of the linear fragments into the genome and the disruption of the gene can be determined based on the selection marker and can be verified by, for example, Southern blot analysis.

[0078] As detailed in the accompanying examples, subsequent to its use in selection, a selectable marker can be removed from the genome of the host cell by, e.g., Cre-loxP systems (see below). This process of marker removal is referred to as "curing" throughout the Examples.

[0079] Alternatively, a gene replacement vector can be constructed in such a way as to include a portion of the gene to be disrupted, where the portion is devoid of any endogenous gene promoter sequence and encodes none, or an inactive fragment of, the coding sequence of the gene. An "inactive fragment" is a fragment of the gene that encodes a protein having, e.g., less than about 10% (e.g., less than about 9%, less than about 8%, less than about 7%, less than about 6%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, less than about 1%, or 0%) of the activity of the protein produced from the full-length coding sequence of the gene. Such a portion of the gene is inserted in a vector in such a way that no known promoter sequence is operably linked to the gene sequence, but that a stop codon and a transcription termination sequence are operably linked to the portion of the gene. This vector can be subsequently linearized in the portion of the gene sequence and transformed into a cell. By way of single homologous recombination, this linearized vector is then integrated in the endogenous counterpart of the gene.

[0080] Expression vectors can be autonomous or integrative.

[0081] A recombinant nucleic acid can be introduced into the cell in the form of an expression vector such as a plasmid, phage, transposon, cosmid or virus particle. The recombinant nucleic acid can be maintained extrachromosomally or it can be integrated into the yeast cell chromosomal DNA. Expression vectors can contain selection marker genes encoding proteins required for cell viability under selected conditions (e.g., URA3, which encodes an enzyme necessary for uracil biosynthesis or TRY1, which encodes an enzyme required for tryptophan biosynthesis) to permit detection and/or selection of cells transformed with the desired nucleic acids (see, e.g., U.S. Pat. No. 4,704,562). Expression vectors can also include an autonomous replication sequence (ARS). For example, U.S. Pat. No. 4,837,148 describes autonomous replication sequences which provide a suitable means for maintaining plasmids in Pichia pastoris.

[0082] Integrative vectors are disclosed, e.g., in U.S. Pat. No. 4,882,279. Integrative vectors generally include a serially arranged sequence of at least a first insertable DNA fragment, a selectable marker gene, and a second insertable DNA fragment. The first and second insertable DNA fragments are each about 200 (e.g., about 250, about 300, about 350, about 400, about 450, about 500, or about 1000 or more) nucleotides in length and have nucleotide sequences which are homologous to portions of the genomic DNA of the species to be transformed. A nucleotide sequence containing a gene of interest (e.g., a gene encoding a protein having N-glycosylation activity) for expression is inserted in this vector between the first and second insertable DNA fragments whether before or after the marker gene. Integrative vectors can be linearized prior to yeast transformation to facilitate the integration of the nucleotide sequence of interest into the host cell genome.

[0083] An expression vector can feature a recombinant nucleic acid under the control of a yeast (e.g., Yarrowia lipolytica, Arxula adeninivorans, or other related dimorphic yeast species) promoter, which enables them to be expressed in yeast. Suitable yeast promoters include the TEF1, HP4, GAP, PDX2, ADC1, TP1, ADH2, PDX, and Gal10 promoter. See, e.g., Madzak et al., (2000) J. Mol. Microbiol. Biotechnol. 2:207-216; Guarente et al. (1982) Proc. Natl. Acad. Sci. USA 79(23): 7410. Additional suitable promoters are described in, e.g., Zhu and Zhang (1998) Bioinformatics 15(7-8):608-611 and U.S. Pat. No. 6,265,185. Where the expression vector is to be introduced into an animal cell, such as a mammalian cell, the expression vector can feature a recombinant nucleic acid under the control of an animal cell promoter suitable for expression in the host cell of interest. Examples of mammalian promoters include the SV40 and cytomegalovirus (CMV) promoters.

[0084] A promoter can be constitutive or inducible (conditional). A constitutive promoter is understood to be a promoter whose expression is constant under the standard culturing conditions. Inducible promoters are promoters that are responsive to one or more induction cues. For example, an inducible promoter can be chemically regulated (e.g., a promoter whose transcriptional activity is regulated by the presence or absence of a chemical inducing agent such as an alcohol, tetracycline, a steroid, a metal, or other small molecule) or physically regulated (e.g., a promoter whose transcriational activity is regulated by the presence or absence of a physical inducer such as light or high or low temperatures). An inducible promoter can also be indirectly regulated by one or more transcription factors that are themselves directly regulated by chemical or physical cues.

[0085] Genetic engineering of a cell also includes activating an endogenous gene (e.g., a gene encoding a protein having N-glycosylation activity) that is present in the host cell, but is normally not expressed in the cells or is not expressed at significant levels in the cells. For example, a regulatory sequence (e.g., a gene promoter or an enhancer) of an endogenous gene can be modified such that the operably linked coding sequence exhibits increased expression. Homologous recombination or targeting can be used to replace or disable the regulatory region normally associated with the gene with a regulatory sequence which causes the gene to be expressed at levels higher than evident in the corresponding non-genetically engineered cell, or causes the gene to display a pattern of regulation or induction that is
different than evident in the corresponding non-genetically engineered cell. Suitable methods for introducing alterations of a regulatory sequence (e.g., a promoter or enhancer) of a gene are described in, e.g., U.S. Application Publication No. 20030147868.

It is understood that other genetically engineered modifications also can be conditional. For example, a gene can be conditionally deleted using, e.g., a site-specific DNA recombinase such as the Cre-loxP system (see, e.g., Grossen et al. (2002) *Ann. Rev. Genetics* 36:153-173 and U.S. Application Publication No. 20060014264).

A recombinant nucleic acid can be introduced into a cell described herein using a variety of methods such as the spheroplast technique or the whole-cell lithium chloride yeast transformation method. Other methods useful for transformation of plasmids or linear nucleic acid vectors into cells are described in, for example, U.S. Pat. No. 4,929,555; Hinnen et al. (1978) *Proc. Nat. Acad. Sci. USA* 75:1929; Ito et al. (1983) *J. Bacteriol.* 153:163; U.S. Pat. No. 4,879,231; and Sreekrishna et al. (1987) *Gene* 59:115. Electroportation and PEG1000 whole cell transformation procedures may also be used, as described by Cregg and Russel, Methods in Molecular Biology: *Pichia* Protocols, Chapter 3, Humana Press, Totowa, N.J., pp. 27-39 (1998). Transformation of animal cells can feature, for example, the introduction of a vector to the cells using calcium phosphate, electroporation, heat shock, liposomes, microinjection reagents such as FUGENE® or LIPOFECTAMINE®, or by contacting naked nucleic acid vectors with the cells in solution (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual Second Edition* vol. 1, 2 and 3, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., USA, November 1989.

Transformed yeast cells can be selected for by using appropriate techniques including, but not limited to, culturing auxotrophic cells after transformation in the absence of the biochemical product required (due to the cell’s auxotrophy), selection for and detection of a new phenotype, or culturing in the presence of an antibiotic which is toxic to the yeast in the absence of a resistance gene contained in the transformants. Transformants can also be selected and/or verified by integration of the expression cassette into the genome, which can be assessed by, e.g., Southern blot or PCR analysis.

Prior to introducing the vectors into a target cell of interest, the vectors can be grown (e.g., amplified) in bacterial cells such as *Escherichia coli* (*E. coli*). The vector DNA can be isolated from bacterial cells by any of the methods known in the art which result in the purification of vector DNA from the bacterial milieu. The purified vector DNA can be extracted extensively with phenol, chloroform, and ether, to ensure that no *E. coli* proteins are present in the plasmid DNA preparation, since these proteins can be toxic to mammalian cells.

Genetic engineering, as described herein, can be used to express (e.g., overexpress), introduce modifications into, or delete any number of genes encoding proteins having N-glycosylation activity. Such proteins include, for example, OCH1, ALG3, α-1,3-glucosyltransferase, Gnt1, mannosidase II, Gnt1, glucose isomerase, or Gal T. The genes encoding proteins having N-glycosylation activity can be from any species containing such genes. Exemplary fungal species from which genes encoding proteins having N-glycosylation activity can be obtained include, without limitation, *Pichia anomala*, *Pichia bovis*, *Pichia canadiensis*, *Pichia carsonii*, *Pichia farinose*, *Pichia fermentans*, *Pichia fluxuum*, *Pichia membranaefaciens*, *Pichia membra neaefaciens*, *Candida valida*, *Candida albicans*, *Candida ascalaphidurum*, *Candida amphixiae*, *Candida antarctica*, *Candida antarctica*, *Candida atmosphaerica*, *Candida blattae*, *Candida carpo phila*, *Candida cerambycidae*, *Candida chautadi*, *Candida coryalalis*, *Candida dosseyi*, *Candida dubliniensis*, *Candida ergatensis*, *Candida fructus*, *Candida glabrata*, *Candida fermentati*, *Candida guilliermondii*, *Candida haemulonii*, *Candida insectamens*, *Candida insectorum*, *Candida intermedia*, *Candida jeffersii*, *Candida kefir*, *Candida krusei*, *Candida lusitaniae*, *Candida lysosiphila*, *Candida maltosa*, *Candida membranifaciens*, *Candida milleri*, *Candida oleophila*, *Candida oregonensis*, *Candida parapsilosis*, *Candida quercitrina*, *Candida shehatae*, *Candida tenuis*, *Candida tropicalis*, *Candida tsuchiya*, *Candida sinalaborantium*, *Candida sojae*, *Candida viswanathii*, *Candida utilis*, *Pichia membranaefaciens*, *Pichia silvestris*, *Pichia membra neaefaciens*, *Pichia chodati*, *Pichia membra neaefaciens*, *Pichia membranaefaciens*, *Pichia minuscula*, *Pichia pastoris*, *Pichia pseudopoly morpha*, *Pichia quercium*, *Pichia robertii*, *Pichia saitoi*, *Pichia silvestris*, *Pichia strasburgensis*, *Pichia terricola*, *Pichia vanrijii*, *Pseudozyma antarctica*, *Rhodosporidium toruloides*, *Rhodotorula glutinis*, *Saccharomyces bayanus*, *Saccharomyces bayanus*, *Saccharomyces cerevisiae*, *Saccharomyces bisporus*, *Saccharomyces chevalieri*, *Saccharomyces delbrueckii*, *Saccharomyces exiguis*, *Saccharomyces fennettatii*, *Saccharomyces fragilis*, *Saccharomyces marxianus*, *Saccharomyces mellis*, *Saccharomyces rosei*, *Saccharomyces rouxii*, *Saccharomyces uvarum*, *Saccharomyces我们将保持上述信息的副本。
lis, Torulaspora delbrueckii, Torulaspora globosa, Trigonopsis variabilis, Williopsis californica, Williopsis saturnus, Zygossacharomyces bisporus, Zygossacharomyces melliis, Zygossacharomyces rouxii, or any other fungi (e.g., yeast) known in the art or described herein. Exemplary lower eukaryotes also include various species of Aspergillus including, but not limited to, Aspergillus caesiellis, Aspergillus candidus, Aspergillus carnea, Aspergillus clavatus, Aspergillus deflectus, Aspergillus flavus, Aspergillus funigatus, Aspergillus glaucus, Aspergillus nidulans, Aspergillus niger, Aspergillus ochraceus, Aspergillus oryzae, Aspergillus parasiticus, Aspergillus penicilloides, Aspergillus restrictus, Aspergillus sojae, Aspergillus sydowi, Aspergillus tamari, Aspergillus terreus, Aspergillus stutus, or Aspergillus versicolor. Exemplary protozoal genera from which genes encoding proteins having N-glycosylation activity can be obtained include, without limitation, Blastocerithidium, Cerithidium, Endodipteryx, Herpetomonas, Leishmania, Leptomonas, Phytomonas, Trypanosoma (e.g., T. brucei, T. gambiense, T. rhodesiense, and T. crass); and Wallula. For example, the gene encoding GluT I can be obtained from human (Swiss Protein Accession No. P26572), rat, Arabidopsis, mouse, or Drosophila; the gene encoding GluT II can be obtained from human, rat (Swiss Protein Accession No. Q09326), Arabidopsis, or mouse; the gene encoding Man II can be obtained from human, rat, Arabidopsis, mouse, or Drosophila (Swiss Protein Accession No. Q24451); and the gene encoding GluT can be obtained from human (Swiss Protein Accession No. P15291), rat, mouse, or bovine.

In some embodiments, a genetically engineered cell lacks the OCH1 (GenBank Accession No: AJ563920) gene or gene product (mRNA or protein) thereof. In some embodiments, a genetically engineered cell lacks the ALG3 (GenBank Accession Nos: XM_503488, Genolevers Ref: YALl00003190g) gene or gene product (mRNA or protein) thereof. In some embodiments, a genetically engineered cell expresses a GluNase transferee (e.g., ALG6, GenBank Accession Nos: XM_502922, Genolevers Ref: YALl0017028g) gene. In some embodiments, a genetically engineered cell expresses an α-1,2-mannosidase (e.g., GenBank Accession No: AF212153) protein. In some embodiments, a genetically engineered cell expresses a GluNase transferee (e.g., Swiss Prot Accession No: P26572) protein. In some embodiments, a genetically engineered cell expresses a mannosidase II protein or catalytic domain thereof (e.g., Swiss Prot Accession No: Q24451). In some embodiments, a genetically engineered cell expresses a galactosyltransferase I protein or catalytic domain thereof (e.g., Swiss Prot Accession No: P15291). In some embodiments, the genetically engineered cell expresses a GluNase transferee II protein or catalytic domain thereof (e.g., Swiss Prot Accession No: Q09326). In some embodiments, the genetically engineered cell expresses an alpha or beta subunit (or both the alpha and the beta subunit) of a glucosidase II such as the glucosidase II of Yarrowia lipolytica, Trypanosoma brucei or Aspergillus niger. A genetically engineered cell can have any combination of these modifications.

For example, in some embodiments, a genetically engineered cell can lack the OCH1 gene and express an α-1,2-mannosidase, GlcNAc-transferase I, mannosidase II, and a galactosyltransferase I. In some embodiment, a genetically engineered cell can lack the ALG3 gene, and express an α-1,2-mannosidase, GlcNAc-transferase I, GlcNAc-transferase I, and a galactosyltransferase I. Such a genetically engineered cell further can express an α-1,3-glucosyltransferase and/or express alpha and beta subunits of a glucosidase II and/or lack the OCH1 gene.

One of more of such proteins can be fusion proteins that contain a heterologous targeting sequence. For example, the α-1,2-mannosidase can have an HDEL endoplasmic reticulum (ER)-retention amino acid sequence (see Examples). It is understood that any protein having N-glycosylation activity can be engineered into a fusion protein comprising an HDEL sequence. Other proteins can have heterologous sequences that target the protein to the Golgi apparatus. For example, the first 100 N-terminal amino acids encoded by the yeast Kre2p gene, the first 36 N-terminal amino acids (Swiss Prot Accession No: P38069) encoded by the S. cerevisiae Mnn2 gene, or the first 46 N-terminal amino acids encoded by the S. cerevisiae Mnn2p gene can be used to target proteins to the Golgi. As such, nucleic acids encoding a protein to be expressed in a fungal cell can include a nucleotide sequence encoding a targeting sequence to target the encoded protein to an intracellular compartment. For example, the α-1,2-mannosidase can be targeted to the ER, while the GluD, GluTI, mannosidase, and Gal T can be targeted to the Golgi.

In some embodiments where a protein having N-glycosylation activity is derived from a cell that is of a different type (e.g., of a different species) than the cell into which the protein is to be expressed, a nucleic acid encoding the protein can be codon-optimized for expression in the particular cell of interest. For example, a nucleic acid encoding a protein having N-glycosylation from Trypanosoma brucei can be codon-optimized for expression in a yeast cell such as Yarrowia lipolytica. Such codon-optimization can be useful for increasing expression of the protein in the cell of interest. Methods for codon-optimizing a nucleic acid encoding a protein are known in the art and described in, e.g., Gao et al. (Biotechnol. Prog. 2004) 20(2): 443-448, Kotula et al. (Nat. Biotechn. 1991) 9, 1386-1389, and Bennetzen et al. (J. Biol. Chem. 1982) 257(6):2036-3031. Table 1 shows the codon usage for Yarrowia lipolytica. Data was derived from 2,945,919 codons present in 5,967 coding sequences. The contents of Table 1 were obtained from a Codon Usage Database, which can be found at world wide web at kawazura.or.jp/codon/cgi-bin/showcodon.cgi?species=284591.

<table>
<thead>
<tr>
<th>Yarrowia lipolytica Codon Usage Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU 15.9(46804)</td>
</tr>
<tr>
<td>CCC 20.6(60095)</td>
</tr>
<tr>
<td>CUG 10.4(30576)</td>
</tr>
<tr>
<td>CUA 13.2(38890)</td>
</tr>
<tr>
<td>CUC 22.6(66461)</td>
</tr>
<tr>
<td>CUA 13.2(38890)</td>
</tr>
<tr>
<td>UAA 6.8(20404)</td>
</tr>
<tr>
<td>UAG 9.8(28769)</td>
</tr>
<tr>
<td>UCG 13.2(38890)</td>
</tr>
<tr>
<td>UGC 13.2(38890)</td>
</tr>
<tr>
<td>UGA 9.8(28769)</td>
</tr>
</tbody>
</table>

Where the genetic engineering involves, e.g., changes in the expression of a protein or expression of an exogenous protein (including a mutant form of an endogenous protein), a variety of techniques can be used to determine if the genetically engineered cells express the protein. For example, the presence of mRNA encoding the protein or the protein itself can be detected using, e.g., Northern Blot or RT-PCR analysis or Western Blot analysis, respectively. The intracellular localization of a protein having N-glycosylation activity can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence.

Methods for detecting glycosylation of a target molecule include DNA sequencer-assisted (DNA), fluorophore-assisted carbohydrate electrophoresis (FACE) or surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS). For example, an analysis can utilize DSA-FACE in which, for example, glycoproteins are denatured followed by immobilization on, e.g., a membrane. The glycoproteins can then be reduced with a suitable reducing agent such as dithiothreitol (DTT) or β-mercaptopethanol. The sulfhydryl groups of the proteins can be carbamylated using an acid such as iodoacetamide. Next, the N-glycans can be released from the protein using an enzyme such as N-glycosidase F. N-glycans, optionally, can be reconstituted and derivatized by reductive amination. The derivatized N-glycans can then be concentrated. Instrumentation suitable for N-glycan analysis includes, e.g., the ABI PRISM® 377 DNA sequencer (Applied Biosystems). Data analysis can be performed using, e.g., GENESCANNER® 3.1 software (Applied Biosystems). Optionally, isolated mannos can be further treated with one or more enzymes to confirm their N-glycan status. Additional methods of N-glycan analysis include, e.g., mass spectrometry (e.g., MALDI-TOF-MS), high-pressure liquid chromatography (HPLC) on normal phase, reversed phase and ion exchange chromatography (e.g., with pulsed amperometric detection when glycans are not labeled and with UV absorbance or fluorescence if glycans are appropriately labeled). See also Callewaert et al. (2001) Glycobiology 11(4):275-281 and Freire et al. (2006) Biocatalysis, Chem. 17(2):559-564.

Where any of the genetic modifications of the genetically engineered cell are inducible or conditional on the presence of an inducing cue (e.g., a chemical or physical cue), the genetically engineered cell can, optionally, be cultured in the presence of an inducing agent before, during, or subsequent to the introduction of the nucleic acid. For example, following introduction of the nucleic acid encoding a target protein, the cell can be exposed to a chemical inducing agent that is capable of promoting the expression of one or more proteins having N-glycosylation activity. Where multiple inducing cues induce conditional expression of one or more proteins having N-glycosylation activity, a cell can be contacted with multiple inducing agents.

Target molecules modified to include the desired N-glycan can be isolated from the genetically engineered cell. The modified target molecule can be maintained within the yeast cell and released upon cell lysis or the modified target molecule can be secreted into the culture medium via a mechanism provided by a coding sequence (either native to the exogenous nucleic acid or engineered into the expression vector), which directs secretion of the molecule from the cell. The presence of the modified target molecule in the cell lysate or culture medium can be verified by a variety of standard protocols for detecting the presence of the molecule. For example, where the altered target molecule is a protein, such protocols can include, but are not limited to, immunoblotting or radioimmuno-precipitation with an antibody specific for the altered target protein (or the target protein itself), binding of a ligand specific for the altered target protein (or the target protein itself), or testing for a specific enzyme activity of the modified target protein (or the target protein itself).

In some embodiments, at least about 25% of the target molecules isolated from the genetically engineered cell contain the desired N-glycan. For example, at least about 27%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, or at least about 95%, or at least about 99% of the target molecules isolated from the genetically engineered cell can contain the desired N-glycan.

In some embodiments, in the target molecules produced using the methods described herein, at least 50% (e.g., at least 55, 60, 65, 70, 75, 80, or 85%) of the N-glycans on the glycoprotein can be GlcNAc, Man, GlcNAc, N-glycans. The percentage of GlcNAc, Man, GlcNAc, N-glycans can be estimated from the peak areas in the DSA-FACE electropherograms. See Example 13.

In some embodiments, the isolated modified target molecules can be frozen, lyophilized, or immobilized and
stored under appropriate conditions, e.g., which allow the altered target molecules to retain biological activity.

Cultures of Engineered Cells

[0103] This document also provides a substantially pure culture of any of the genetically engineered cells described herein. As used herein, a “substantially pure culture” of a genetically engineered cell is a culture of that cell in which less than about 40% (i.e., less than about: 35%; 30%; 25%; 20%; 15%; 10%; 5%; 2%; 1%; 0.5%; 0.25%; 0.1%; 0.01%; 0.001%; 0.0001%; or even less) of the total number of viable cells in the culture are viable cells other than the genetically engineered cell, e.g., bacterial, fungal (including yeast), mycoplasma, or protozoan cells. The term “about” in this context means that the relevant percentage can be 15% percent of the specified percentage above or below the specified percentage. Thus, for example, about 20% can be 17% to 23%. Such a culture of genetically engineered cells includes the cells and a growth, storage, or transport medium. Media can be liquid, semi-solid (e.g., gelatinous media), or frozen. The culture includes the cells growing in the liquid or in/on the semi-solid medium or being stored or transported in a storage or transport medium, including a frozen storage or transport medium. The cultures are in a culture vessel or storage vessel or substrate (e.g., a culture dish, flask, or tube or a storage vial or tube).

[0104] The genetically engineered cells described herein can be stored, for example, as frozen cell suspensions, e.g., in buffer containing a cryoprotectant such as glycerol or sucrose, as lyophilized cells. Alternatively, they can be stored, for example, as dried cell preparations obtained, e.g., by fluidized bed drying or spray drying, or any other suitable drying method.

Disorders Treatable by Altered N-Glycosylation Molecules

[0105] The isolated, target molecules modified to contain the desired N-glycan can be used to treat a variety of disorders, including metabolic disorders, cancer, and inflammatory disorders.

[0106] (i) Metabolic Disorders

[0107] A metabolic disorder is one that affects the production of energy within individual human (or animal) cells. Most metabolic disorders are genetic, though some can be “acquired” as a result of diet, toxins, infections, etc. Genetic metabolic disorders are also known as inborn errors of metabolism. In general, the genetic metabolic disorders are caused by genetic defects that result in missing or improperly constructed enzymes necessary for some step in the metabolic process of the cell. The largest classes of metabolic disorders are disorders of carbohydrate metabolism, disorders of amino acid metabolism, disorders of organic acid metabolism (organic acidurias), disorders of fatty acid oxidation and mitochondrial metabolism, disorders of porphyria metabolism, disorders of purine or pyrimidine metabolism, disorders of steroid metabolism disorders of mitochondrial function, disorders of peroxisomal function, and lysosomal storage disorders (LSDs).

[0108] Examples of metabolic disorders that can be treated through the administration of one or more glycosylated molecules (or pharmaceutical compositions of the same) described herein can include hereditary hemochromatosis, oculocutaneous albinism, protein C deficiency, type I hereditary angioedema, congenital sucrose-isomaltase deficiency, Crigler-Najjar type II, Laron syndrome, hereditary Myeloperoxidase, primary hypothyroidism, congenital long QT syndrome, tyroxine binding globulin deficiency, familial hypercholesterolemia, familial chylomicronemia, abetalipoproteinemia, low plasma lipoprotein A levels, hereditary emphysema with liver injury, congenital hypothyroidism, osteogenesis imperfecta, hereditary hypofibrinogenemia, alpha-lantichymotrypsin deficiency, nephrogenic diabetes insipidus, neurohypophyseal diabetes insipidus, adrenoleukodystrophy, type II diabetes, Batten disease, ataxia telangiectasia, ADPKD-autosomal dominant polycystic kidney disease, microvillus inclusion disease, tuberous sclerosis, ocuculocerebrorenal syndrome of Lowe, amyotrophic lateral sclerosis, myelodysplastic syndrome, Buer lymphocyte syndrome, Tangier disease, familial intrahepatic cholestasis, X-linked adrenoleukodystrophy, Scott syndrome, Hermansky-Pudlak syndrome types 1 and 2, Zellweger syndrome, rhizomelic chondrodysplasia puncta, autosomal recessive primary hyperoxaluria, Mohr-Tranebjaerg syndrome, spinal and bulbar muscular atrophy, primary ciliary dyskinesias (Kartagener’s syndrome), giantism and acromegaly, galactocerebrodysplasia, Addison’s disease, adrenal virilism, Cushing’s syndrome, ketoadiposis, primary or secondary aldosteronism, Miller-Dieker syndrome, lissencephaly, motor neuron disease, Usher’s syndrome, Wiskott-Aldrich syndrome, Optiz syndrome, Huntington’s disease, hereditary pancreatitis, anti-phospholipid syndrome, overlap connective tissue disease, Sjögren’s syndrome, stiff-man syndrome, Brugada syndrome, congenital nephritic syndrome of the Finnish type, Dubin-Johnson syndrome, X-linked hypophosphatemia, Pendred syndrome, persistent hyperinsulinemic hypoglycemia of infancy, hereditary spherocytosis, aceruloplasminemia, infantile neuronal ceroid lipofuscinosis, pseudohypoparathyroidism and multiple epiphyseal dysplasia, Stargardt-like macular dystrophy, X-linked Charcot-Marie-Tooth disease, autosomal dominant retinitis pigmentosa, Wolkott-Rallison syndrome, Cushing’s disease, limb-girdle muscular dystrophy, mucopoly-saccharidosis type IV, hereditary familial amyloidosis of Finisc, Anderson disease, sarcoma, chronic myelomonocytic leukemia, cardiomyopathy, facioscapulohumeral dystrophy, Torsion disease, Huntington and spinocerebellar ataxias, hereditary hyperhomocysteinemia, polynuropathy, lower motor neuron disease, pigmented retinitis, seronegative polychratisis, interfasciolar pulmonary fibrosis, Raynaud’s phenomenon, Wegner’s granulomatosis, preeclampsia, CDG-Ia, CDG-Ib, CDG-Ie, CDG-Id, CDG-Ie, CDG-II, CDG-IIa, CDG-IIb, CDG-IIc, CDG-IId, Ehrler-Danlos syndrome, multiple exostoses, Griscelli syndrome (type 1 or type 2), or X-linked non-specific mental retardation. In addition, metabolic disorders can also include lysosomal storage disorders such as, but not limited to, Fabry disease, Faber disease, Gaucher disease, GM1-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM2 activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease (types A, B, and C), Hurler disease, Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, siaidosis type 1, Pompe disease, Pec-nodysostosis, ceroid lipofuscinosis, cholesterol ester storage disease, Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucolipidosis (types II and IV), cystinosis,
sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Geleophysic dysplasia.

[0109] Symptoms of a metabolic disorder are numerous and diverse and can include one or more of, e.g., anemia, fatigue, bruising easily, low blood platelets, liver enlargement, spleen enlargement, skeletal weakening, lung impairment, infections (e.g., chest infections or pneumonias), kidney impairment, progressive brain damage, seizures, extra thick ceroid, coughing, wheezing, excess saliva or mucus production, shortness of breath, abdominal pain, oculated bowel or gut, fertility problems, polyph in the nose, clubbing of the finger/toe nails and skin, pain in the hands or feet, angiokeratoma, decreased perspiration, corneal and lenticular opacities, cataracts, mitral valve prolapse and/or regurgitation, cardiomegaly, temperature intolerance, difficulty walking, difficulty swallowing, progressive vision loss, progressive hearing loss, hypotonia, macroglossia, areflexia, lower back pain, sleep apnea, orthopnea, somnolence, lordosis, or scoliosis. It is understood that due to the diverse nature of the defective or absent proteins and the resulting disease phenotypes (e.g., symptomatic presentation of a metabolic disorder), a given disorder will generally present only symptomatics characteristic to that particular disorder. For example, a patient with Fabry disease can present a particular subset of the above-mentioned symptoms such as, but not limited to, temperature intolerance, corneal whirling, pain, skin rashes, nausea, or diarrhea. A patient with Gaucher syndrome can present with splenomegaly, cirrhosis, convulsions, hypotonia, apnea, osteoporosis, or skin discoloration.

[0110] In addition to the administration of one or more molecules described herein, a metabolic disorder can also be treated by proper nutrition and vitamins (e.g., cofactor therapy), physical therapy, and pain medications.

[0111] Depending on the specific nature of a given metabolic disorder, a patient can present these symptoms at any age. In many cases, symptoms can present in childhood or in early adulthood. For example, symptoms of Fabry disease can present at an early age, e.g., at 10 or 11 years of age.

[0112] As used herein, a subject “at risk of developing a metabolic disorder” is a subject that has a predisposition to develop a disorder, i.e., a genetic predisposition to develop metabolic disorder, or a result of a mutation in a enzyme such as alpha-L-iduronidase, beta-D-galactosidase, beta-glucosidase, beta-hexosaminidase, beta-D-mannosidase, alpha-L-fucosidase, arylsulfatase B, arylsulfatase A, alpha-N-acetylgalactosaminidase, aspartylglucosaminidase, iduronate-2-sulfatase, alpha-glucosaminidase-N-acetytransferase, beta-D-glucuronidase, hyaluronidase, alpha-L-mannosidase, alpha-neuraminidase, phosphotransferase, acid lipase, acid ceramidase, sphingomyelinase, thioesterase, cathepsin K, or lipoprotein lipase. Clearly, subjects “at risk of developing a metabolic disorder” are not all the subjects within a species of interest.

[0113] A subject “suspected of having a disorder” is one having one or more symptoms of a disorder such as any of those described herein.

[0114] (ii) Cancer

[0115] Cancer is a class of diseases or disorders characterized by uncontrolled division of cells and the ability of these to spread, either by direct growth into adjacent tissue through invasion, or by implantation into distant sites by metastasis (where cancer cells are transported through the bloodstream or lymphatic system). Cancer can affect people at all ages, but risk tends to increase with age. Types of cancers can include, e.g., lung cancer, breast cancer, colon cancer, pancreatic cancer, renal cancer, stomach cancer, liver cancer, bone cancer, hematological cancer, neural tissue cancer, melanoma, thyroid cancer, ovarian cancer, testicular cancer, prostate cancer, cervical cancer, vaginal cancer, or bladder cancer.

[0116] As used herein, a subject “at risk of developing a cancer” is a subject that has a predisposition to develop a cancer, i.e., a genetic predisposition to develop cancer such as a mutation in a tumor suppressor gene (e.g., mutation in BRCA1, p53, RB, or APC) or has been exposed to conditions that can result in cancer. Thus, a subject can also be one “at risk of developing a cancer” when the subject has been exposed to mutagenic or carcinogenic levels of certain compounds (e.g., carcinogenic compounds in cigarette smoke e.g., Acrolein, Arsenic, Benzene, Benz[a]anthracene, Benzo[a]pyrene, Polonium-210 (Radium), Urethane, or Vinyl Chloride). Moreover, the subject can be “at risk of developing a cancer” when the subject has been exposed to, e.g., large doses of ultraviolet light or X-irradiation, or exposed (e.g., infected) to a tumor-causing/associated virus such as papillomavirus, Epstein-Barr virus, hepatitis B virus, or human T-cell leukemia-lymphoma virus. From the above it will be clear that subjects “at risk of developing a cancer” are not all the subjects within a species of interest.

[0117] A subject “suspected of having a cancer” is one having one or more symptoms of a cancer. Symptoms of cancer are well-known to those of skill in the art and include, without limitation, breast lumps, nipple changes, breast cysts, breast pain, weight loss, weakness, excessive fatigue, difficulty eating, loss of appetite, chronic cough, worsening breathlessness, coughing up blood, blood in the urine, blood in stool, nausea, vomiting, liver metastases, lung metastases, bone metastases, abdominal fullness, bloating, fluid in peritoneal cavity, vaginal bleeding, constipation, abdominal distension, perforation of colon, acute peritonitis (infection, fever, pain), pain, vomiting blood, heavy sweating, fever, high blood pressure, anemia, diarrhea, jaundice, dizziness, chills, muscle spasms, colon metastases, lung metastases, bladder metastases, liver metastases, bone metastases, kidney metastases, and pancreas metastases, difficulty swallowing, and the like. From the above it will be clear that subjects “suspected of having a cancer” are not all the subjects within a species of interest.

[0118] In addition to the administration of one or more altered N-glycosylation molecules described herein, a cancer can also be treated by chemotherapeutic agents, ionizing radiation, immunotherapy agents, or hyperthermotherapy agents. Chemotherapeutic agents include, e.g., cisplatin, carboplatin, procarbazine, meclofenathamine, cyclophosphamide, camptothecin, adriamycin, ifosfamide, melphalan, chlorambucil, bisulfan, nitrosurea, dacarbazine, daunorubicin, doxorubicin, bleomycin, plicomycin, mitomycin, etoposide, vinblastin, podophyllotoxin, tamoxifen, taxol, transplatinum, 5-fluorouracil, vincristin, viabilitin, and methotrexate.

[0119] (iii) Inflammatory Disorders

[0120] An “inflammatory disorder,” as used herein, refers to a process in which one or more substances (e.g., substances not naturally occurring in the subject), via the action of white blood cells (e.g., B cells, T cells, macrophages, monocytes, or dendritic cells) inappropriately trigger a pathological response, e.g., a pathological immune response. Accordingly, such cells involved in the inflammatory response are referred
to as “inflammatory cells.” The inappropriately triggered inflammatory response can be one where no foreign substance (e.g., an antigen, a virus, a bacterium, a fungus) is present in or on the subject. The inappropriately triggered response can be one where a self-component (e.g., a self-antigen) is targeted (e.g., an autoimmune disorder such as multiple sclerosis) by the inflammatory cells. The inappropriately triggered response can also be a response that is inappropriate in magnitude or duration, e.g., anaphylaxis. Thus, the inappropriately targeted response can be due to the presence of a microbial infection (e.g., viral, bacterial, or fungal). Types of inflammatory disorders (e.g., autoimmune disease) can include, but are not limited to, osteoarthritis, rheumatoid arthritis (RA), spondyloarthropathies, POEMS syndrome, Crohn’s disease, multicentric Castleman’s disease, systemic lupus erythematosus (SLE), multiple sclerosis (MS), muscular dystrophy (MD), insulin-dependent diabetes mellitus (IDDM), dermatomyositis, polymyositis, inflammatory neuropsychiatric such as Guillain Barre syndrome, vasculitis such as Wegener’s granulomatosis, polyarteritis nodosa, polymyalgia rheumatica, temporal arteritis, Sjogren’s syndrome, Behchet’s disease, Churg-Strauss syndrome, or Takayasu’s arteritis. Also included in inflammatory disorders are certain types of allergies such as rhinitis, sinusitis, urticaria, hives, angioedema, atopic dermatitis, food allergies (e.g., a nut allergy), drug allergies (e.g., penicillin), insect allergies (e.g., allergy to a bee sting), or mastocytosis. Inflammatory disorders can also include ulcerative colitis and asthma.

[0121] A subject “at risk of developing an inflammatory disorder” refers to a subject with a family history of one or more inflammatory disorders (e.g., a genetic predisposition to one or more inflammatory disorders) or one exposed to one or more inflammation-inducing conditions. For example, a subject can have been exposed to a viral or bacterial superantigen such as, but not limited to, staphylococcal enterotoxins (SEs), a streptococcus pyogenes exotoxin (SPE), a staphylococcus aureus toxic shock syndrome toxin (TSST-1), a streptococcal mitogenic exotoxin (SME) and a streptococcal superantigen (SSA). From the above it will be clear that subjects “at risk of developing an inflammatory disorder” are not all the subjects within a species of interest.

[0122] A subject “suspected of having an inflammatory disorder” is one who presents with one or more symptoms of an inflammatory disorder. Symptoms of inflammatory disorders are well known in the art and include, but are not limited to, redness, swelling (e.g., swollen joints), joints that are warm to the touch, joint pain, stiffness, loss of joint function, fever, chills, fatigue, loss of energy, headaches, loss of appetite, muscle stiffness, insomnia, itchiness, stuffy nose, sneezing, coughing, one or more neurologic symptoms such as dizziness, seizures, or pain. From the above it will be clear that subjects “suspected of having an inflammatory disorder” are not all the subjects within a species of interest.

[0123] In addition to the administration of one or more molecules described herein, an inflammatory disorder can also be treated by non-steroidal anti-inflammatory drug (NSAID), a disease-modifying anti-rheumatic drug (DMARD), a biological response modifier, or a corticosteroid. Biological response modifiers include, e.g., an anti-TNF agent. Non-limiting examples of anti-TNF agents include a soluble TNF receptor or an antibody specific for TNF such as adalimumab, infliximab, or etanercept.

[0124] Methods suitable for treating (e.g., preventing or ameliorating one or more symptoms of) any of the disorders described herein using any of the altered N-glycosylation molecules (or pharmaceutical compositions thereof) are set forth in the following section.

Pharmaceutical Compositions and Methods of Treatment

[0125] A target molecule modified to have the desired N-glycan can be incorporated into a pharmaceutical composition containing a therapeutically effective amount of the molecule and one or more adjuvants, excipients, carriers, and/or diluents. Acceptable diluents, carriers and excipients typically do not adversely affect a recipient’s homeostasis (e.g., electrolyte balance). Acceptable carriers include biocompatible, inert or bioabsorbable salts, buffering agents, oligo- or polysaccharides, polymers, viscosity-improving agents, preservatives and the like. One exemplary carrier is physiologic saline (0.15 M NaCl, pH 7.0 to 7.4). Another exemplary carrier is 50 mM sodium phosphate, 100 mM sodium chloride. Further details on techniques for formulation and administration of pharmaceutical compositions can be found in, e.g., Remington’s Pharmaceutical Sciences (Mack Publishing Co., Easton, Pa.). Supplementary active compounds can also be incorporated into the compositions.

[0126] Administration of a pharmaceutical composition containing molecules with N-glycans can be systemic or local. Pharmaceutical compositions can be formulated such that they are suitable for parenteral and/or non-parenteral administration. Specific administration modalities include subcutaneous, intravenous, intramuscular, intraperitoneal, transdermal, intra-thecal, oral, rectal, buccal, topical, nasal, ophthalmic, intra-articular, intra-arterial, sub-arachnoid, bronchial, lymphatic, vaginal, and intra-uterine administration.

[0127] Administration can be by periodic injections of a bolus of the pharmaceutical composition or can be uninterrupted or continuous by intravenous or intraperitoneal administration from a reservoir which is external (e.g., an IV bag) or internal (e.g., a bioerodable implant, a bioartificial organ, or a colony of implanted altered N-glycosylation molecule production cells). See, e.g., U.S. Pat. Nos. 4,407,957, 5,798,113, and 5,800,828. Administration of a pharmaceutical composition can be achieved using suitable delivery means such as a pump (see, e.g., Annals of Pharmacotherapy, 27:912 (1993); Cancer, 41:1270 (1993); Cancer Research, 44:1698 (1984); microencapsulation (see, e.g., U.S. Pat. Nos. 4,352,883; 4,353,888; and 5,084,350); continuous release polymer implants (see, e.g., Sabel, U.S. Pat. No. 4,883,666); microencapsulation (see, e.g., U.S. Pat. Nos. 5,284,761, 5,158,881, 4,976,859 and 4,986,733 and published PCT patent applications WO92/19195, WO 95/05452); injection, either subcutaneously, intravenously, intra-arterially, intramuscularly, or to other suitable site; or oral administration, in capsule, liquid, tablet, pill, or prolonged release formulation.

[0128] Examples of parenteral delivery systems include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantaable infusion systems, pump delivery, encapsulated cell delivery, liposomal delivery, needle-delivered injection, needle-less injection, nebulizer, aerosolizer, electroproporation, and transdermal patch.

[0129] Formulations suitable for parenteral administration conveniently contain a sterile aqueous preparation of the altered N-glycosylation molecule, which preferably is isotonic with the blood of the recipient (e.g., physiological saline solution). Formulations can be presented in unit-dose or multi-dose form.
Formulations suitable for oral administration can be presented as discrete units such as capsules, cachets, tablets, or lozenges, each containing a predetermined amount of the altered N-glycosylation molecule; or a suspension in an aqueous liquid or a non-aqueous liquid, such as a syrup, an elixir, an emulsion, or a draught.

A molecule having N-glycans suitable for topical administration can be administered to a mammal (e.g., a human patient) as, e.g., a cream, a spray, a foam, a gel, an ointment, a salve, or a dry rub. A dry rub can be rehydrated at the site of administration. Such molecules can also be infused directly into (e.g., soaked into and dried) a bandage, gauze, or patch, which can then be applied topically. Such molecules can also be maintained in a semi-liquid, gelled, or fully-liquid state in a bandage, gauze, or patch for topical administration (see, e.g., U.S. Pat. No. 4,307,717).

Therapeutically effective amounts of a pharmaceutical composition can be administered to a subject in need thereof in a dosage regimen ascertainable by one of skill in the art. For example, a composition can be administered to the subject, e.g., systemically at a dosage from 0.01 μg/kg to 10,000 μg/kg body weight of the subject, per dose. In another example, the dosage is from 1 μg/kg to 100 μg/kg body weight of the subject, per dose. In another example, the dosage is from 1 μg/kg to 30 μg/kg body weight of the subject, per dose, e.g., from 5 μg/kg to 10 μg/kg body weight of the subject, per dose.

In order to optimize therapeutic efficacy, a molecule containing an N-glycan can be first administered at different dosing regimens. The unit dose and regimen depend on factors that include, e.g., the species of mammal, its immune status, the body weight of the mammal. Typically, levels of such a molecule in a tissue can be monitored using appropriate screening assays as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen.

The frequency of dosing for a molecule is within the skills and clinical judgement of medical practitioners (e.g., doctors or nurses). Typically, the administration regime is established by clinical trials which may establish optimal administration parameters. However, the practitioner may vary such administration regimes according to the subject’s age, health, weight, sex and medical status. The frequency of dosing can be varied depending on whether the treatment is prophylactic or therapeutic.

Toxicity and therapeutic efficacy of such molecules or pharmaceutical compositions thereof can be determined by known pharmaceutical procedures in, for example, cell cultures or experimental animals. These procedures can be used, e.g., for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Pharmaceutical compositions that exhibit high therapeutic indices are preferred. While pharmaceutical compositions that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to normal cells (e.g., non-target cells) and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in appropriate subjects (e.g., human patients). The dosage of such pharmaceutical compositions lies generally within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For a pharmaceutical composition used as described herein (e.g., for treating a metabolic disorder in a subject), the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the 1050 (i.e., the concentration of the pharmaceutical composition which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

As defined herein, a “therapeutically effective amount” of a molecule containing an N-glycan is an amount of the molecule that is capable of producing a medically desirable result (e.g., amelioration of one or more symptoms of a metabolic disorder) in a treated subject. A therapeutically effective amount (i.e., an effective dosage) can includes milligram or microgram amounts of the compound per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram).

The subject can be any mammal, e.g., a human (e.g., a human patient) or a non-human primate (e.g., chimpanzee, baboon, or monkey), a mouse, a rat, a rabbit, a guinea pig, a gerbil, a hamster, a horse, a type of livestock (e.g., cow, pig, sheep, or goat), a dog, a cat, or a whale.

A molecule or pharmaceutical composition thereof described herein can be administered to a subject as a combination therapy with another treatment, e.g., a treatment for a metabolic disorder (e.g., a lysosomal storage disorder). For example, the combination therapy can include administering to the subject (e.g., a human patient) one or more additional agents that provide a therapeutic benefit to the subject who has, or is at risk of developing, (or suspected of having) a metabolic disorder (e.g., a lysosomal storage disorder). Thus, the compound or pharmaceutical composition and the one or more additional agents can be administered at the same time. Alternatively, the molecule can be administered first and the one or more additional agents administered second, or vice versa.

It will be appreciated that in instances where a previous therapy is particularly toxic (e.g., a treatment for a metabolic disorder with significant side-effect profiles), administration of a molecule described herein can be used to offset and/or lessen the amount of the previously therapy to a level sufficient to give the same or improved therapeutic benefit, but without the toxicity.

Any of the pharmaceutical compositions described herein can be included in a container, pack, or dispenser together with instructions for administration.

The following are examples of the practice of the invention. They are not to be construed as limiting the scope of the invention in any way.

**EXAMPLES**

Table 2 contains a list of all of the strains used in the experiments described below. In Table 2, MH–HDEL–tagged α-1,2-mannosidase; ξ–random integration via zeta sequences, docking Ψ–integration into a specific locus; and (H)–hygromycin resistant.
<table>
<thead>
<tr>
<th>Number</th>
<th>Short name</th>
<th>Description</th>
<th>Markers</th>
<th>Expected N-Glycans</th>
</tr>
</thead>
<tbody>
<tr>
<td>G013</td>
<td>PoiId Imaga Δocho1 (URA3) cl 26.1</td>
<td>PoiId Imaga transformed with SpeBlI-1107P-digested pYLOCHIEPETOPO</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Mainly Man&lt;sub&gt;3&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G014</td>
<td>PoiId Imaga Δocho1 (cured) cl 7</td>
<td>PoiId Imaga Δocho1 (G013) cured from the URA3 marker using pUB4-Cer</td>
<td>ura3&lt;sup&gt;−&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Mainly Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G016</td>
<td>PoiId Imaga Δocho1 TefM1 (ΔNot) cl 1.4</td>
<td>PoiId Imaga Δocho1 (cured) (G014) transformed with NotI-digested pY115XL2preManHDEL(Y1)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G018</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot) cl 11.2</td>
<td>PoiId Imaga Δocho1 (cured) (G014) transformed with NotI-digested pY115XL2preManHDEL(Y1)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G036</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 2.2 (1 copy Man1DEL)</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Not) cl 11.2 (G018) cured from the URA3 marker using pBRQ2</td>
<td>ura3&lt;sup&gt;−&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G039</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 24.1</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 2.2 (G036) transformed with NotI-digested pY1A1G3P7-ALG6</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G040</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) TefhGlnTII cl 5.4 (1 copy Gnt1)</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 22.6 (G036) transformed with NotI-digested pY131XmK86GlnTII</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G043</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) TefManIII cl 15</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 22.6 (G043) transformed with NotI-digested pY131XmK86ManIII (LEU2 ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; LEU2&lt;sup&gt;+&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G044</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) TefGalTII cl 12</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 22.6 (G044) transformed with NotI-digested pY131XmK86GalTT (ADE2 ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ADE2&lt;sup&gt;+&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>GalGlcNAcMan&lt;sub&gt;3&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G045</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 2.16 (cured)</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (ΔNot)(cured) 24.1 (G039) cured from the URA3 marker using pBRQ2</td>
<td>ura3&lt;sup&gt;−&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and Man&lt;sub&gt;4&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G046</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (docking Δleu2)</td>
<td>PoiId Imaga Δocho1 (cured) (G014) transformed with NotI-digested JME926 pPTieu2- ADE2ex-Hpd4ManHDEL(Y1)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt; Hyg&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G047</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 2.16 (cured) TefhGlnTI clone 7.3</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 24.1 (G039) transformed with NotI-digested pY131XmK86GlnTII (Hyp&lt;sup&gt;6&lt;/sup&gt; ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G048</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 2.16 (cured) TefhGlnTI TefhGlnTII cl. 24.3</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 24.1 (G039) transformed with NotI-digested pY131XmK86GlnTII and pY131XmK86GlnTII (ADE2 Ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt; Hyg&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G050</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 2.16 (cured) TefhGlnTI TefhGlnTII e12 (G047)</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 24.1 (G039) transformed with NotI-digested pY131XmK86GlnTII and pY131XmK86GlnTII (ADE2 Ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt; Hyg&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G051</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 2.16 (cured) Hpd4bGlnTI (Δc) cl 4.5</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 24.1 (G039) transformed with NotI-digested pY131XmK86GlnTII and pY131XmK86GlnTII (ADE2 Ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt; Hyg&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>G052</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) Hpd4bGlnTI (Δc) cl one 16</td>
<td>PoiId Imaga Δocho1 Hpd4M1 (Δcured) 24.1 (G039) transformed with NotI-digested pY131XmK86GlnTII and pY131XmK86GlnTII (ADE2 Ex)</td>
<td>URA3&lt;sup&gt;+&lt;/sup&gt; leu2&lt;sup&gt;−&lt;/sup&gt; ade2&lt;sup&gt;−&lt;/sup&gt; gat2&lt;sup&gt;−&lt;/sup&gt;</td>
<td>Glc&lt;sub&gt;1&lt;/sub&gt;Man&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt; and GlcNAcMan&lt;sub&gt;2&lt;/sub&gt;GlcNAc&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>Number</td>
<td>Short name</td>
<td>Description</td>
<td>Markers</td>
<td>Expected N-Glycans</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
</tr>
<tr>
<td>G053</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (docking Δααρβ)</td>
<td>PoI'd ΔImga Δsccl1 (cured) (G018) transformed with NotI-digested OXYYP289-pFTAgp1-Len2Ex-Hp4dMHΔDIEL(F107)</td>
<td>ura3·· LEU2· ade2· gat2·</td>
<td>Man3,GlcNAc2</td>
</tr>
<tr>
<td>G054</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (docking Δααρβ) (cured)</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (docking Δααρβ1) (G053) cured from the LEU2 marker using pUBH-Cre</td>
<td>ura3·· leu2· ade2· gat2·</td>
<td>Man3,GlcNAc2</td>
</tr>
<tr>
<td>G055</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (docking Δαεε2·cured)</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (docking Δαεε2·) (G046) cured from the ADE2 marker using pRRQ2</td>
<td>ura3·· leu2· ade2· gat2·</td>
<td>Man3,GlcNAc2</td>
</tr>
<tr>
<td>G056</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (Δεεε) clone E</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested pYLIppmΔXα3GmTS</td>
<td>URA3·· leu2· ade2· gat2·</td>
<td>Glc13,Man3,GlcNAc2 and GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G057</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2·) clone G</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested JME925 pF∂βa-2·URA3ex·Hp4dGmTS</td>
<td>URA3·· leu2· ade2· gat2·</td>
<td>Glc13,Man3,GlcNAc2 and GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G058</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (Δεεε·cured)</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (Δεεε) (G046) cured from the URA3 marker using pRRQ2</td>
<td>ura3·· leu2· ade2· gat2·</td>
<td>Man3,GlcNAc2 and GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G059</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2·cured)</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2) (G057) cured from the URA3 marker using pRRQ2</td>
<td>ura3·· leu2· ade2· gat2·</td>
<td>Man3,GlcNAc2 and GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G060</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (Δεεε) clone 6</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2) (G057) transformed with NotI-digested Zeta-LEU2Ex-Hp4dG2αβExAnGmTS a + b (alt)</td>
<td>URA3·· LEU2· ade2· gat2·</td>
<td>GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G061</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3) clone 18</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (Δεεε·cured) Δalg3ALG6 Hp4dGmTI (docking Δαεε2) (G057) transformed with NotI-digested JME925 pF∂βa-2·Len2Ex-Hp4dL2αβExAnGmTS a + b (alt)</td>
<td>URA3·· LEU2· ade2· gat2·</td>
<td>GlcNAcMan,GlcNAc3</td>
</tr>
<tr>
<td>G070</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3) Hp4dGmTI (Δεεε) cl 6</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3) (G091) transformed with NotI-digested pYLIppmΔXα3GmTS</td>
<td>URA3·· LEU2· ADE2· gat2·</td>
<td>GlcNAc3,Man3,GlcNAc2</td>
</tr>
<tr>
<td>G071</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3·) Hp4dGmTI (integration in Asp1 locus was added) cl 8</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3·) (G091) transformed with NotI-digested OXYYP289-pF∂αp1-ADE2Ex-Hp4dGmTI</td>
<td>URA3·· LEU2· ADE2· gat2·</td>
<td>GlcNAc3,Man3,GlcNAc2</td>
</tr>
<tr>
<td>G096</td>
<td>PoI'd ΔImga Δsccl1 Hp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3·) Hp4dGmTI (integration in Asp1 locus was added) Δ Hp4dPP-HC/CL clone 13</td>
<td>PoI'd ΔImga Δsccl1 ΔHp4dMH (1 copy··Δεεε·cured) Δalg3ALG6 (cured) ΔHp4dGmTI (docking Δαεε2·) ΔHp4dG6s2αβ· (docking Δαρα3·) Δ Hp4dGmTI (docking Δααρβ) (G071) transformed with NotI-digested pYLIppmΔXα3GmTI (ADE2Ex)</td>
<td>URA3·· LEU2· ADE2· GUT2·</td>
<td>GlcNAc3,Man3,GlcNAc2</td>
</tr>
</tbody>
</table>
Example 1

Yarrowia lipolytica OCH1 Disruption

[0144] The generation of a glyco-engineered protein expression strain was done in Yarrowia lipolytica strain pO1L1 nuga (a strain having the auxotrophies leu2-, ura3-, gln2- and ade2-). A strategy to knock out the OCH1 (GenBank Accession No: AJ563920) gene in Yarrowia lipolytica was set up as described for the lip2 gene (Fickers et al., 2003 J Microbiol Methods, 55(3):727-37). The gene construction strategy followed for the OCH1 gene is described in U.S. Patent Publication No. 20090069932-A1. The resulting vector was called pYIOCH1 PUT TOPO (FIG. 1B).

[0145] The OCH1 KO fragment was isolated from the plasmid by a Spel/Bst1 1071 restriction digest and transformed to Yarrowia lipolytica strain pO1L1 nuga. Several uncll prototrophic strains were obtained by PCR on genomic DNA (gDNA) using primers Yloch1 prom fw (5’-TGGCTATACGCTTCAGGC-3’, SEQ ID NO:1) and Yloch1 term rev (5’-CTCTGATATGCTTGATAGTGC-3’, SEQ ID NO:2) to analyze the genomic integration of the plasmid. A fragment of the correct size (i.e., 2328 by vs. 1894 bp in the wild type) was amplified for several clones tested. The knock-out of the OCH1 gene also was confirmed by N-glycan analysis of the total glycoprotein pool secreted into the growth medium (=secretome): the Manα2GlcNAc2 structure has become the predominant N-glycan within the sugar profile (FIG. 2). This profile differs from that of the wild-type strain, which contains a higher amount of Manα2GlcNAc2—the latter most probably containing an additional mannose as a result of Och1p activity—as well as some structures with an even higher number of mannose residues.

[0146] To remove the URA3 gene, a positive Aoch1 clone (called G013, see Table 2) was transformed with the episomal plasmid pUB4-Cre (Fickers et al., 2003, supra) that contains an expression cassette for the Cre recombinase. Removal of the URA3 gene was screened for by PCR on gDNA using primers Yloch1 prom fw and Yloch1 term rev (see above). Clones in which the URA3 marker was excised no longer resulted in the amplification of a 2328 bp band; instead a PCR-fragment of 1075 bp (excl. URA3) was obtained. Positive clones were checked at the N-glycan level of the secretome and show a profile very similar to that of the non-cured strain (FIG. 2). One of the cured strains (called G014, see Table 2) was selected for further N-glycan engineering.

Example 2

Overexpression of an ER-Retained α-1,2-Mannosidase by Either Random Integration or Targeted/Docked Integration

[0147] To enable the generation of Manα2GlcNAc2 attached to glycoproteins expressed by a Aoch1 strain, an α-1,2-mannosidase was expressed to cleave Manα2GlcNAc2 to Manα2GlcNAc2 (i.e., a Golgi type α-1,2-mannosidase activity). Such a mannosidase should be targeted to the secretory system. Trichoderma reesei α-1,2-mannosidase (Genbank accession no. AF212153), fused to the S. cerevisiae prepore mating factor and tagged with a HDEL signal (SEQ ID NO:21) to localize it into the ER, is able to trim Manα2GlcNAc2 to Manα2GlcNAc2 in vivo in Pichia pastoris as well as in Trichoderma reesei and Aspergillus niger. Expression constructs were made where a codon-optimized version of the HDEL-tagged T. reesei α-1,2-mannosidase was fused to the Y. lipolytica LIP2 pre signal sequence and placed under the transcriptional control of either the TEF1, Hpd4 (Madzak et al., 2000, J. Mol. Microbiol. Biotechnol. 2:207-216), GAP or PDX2 promoter. The construction strategy of these plasmids is described in U.S. Patent Publication No. 20090069232-A1.

[0148] Two of these vectors, pYLHUXd1.2preManHDEL and pYLTXd1.2preManHDEL (FIG. 3)—with the mannosidase under the transcriptional control of the Hpd4 resp. TEF1 promoter, were used to transform strain G014 (derived from Example 1). The vectors were digested with NotI to allow random integration into the genome via the zeta sequences. URA3 prototrophic transformants were selected for N-glycan analysis. Several transformants show a clear conversion of Manα2GlcNAc2 towards Manα2GlcNAc3 (FIG. 4). Since clones expressing the mannosidase under TEF1 promoter control showed a slow and clumpy growth phenotype (one of these clones was called G016), further steps in glyco-engineering were done in a strain background where the gene is under Hpd4 transcriptional control.

[0149] One positive clone expressing the ManHDEL under control of the Hpd4 promoter (G018) was chosen, from which the URA3 marker was cured via transient transformation of plasmid pRRQ2 (Richard et al., 2001 J Bacterial. 183:3098-3107), expressing the Cre-recombinase. Several ura3-clones were selected after the procedure and one clone (G036), showing a clear Manα2GlcNAc2 profile on the secretome, was used for further engineering work (FIG. 4). Southern analysis of this clone revealed the presence of one randomly integrated mannosidase expression cassette. This Southern analysis was performed on HindIII digested genomic DNA using a DIG-labeled mannosidase-specific PCR fragment that was generated using primers Man for (5’-GGCGTTCCAGACCTTGTGGAGCCTACCA-3’, SEQ ID NO:22) and Man rev (5’-GGCGTGCAGGTGCAGTGCAGTAAGAGATCC-3’, SEQ ID NO:23).

[0150] In an alternative strategy, two constructs were generated that allowed targeted integration of the Hpd4-driven mannosidase expression cassette into either the LEU2 or AXPl locus of the Yarrowia genome. Construction of these plasmids, JME926 pPTLLE2-2ADE2sec-Hpd4ManHDEL (Y1) and OXY289 pPTAXpl1-LEU2sec-Hpd4ManHDEL (Y1), is described in FIG. 5. Prior to transformation to strain G014, both constructs were digested with NotI and the respective expression cassettes were isolated. Selected ADE2 prototrophic clones had potentially integrated the mannosidase expression cassette into the LEU2 locus, whereas LEU2 prototrophs potentially had integrated the cassette into the AXPl locus. The transformants were checked by Southern analysis to assess proper targeting into the genome. This was performed on BamHI digested (integration in LEU2 locus) or HindIII digested (integration in AXPl locus) genomic DNA using a DIG-labeled mannosidase-specific PCR fragment that was generated using primers Man for (5’-GGCGTTCCAGACCTTGTGGAGCCTACCA-3’, SEQ ID NO:22) and Man rev (5’-GGCGTGCAGGTGCAGTGCAGTAAGAGATCC-3’, SEQ ID NO:23). The selected clones also were checked for the nature of the N-glycans synthesized into the secreted glycoproteins. In most cases, correctly targeted Hpd4-driven α-1,2-mannosidase expression resulted into the synthesis of predominantly Manα2GlcNAc2 oligosaccharides (FIG. 6). For each targeting locus, one mannosidase expressing clone (G046 in case of LEU2 docking; G053 in
case of AXP1 docking) was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2 for strain G046 and pUB4-Cre for strain G053. The resulting cured strains (G055 and G054, respectively) were re-checked via Southern blotting and their Man₅GlcNAc₂ profile confirmed via N-glycan analysis using DSA-FACE.

Example 3
Expression of GlcNAc-Transferase I

[A0151] A Yarrowia codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 100 N-terminal amino acids of the S. cerevisiae Kex2 protein (SwissProt Acc No P27809) followed by the catalytic domain of human GlcNAc-transferase I (SwissProt Acc No P02672) (Fig. 7, SEQ ID NO:3 and SEQ ID NO:4). The yeast Kex2p 100 N-terminal amino acids are localized as a Golgi localization signal for the catalytic GnT I domain. In this way, it is ensured that the GnT I fusion protein is localized later in the secretion pathway than the ER-retained HDEL-tagged α-1,2-mannosidase in order to enable the enzyme converting the protein-linked N-glycans from Man₅GlcNAc₂ to GlcNAcMan₅GlcNAc₂. The codon optimized synthetic gene for the expression of the fusion protein was placed under the transcriptional control of either the TEF1 or the Hp4d promoter, resulting into the plasmids pYLTmAAXGnT1 and pYLT1Hp4mAXGnT1. The construction strategy is shown in FIG. 8. Functional expression of the Kre2-GnT I fusion protein should result in the addition of a β-1,2-linked GlcNAc residue onto the available Man₅GlcNAc₂ glycans resulting in the synthesis of GlcNAcMan₅GlcNAc₂.

[A0152] The plasmids pYLTmAAXGnT1 and pYLT1Hp4mAXGnT1 were NotI digested before transformation to strain G036 (cf. Example 2), known to produce Man₅GlcNAc₂ N-glycans on its secreted proteins. Transformants were selected for uracil prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a clear change in the N-glycan pattern: the Man₅GlcNAc₂ was significantly reduced and a new peak, representing an N-glycan with higher molecular weight (about one glucose unit extra), appeared. Treatment of the isolated N-glycans with Jack Bean β-N-acetylhexosaminidase, an enzyme capable of removing terminal β-linked GlcNAc residues, indicated that the new N-glycan is GlcNAcMan₅GlcNAc₂: the new peak disappeared and was completely converted into Man₅GlcNAc₂ (Fig. 9). Depending on the cultivation method used, about 70% of the total N-glycan pool proved to be GlcNAcMan₅GlcNAc₂ (with approximately 77% of the available Man₅GlcNAc₂ being converted).

[A0153] One transformant expressing the Kre2-GnT I fusion protein under control of the TEF1 promoter was named strain G040 and selected for further use. Genomic analysis of this strain via Southern blot indicated the presence of one expression cassette. Southern analysis was done on BamHI digested genomic DNA using a DIG-labeled GnT I-specific PCR fragment that was generated using primers 5′-GGATCAGCTACAACATGGGCCTTTGTTCTG-3′ (SEQ ID NO:5) and 5′-TGCCTCCTAGCTGTAATCAAGGGGTC-3′ (SEQ ID NO:6). Analysis of the glycosylation profile on the secretome of strain G040 versus strains carrying 1 to 3 copies (confirmed by the same southern blot) of the Hp4d-driven Kex2-GnT I expression cassette, did not show significant difference in GlcNAc-transfer capacity.

Example 4
Expression of Mannosidase II

[A0154] A Yarrowia codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the S. cerevisiae Man2 protein (SwissProt Acc No P38069) followed by the catalytic domain of Drosophila melanogaster mannosidase II (SwissProt Acc No Q24451) (Fig. 10, SEQ ID NO:7 and SEQ ID NO:8). The yeast Man2p 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic Man II domain. In this way, it is ensured that the Man2-Man II fusion protein is localized at the same or even a later position in the secretion pathway than the Kre2-GnT I fusion protein and is therefore able to convert GlcNAcMan₅GlcNAc₂ into GlcNAcMan₅GlcNAc₂. The Yarrowia codon optimized synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the TEF1 promoter, resulting into the plasmids pYLTmAXDmManII and pYLTmMAXDmManII (LEU2cα). The construction strategy is shown in FIG. 11.

[A0155] Plasmid pYLTmAXDmManII (LEU2cα) was NotI digested before transformation to strain G040 (see Example 3), which was known to produce GlcNAcMan₅GlcNAc₂ N-glycans on its secreted proteins. Transformants were selected for leucine prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a change in the N-glycan pattern: a new peak representing an N-glycan with a lower molecular weight of about two glucose units appeared, which could indicate the formation of GlcNAcMan₅GlcNAc₂ and thus partial mannosidase II activity. Also another peak appears, running at almost the same position as Man₅GlcNAc₂ (i.e. a shoulder to the peak), potentially representing GlcNAcMan₅GlcNAc₂. The latter structure could be the result of a partial trimming event, where the mannosidase II activity has only removed one mannose residue instead of two. Treatment of the isolated N-glycans with Jack Bean β-N-acetylhexosaminidase resulted in a leftward shift of the glycan pattern with about one glucose unit and thus a higher electrophoretic mobility due to the loss of a terminal GlcNAc residue (Fig. 12). This further confirms the generation of GlcNAcMan₅GlcNAc₂ and GlcNAcMan₅GlcNAc₂ from GlcNAcMan₅GlcNAc₂ due to the expression of a functional mannosidase II activity. Depending on the cultivation method used, about 15% of the total N-glycan pool proved to be GlcNAcMan₅GlcNAc₂: approximately 35% of the available GlcNAcMan₅GlcNAc₂ lost 1 or 2 mannose residues, with 20% being completely trimmed towards GlcNAcMan₅GlcNAc₂.

Example 5
Expression of Galactosyltransferase I

[A0156] Synthesis of N-glycans with terminal galactose residues not only depends on the presence of a functional and well-localized galactosyltransferase within the secretion pathway, but also on the availability of UDP-Gal, the donor substrate that is used by the enzyme. Although UDP-Glc and UDP-GlcNAc are generally thought to be sufficiently available in the Golgi apparatus of yeast organisms, this is less known for UDP-Gal. To overcome potential UDP-Gal deficiency during glyco-engineering, attempts have been made previously in Pichia pastoris to target a fusion protein of the
**Schizosaccharomyces pombe** UDP-Glc-4-epimerase (encoded by the GAL10 like gene SPBC365.14c-SwissProt Acc No Q9Y7x5) and the catalytic domain of the human β-1,4-galactosyltransferase I (Gaff I) (SwissProt Acc No P15291) into the yeast Golgi apparatus (Bobrowicz et al., Glycobiology 14(9):757-766, 2004). Localization of the Gal10p-Gaff I fusion protein within the secretion pathway, preferably at a position where GlcNAc-transfer and mannosidase II activity has already acted on the N-glycans of proteins destined for secretion, was accomplished by using the first 46 N-terminal amino acids of *S. cerevisiae* Mnn2p as N-terminal targeting signal. 

**0157** Hence, a *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 46 N-terminal amino acids of the *S. cerevisiae* Mnn2 protein, followed by the *S. pombe* Gal10-like protein and the catalytic domain of human GalT I (FIG. 13). The resulting synthetic gene was placed under the transcriptional control of the TEF1 promoter, resulting in the plasmids pYLTmXXSpGal10hGatM and pYLTmXXSpGal10hGatT1 (ADE2ex). The construction strategy is shown in FIG. 14.

**0158** Plasmid pYLTmXXSpGal10hGatT1 (ADE2ex) was NotI digested before transformation to strain G040 (see Example 3), known to produce GlcNAcMan6GlcNAc2, N-glycans on its secreted proteins. Transformants were selected for their adhesion prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a change in the N-glycan pattern: a new peak appears, running at a position between Man5GlcNAc2 and Man6GlcNAc2 (FIG. 15). Treatment of the N-glycans with *Streptococcus pneumoniae* β-1,4-galactosidase indicates that the peak represents GalGlcNAcMan6GlcNAc2 since this in vitro digest results in the disappearance of this new peak and an equally high increase in GlcNAcMan6GlcNAc2.

**0159** Using this set-up and depending on the growth conditions, about 75% of GlcNAcMan6GlcNAc2 was converted into GalGlcNAcMan6GlcNAc2. The total amount of the galactosylated structure accounted for about 25% of the total N-glycan pool. From an in vitro α-1,2-mannosidase digest it is clear, however, that a significant amount of high-mannose N-glycans was not converted to Man6GlcNAc2 (FIG. 15). Depending on the cultivation medium used, the conversion rate of Man6GlcNAc2 towards GalGlcNAcMan6GlcNAc2 also is lower than that observed in the G040 parent strain. This is most probably related to the slower growth rate observed for transformants of this Mnn2-Gal10-GatT1 fusion protein.

**Example 6**

**Knock-Out of YIALG3 and Simultaneous Overexpression of YIALG6**

**0160** To allow the generation of a Man6GlcNAc2 platform, the ALG3 gene of strain G036 (pol1Δ INGA ΔOCH1 Hp4dr-driven α-1,2-mannosidase) needs to be inactivated. This results into the loss of the ER-localized Alg3p α-1,6-mannosyltransferase activity and changes the composition of the lipid-linked N-glycan precursor structure. Transfer of this structure to an N-glycosylation site of a nascent polypeptide chain makes it possible to convert the yeast glycosylation profile into mammalian-like N-glycan structures without the need to express the Mannosidase II. However, since this new lipid-linked structure is not transferred as efficiently to nascent polypeptides, the *Yarrowia* ALG6 gene (encoding an ER-localized Alg6p α-1,3-glucosyl transferase) needs to be overexpressed simultaneously to reduce potential protein underglycosylation as much as possible.

**0161** A vector called pYLAglg3PUT-ALG6 (FIG. 16) was constructed previously to allow simultaneous knock-out of YIALG3 and Hp4dr-driven overexpression of YIALG6. See U.S. Patent Publication No. 20090069232-A1. A NotI/PvuII fragment of this vector, containing this knock-out/knock-in cassette, was transformed into *Yarrowia lipolytica* G036 and transformants were selected based on their uracil prototrophy. Clones that had correctly integrated the construct were directly screened via N-glycan analysis on the secretome. Out of 80 screened clones, 2 clones showed an N-glycosylation profile that could fit with the inactivation of YIALG3 in a strain expressing an ER-located α-1,2-mannosidase. Apart from a fraction Man6GlcNAc2 glycan, there was still some Man5GlcNAc2 and Man6GlcNAc2 as well as a significant amount of glucosylated N-glycans (GlcMan5,GlcNAc2 and GlcMan6,GlcNAc2). The latter are the result of an inefficient trimming by glucosidase II (Griffin and Robbins, J. Biol. Chem. 255, 2255-2258, 1980). The nature of the structures of Man6GlcNAc2 and Man5GlcNAc2 was confirmed by in vitro treatment of the N-glycans with α-1,2-mannosidase (FIG. 17). Depending on the growth conditions used, the level of Man6GlcNAc2 could increase to up to 60% of the total N-glycan pool, with the glucosylated peaks being insensitive towards α-1,2-mannosidase and only slightly sensitive towards Jack Bean α-mannosidase treatment (aspecific c-mannosidase that cleaves on α-1,2-, α-1,3- and α-1,6-linked mannose residues). In contrast, the latter enzyme converts the generated Man5GlcNAc2 into Man6GlcNAc2 (FIG. 17).

**0162** One of the two positive transformants was called G039 and used for further glyco-engineering work. The strain was transformed transiently with pRRQ2 expressing the Cre-recombinase to allow the curing of the URA3 marker that was introduced upon transformation of G036 with vector pYLAglg3PUT-ALG6. Analysis shows that the glycosylation profile remains the same after curing. One cured strain was selected for further use and designated G045.

**Example 7**

Expression of GlcNAc-Transferase I in a Man6GlcNAc2 Producing Strain

**0163** Similar to what was done in example 3, the introduction of a GnT I activity was accomplished via the expression of the Kec2-GnT I fusion protein. Random integration of such an expression construct for GnT I was accomplished in three ways: 1) the non cured strain G039 (see Example 6) was transformed with the NotI digested vector pYLTmAXhGnT1 (Hygr ex) and GnT I expressing clones were initially selected based on their ability to survive 300 µg/ml of hygromycin added to the selection plates, 2) the cured strain G045 (see Example 6) was transformed with the NotI digested vector pYLTmAXhGnT1 (see also Example 3) and GnT I expressing clones were initially selected based on their uracil prototrophy or 3) the cured strain G045 (see Example 6) was transformed with the NotI digested vector pYLP1H4 mAXhGnT1 and GnT I expressing clones were initially selected based on their uracil prototrophy. The construction strategy for pYLTmAXhGnT1 (Hygr ex) is shown in FIG. 18. When using plasmids pYLTmAXhGnT1 (Hygr ex) and pYLTmAXhGnT1, the expression of GnT I was under the transcriptional
control of the TEF1 promoter; when using plasmid pYLHp4 mXhGnTI, the GnT1 expression was under the control of the Hp4d promoter.

[0164] Transformation of G039 with pYLtmxAxhGnTI (Hygr cv) resulted in three clones that only emerged on the culture plates after a longer incubation period than what was expected. However, analysis of the N-glycosylation profile of the secretome of these clones showed a clear change in the N-glycan pattern: the Man3GlcNAc2 present in the non-transformed G039 strain was significantly reduced or almost completely absent while a new peak, representing an N-glycan with higher molecular weight (above one glucose unit extra), appeared. Treatment of the isolated N-glycans with Jack Bean β-N-acetylhexosaminidase, an enzyme capable of removing terminal β-linked GlcNAc residues, indicated that the new N-glycan indeed is GlcNAcMan3GlcNAc2. The new peak disappeared and was completely converted into Man3GlcNAc2 (Fig. 19). One of the evaluated transformants was used for further glyco-engineering work and named G047. Similar results were also obtained when the cured strain G045 was transformed with pYLtmxAxhGnTI (G048) or with pYLHp4 mXhGnTI (G056). Strain G056 was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2. The resulting strain was called G058.

[0165] Depending on the cultivation method used, about 70% of the total N-glycan pool of strain G047 proved to be GlcNAcMan3GlcNAc2 with some Glc3Man3GlcNAc2 and almost no Man4GlcNAc2 was present (conversion rate >=90%) (Fig. 19). Regardless of the high conversion rate, only one copy of the GnT1 expression cassette could be identified in this strain via Southern blot. Southern analysis was done on BamH1-digested genomic DNA using a DIG-labeled GnT1 specific PCR fragment that was generated using primers 5'-GGATGATCACACTAGGTGCCCTTCTGTCGTTGGGAATGTCTG-3', SEQ ID NO:13) reverse primer Ver2ade2 (5'-CTACGTGGCGAAATGCGATACACCGCCCTGTTCTAGTG-3', SEQ ID NO:14) and further confirmed via Southern blotting. This was done on BamH1/SpeI-digested genomic DNA using a DIG-labeled GnT1 specific PCR fragment that was generated using primers 5'-GGATGATCACACAAATGGCGCCTGTTCTAGTG-3', SEQ ID NO:15) and 5'-TGCTCTAGACTTGCCCAAGGCGGTC-3' (SEQ ID NO:16). Synthesis of GlcNAcMan3GlcNAc2 onto the secretome was confirmed by N-glycan analysis and in vitro Jack Bean β-N-acetylhexosaminidase treatment (Fig. 21). One GnT1 expressing transformant (called G057) was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2. The resulting strain was called G059.

[0167] A Yarrowia codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the S. cerevisiae Man2 protein (SwissProt Acc No P38069) followed by the catalytic domain of rat GlcNAc-transferase II (GnT II) (SwissProt Acc No Q09325) (Fig. 22, SEQ ID NO:17 and SEQ ID NO:18). The yeast Man2 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic GnT II domain. In this way, it was ensured that the Man2-GnT II fusion protein was localized at the same or even a later position in the secretion pathway than the Kre2-GnT1 (and Man2-Man) fusion protein and was therefore able to convert GlcNAcMan3GlcNAc2 into GlcNAcMan3GlcNAc2. The synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the TEF1 promoter, resulting in the plasmids pYLtmXMnGnTII and pYLtmXMnGnTII (ADE2ex). The construction strategy is shown in Fig. 23.

[0168] A strain expressing the GnT II activity was generated in two different ways: 1) strain G045 (see Example 6) was transformed simultaneously with NotI-digested pYLtmxAxhGnTI and NotI-digested pYLtmXMnGnTII (ADE2 ex) and transformants were selected based on their uracil and adenine protoporphyrin or 2) strain G047 (Example 7) was transformed with NotI-digested pYLtmAXNnGnTII (ADE2 ex) and transformants were selected based on their adenine protoporphyrin. Integration of the expression cassette was checked using forward primer TepPromFW 5'-GTCCTCCGATACTCGTTGACCAAGGCGGTC-3' (SEQ ID NO:19) and reverse primer Lip2TermRV 5'-AGGTAGATTGAAAGAGGATG-3' (SEQ ID NO:20). N-glycan analysis on the secretome in combination with in vitro treatment of the isolated sugars with Jack Bean β-N-acetylhexosaminidase indicated that several transformants were capable of producing GlcNAcMan3GlcNAc2 and thus of expressing a functional GnT II activity (Fig. 24). In one selected condition, about 40% of the total N-glycan pool consisted of GlcNAcMan3GlcNAc2. The conversion rate of the substrate GlcNAcMan3GlcNAc2 to GlcNAcMan3GlcNAc2 was 90%. The final selected strains were called G050 (double transformation of G045) and G051 (GnT II expression in G047).

Example 9

Expression of Glucosidase II Alpha and Beta Subunits (Gh26t and Gh26b)

[0169] Based on the experiments described in Examples 6 to 8, the strategy involving the knock-out of YIALG3 and simultaneous overexpression of YIALG6 results into the generation of N-glycans carrying one or two terminal glucose residues (Glc,Man5GlcNAc2). The presence of these glucose residues hampers the conversion towards Man3GlcNAc2 by the ER-localized HDEL-tagged α-1,2-mannosidase. In order for the glucose residues to be removed, the glucosidase II activity within the ER needs to be increased. In a background without α-1,2-mannosidase expression, overexpression of the Aspergillus niger glucosidase 11 alpha and beta subunit resulted in the highest conversion of Glc,Man5GlcNAc2 into Man5GlcNAc2 (U.S. Patent Publication No. 2009009232-A1). Constructs for the overexpression of
the *A. niger* gls2 subunits were produced as follows: 1) a *Yarrowia* codon-optimized cDNA was generated for the expression of the mature (lacking the signal peptide) *A. niger* gls2α and gls2β subunits; 2) the cDNA’s were cloned in-frame to the *E. coli* LIP2pre-sequence; 3) the resulting LIP2pre-gls2α and LIP2pre-gls2β sequences were cloned under the transcriptional control of the constitutive TEF1 promoter. The resulting plasmids were called pYLTU5KL2preAnGlcIIα and pYLm2EXTTEFprep2AnGlcIIβ (FIG. 25).

[0170] Based on these plasmids, new constructs were generated for the simultaneous overexpression of the *A. niger* gls2α and gls2β subunits under either TEF1 promoter control (vector JME923 pPTura3-1LEU2ex-TefL2preAnGlcIIα+b[alt] for targeted integration—FIG. 26) or Hp4d promoter control (vector JME923 pPTura3-1LEU2ex-Hp4dL2preAnGlcIIα+b[alt] for targeted integration and vector Zeta-LEU2ex-Hp4dL2preAnGlcIIα+b[alt] for random integration—FIG. 27).

[0171] Strain G057 (see example 7) was transformed with NotI digested plasmids JME923 pPTura3-1LEU2ex-Hp4dL2preAnGlcIIα+b[alt] and Zeta-LEU2ex-Hp4dL2preAnGlcIIα+b[alt] and transformants were selected based on their leucine prototrophy. Several clones were analyzed genotypically via PCR and Southern analysis to evaluate the integration of the gls2α and gls2β expression cassette. PCR-analysis and DIG probe generation for the gls2α subunit was done using primers AnGls2α-FW (5'-GCTGAGACTCTCTTCTTAACC-3') (SEQ ID NO:24) and AnGls2α-RV (5'-GGTCTCCCTGAGACAGGAGG-3') (SEQ ID NO:25); for the gls2β subunit we made use of primers AnGls2β-FW (5'-CCAAAGTCTCAAGGAGCAACC-3') (SEQ ID NO:26) and AnGls2β-RV (5'-CCCTTGGAGACCTTAGAGG-3') (SEQ ID NO:27). Southern analysis to check for targeted integration of the dual Hp4dGls2αβ expression cassette was done on Eco47III-digested DNA when using the gls2α probe, and on SplI/SfiI-digested DNA when using the gls2β probe. The majority of the selected clones showed correct integration of the dual expression cassette into the URA3 locus. Southern analysis for random integration of the dual Hp4dGls2αβ expression cassettes was checked on PvuII-digested DNA with both probes. In all cases, only one copy of the dual expression cassette was integrated.

[0172] Next, N-glycan analysis was performed on several clones confirmed to have the dual Hp4dGls2αβ expression cassette (correctly) integrated. N-glycosylation was examined on total secreted protein after three days of falcon cultivation. Several clones showed a significant reduction of the glycosylated sugars and an increase of Mann,GlcNAc2, and GlcNAcManα,GlcNAc2. The profiles of a clone that has integrated the dual expression cassette randomly (strain G060) on the one hand and in a targeted way (strain G061) on the other, are shown in FIG. 28. The two smaller peaks represent Manα,GlcNAc2 and Manα,GlcNAc, since they shift to Manα,GlcNAc upon treatment with α-1,2-mannosidase and Jack Bean mannosidase. The latter treatment also results in a partial conversion of the remaining GlcNAc2Manα,GlcNAc into GlcNAcManα,GlcNAc and of GlcNAcManα,GlcNAc into GlcNAcManα,GlcNAc. Presence of Manα,GlcNAc and Manα,GlcNAc, however indicates incomplete conversion towards Manα,GlcNAc by the heterologously co-expressed HDEL-tagged α-1,2-mannosidase. Similarly, the presence of Manα,GlcNAc indicates incomplete transfer of a GlcNAc-residue by recombinant human GntI to obtain GlcNAcManα,GlcNAc. However, based on results described above (e.g. G047 cultivation in Example 7, FIG. 19), it is clear that differences in cultivation conditions can increase the conversion rates significantly and thus improve the end result.

Example 10

Expression of GlcNAc-Transferase II in the GlcNAcManα,GlcNAc-Producing Strain G061

[0173] As described in Example 8, a *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the *S. cerevisiae* Mnn2 protein (SwissProt Acc No P38069) followed by the catalytic domain of rat GlcNAc-transferase II (Gnt II) (SwissProt Acc No Q09326) (FIG. 22, SEQ ID NO:17 and SEQ ID NO:18, respectively). The yeast Mnn2 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic Gnt II domain. In this way, it was ensured that the Mnn2-Gnt II fusion protein was localized to the same or even a later position in the secretion pathway than the Kex2-Gnt II fusion protein and was therefore able to convert GlcNAcManα,GlcNAc into GlcNAcManα,GlcNAc. The synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the Hp4d promoter resulting in plasmid pYLIHp4dMxARTGnTII, which was used for random integration of the Hp4d-driven Gnt II expression cassette into the *Yarrowia* genome. In an alternative strategy, construct OXYP289 pPTAprl-AD2ex-Hp4dGnTII was generated to allow targeted integration of the Hp4d-driven Gnt II expression cassette into the *Apx1* locus of the *Yarrowia* genome.

[0174] Prior to transformation of strain G061 (see Example 9), the plasmids were NotI digested and the targeting/expression cassette was isolated. Transformants were selected based on their adenine prototrophy. Correct integration of the expression cassette into the ADE2 locus was confirmed by Southern blot analysis after digesting the genomic DNA with Xmnl. A DIG-labeled probe with specificity for the Gnt II coding sequence was generated using forward primer rGnt-2FW (5'-GCCAGAGCTGCGCTAAAGC-3') (SEQ ID NO: 28) and reverse primer rGnt-RV (5'-CCCTTGACGTCACCTGTGC-3') (SEQ ID NO: 29). This strategy produces a band of 3172 bp when the gene is successfully integrated into the Apx1 locus.

[0175] In an alternative strategy, correct integration into the Apx1 locus can be examined via a PCR reaction on genomic DNA using the forward primer AXPVer1b (5'-GCGTGAACGGCAGTCGATGCAGGCAAATCC-3') (SEQ ID NO: 30) and the reversed primer AXPVer2b (5'-CAAGAGCTCAGGCTCGAGATCTCCA-TC-3') (SEQ ID NO: 31). In case of correct targeting into the Apx1 locus, a PCR fragment of 6489 bp is expected.

[0176] N-glycan analysis on the secretome, in combination with in vitro treatment of the isolated sugars with Jack Bean β-N-acetylhexosaminidase or *T. reesei* α-1,2-mannosidase, indicated that several transformants were capable of producing GlcNAcManα,GlcNAc and thus of expressing a functional Gnt II activity (FIG. 29). The analyses indicated that about 25 to 30% of the total N-glycan pool consisted of GlcNAcManα,GlcNAc, with a GlcNAcManα,GlcNAc to GlcNAcManα,GlcNAc conversion rate of about 90%. The final selected strains were called G070 (integration of
Example 11

Construction of a Tandem Plasmid for Simultaneous Hp4d-Driven Expression of the Anti-HER2 Heavy Chain (HC) and Light Chain (LC) into Yarrowia lipolytica

[0177] The amino acid sequences for the anti-HER2 antibody heavy and light chains were obtained from Carter et al., Proc Natl Acad Sci USA, 89(10): 4285-4289 (1992); and Ward et al., Appl Environ Microbiol, 70(5): 2577-2576 (2004). The relevant amino acid sequences were reverse translated, codon-optimized for Yarrowia lipolytica, and synthesized by GenArt, Regensburg Germany. Regions of very high (>80%) or very low (<30%) GC content were avoided where possible. During the optimization processes, the following cis-acting sequence motifs were avoided: internal TATA-boxes, chi-sites and ribosomal entry sites, AT-rich or GC-rich sequence stretches, repeat sequences and RNA secondary structures as well as (cyclic) splice donor and acceptor sites. In order to allow secretion of the epitope proteins, the coding sequence of the lip2 protein ‘prepro’ signal (followed by that of a peptide linker ‘GGG’) was added to the 5' region of the coding sequences, ‘GGG’ was added to enhance the changes for correct Kex2 processing. FIG. 3A contains the nucleotide sequence of the synthetic preprolip2-LC (=750 bp) (SEQ ID NO: 32). FIG. 3B contains the amino acid sequence of the preprolip2-LC (=250 Aa; MW=27.011 Da; pl=8.46) (SEQ ID NO: 33). FIG. 3A contains the nucleotide sequence of the synthetic preprolip2-HC (=1458 bp) (SEQ ID NO: 34). FIG. 3B contains the amino acid sequence of the preprolip2-HC (=486 Aa; MW=52.853 Da; pl=8.65) (SEQ ID NO: 35). The coding sequences for preprolip2-HC and -LC were introduced into the same vector, called pYLHp4L2preproHerHC&LC (GUT2exy)-ori2.

Example 12

Expression of the Anti-HER2 Antibody HC and LC into Yarrowia lipolytica Strains with a Varying Degree of Glyco-Engineering

[0178] Plasmid pYLHp4L2preproHerHC&LC (GUT2exy)-ori2 was digested with NotI and the HC/LC-tandem expression cassette was isolated before transforming Yarrowia lipolytica strains G045, G057, G061 and G071 (see Table 2). Transformants containing the randomly integrated HC/LC-expression cassette were selected based on their ability to grow on glycerol as the sole carbon source. Expression analysis of the HC and LC was done via western blotting after a 4 day shake flask cultivation of the selected transformants in rich medium containing glycerol as the only carbon source (SuperT/glycerol medium: 0.5% yeast extract; 2% malt extract; 1% tryptone; 1.5% glycerol; 200 mM phosphate buffer pH 6.8). LC-detection was performed using a mouse monoclonal to Kappa Free Light Chains (4C11) (Abeam) while HC-detection was done using mouse monoclonal anti-human IgG (g-chain specific) (Sigma).

[0179] The N-glycans of the secretome of the anti-HER2 antibody producing strains showed a similar profile as the corresponding glyco-engineered strains that were not expressing any HC and LC (FIG. 32). The percentages of N-glycans in strains with the G045, G057, G061, and G071 background were determined after a 6-day shake flask cultivation in SuperT/glycerol medium. In a G045 background, 54.6% of the N-glycans were Man$_2$GlcnAc$_2$. In the G057 background, 47.5% of the N-glycans were GlcnAc$_2$Man$_2$GlcnAc$_2$. In a G061 background, 58.9% of the N-glycans were GlcnAc$_2$Man$_2$GlcnAc$_2$. In a G071 background, 37.6% of the N-glycans were GlcnAc$_2$Man$_2$GlcnAc$_2$.

Example 13

Fermentation of Yarrowia Strain G096, a GlcnAc$_2$Man$_2$GlcnAc$_2$, Synthesizing Strain Expressing the Anti-HER2 Antibody HC and LC

[0180] Several pYLHp4L2preproHerHC&LC (GUT2exy)-ori2 transformants of Yarrowia lipolytica G071, a strain capable of synthesizing GlcnAc$_2$Man$_2$GlcnAc$_2$, were analyzed for HC and LC expression levels. One of these clones, G096, was chosen for further analysis.

[0181] Fermentation was done in a 14-litre stirred tank bioreactor (MAVAG AG) equipped with a process control and management system (Lucillus PIMS). The relative partial oxygen pressure in the medium, the CO$_2$ and O$_2$ concentrations in the exhaust gas, pH value, temperature, reactor over-pressure, reactor weight, feed weight and base weight were all monitored on-line. Foam generation was counteracted by adding the antifoaming agent polypropylene glycol (PPG). Adjustments in pH were done by either the addition of a 25% ammonia solution or a 8.5% phosphoric acid solution.

[0182] A seed culture of G096 was grown at 28°C. In a shake flask containing rich medium. The seed culture was inoculated into the fermentor containing mineral medium to start a batch phase at 28°C with unrestricted growth, using glycerol as only carbon source. This phase was used to rapidly reach a high biomass concentration. From that point onward, the process was shifted to an exponential glycerol fed batch (with glycerol as sole carbon and energy source; pH 6), with a constant growth rate of 0.02. As an example, the results for a fed batch fermentation at 28°C are described below.

[0183] The fed-batch phase lasted for 148 hours. At different time-points of the fermentation, samples were taken to follow up the following parameters: 1) expression of the LC and HC protein backbones via western blot; 2) expression of functional anti-HER2 antibody via an ELISA; and 3) evolution of the N-glycosylation profile of the secretome. The full-length HC expression level reached a maximum around timepoints 7 (39 hrs) and remains approximately equal from then onwards. The LC expression reached a maximum between time-points 7 (39 hrs) and 10 (73 hrs), but decreased somewhat in the later time-points. Some LC-dimers were produced between time-points 5 (25 hrs) and 9 (62 hrs), but disappeared again from that point onwards.

[0184] A functional ELISA was developed to measure the production of anti-HER2 antibody that has at least one functional antigen binding domain. Plates were coated with a recombinant variant of the natural HER2 antigen, the recombinant human ErbB2/Fc chimera (R&D systems). Then a dilution of the medium, harvested at different time-points, was added to the coated plates. Assessment of the amount of antigen binding protein was done using a HRP-conjugated anti-human kappa LC antibody (Sigma). The evolution of the amount of ErbB2/Fc chimera binding protein (a measure of the amount of secreted functional anti-HER2 antibody) within the fed-batch fermentation is shown in FIG. 33. The
data show a gradual increase in the levels of anti-HER2 antibody, with a maximum of 10 to 12 mg/L at the end of the production phase.

N-glycan analysis was done on samples taken at several time-points during the fed-batch fermentation. The results are shown in FIGS. 34A and 34B. At the beginning of the fed-batch phase, there was a significant amount of glucose-containing N-glycans present. From time-point 6 onward (34 hrs after start of exponential feeding), the level of glucosylated N-glycans decreased significantly with hardly any left at the time of harvest (time-point 18, 148 hrs). This indicated that proteins originally carrying glucose-containing N-glycans, were diluted out by the end of the fermentation. At that point about 86% of the N-glycans isolated from the secretome had the structure GlcNAc2Man3GlcNAc2.

Other Embodiments

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

**SEQUENCE LISTING**

```
19
tcgtaatac gctcgtagc

29
tactctgata ctgtagata ctgtagac

```

```
5
Met Ala Leu Phe Leu Ser Lys Arg Leu Leu Arg Phe Thr Val Ile Ala
Gly Ala Val Ile Val Leu Leu Thr Leu Asn Ser Asn Ser Arg Thr
20
Gln Gin Tyr Ile Pro Ser Ser Ile Ser Ala Ala Phe Asp Phe Thr Ser
35
Gly Ser Ile Ser Pro Glu Gin Gin Val Ile Ser Glu Glu Asn Asp Ala
50
Lys Lys Leu Glu Gin Ser Ala Leu Asn Ser Glu Ala Ser Glu Asp Ser
65
Glu Ala Met Asp Glu Glu Ser Lys Ala Leu Lys Ala Ala Glu Lys
85
Ala Asp Ala Pro Pro Ala Val Ile Pro Ile Leu Val Ile Ala Cys Asp
100```
<210> SEQ ID NO 4
<211> LENGTH: 1332
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:3
<400> SEQUENCE: 4
atggccctgt ttctgtctaa gcgcctgctg caggtccaccg tgatcgcggc ttgctgtgcc
60
gtctgctgc ttgacctgaa ctctaccctc cgaaccaccgc agtacatccc ctctttctac 120
-continued

tctgccccct tgcacttca cacctactct tatacttcct ccaccacctc cagcagcgcct gatctctctag 180
gagagacgac ccagacacagc gcctctgcct tctgcactctt gcagcagcct 240
gagcgcacg cagcagcactc aggcacgtg aagcagctgt cgcagacgacct 300
ccagcgcactc tctccctgcct aaggcacgtg ttcctctgtc cagacgacct 360
gacacgcgtc tgcacttca gacccctgcct ccagcagcctt gcagcagcctc 420
tctgcactctc aggcaggccct ccagcagcctt gcagcagcctg gcagcagcctc 480
cagcagcctt tatactttc ctcaccct ccacgcagcct ccagcagcct 540
tacaagcactc ccaccactc ccgcagcgcct ccagcagcgcct cctctctctgt gttctctgcct 600
cagcgcagcct cactcctctc gcacgcagcct ccagcagcgcct cagcagcgcct 660
cgagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct cagcagcgcct 720
gacacgcgcct gcagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct 780
ttgttgtgtgag gcagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct 840
tgagcgcgcct gcagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct 900
tgagcgcgcct gcagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct 960
cagcgcgcct gcagcagcgcct cactcctctc gcacgcagcct ccagcagcgcct 1020
<210> SEQ ID NO 5
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 5

gagagacgac ccagacacagc gcctctgcct tctgcactctt gcagcagcct 28

tgtctcagc tagttcagc aggggtc 27

tgtctcagc tagttcagc aggggtc 27

tgtctcagc tagttcagc aggggtc 27

<210> SEQ ID NO 7
<211> LENGTH: 1069
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: fusion of N-terminal portion of the S. cerevisiae Mn2 protein and catalytic domain of Drosophila melanogaster mannosidase II
<400> SEQUENCE: 7
Ala Gly Gln Ala Glu Phe Pro Thr Leu Ser Gly Asp Phe Phe Thr Tyr
405 410 415
Ala Asp Arg Ser Asp Asn Tyr Trp Ser Gly Tyr Thr Ser Arg Pro
420 425 430
Tyr His Lys Arg Met Asp Arg Val Leu Met His Tyr Val Arg Ala Ala
435 440 445
Glu Met Leu Ser Ala Trp His Ser Trp Asp Gly Met Ala Arg Ile Glu
450 455 460
Glu Arg Leu Glu Gln Ala Arg Arg Glu Leu Ser Leu Phe Glu His His
465 470 475 480
Asp Gly Ile Thr Gly Thr Ala Lys Thr His Val Val Asp Tyr Glu
485 490 495 500 505 510
Gln Arg Met Gln Glu Ala Leu Lys Ala Cys Glu Met Val Met Gln Gln
515 520 525
Ser Val Tyr Arg Leu Leu Thr Lys Pro Ser Ile Tyr Ser Pro Asp Phe
530 535 540
Ser Phe Ser Tyr Phe Thr Leu Asp Ser Arg Ser Trp Pro Gly Ser Gly
545 550 555 560
Val Glu Asp Ser Arg Thr Thr Ile Leu Gly Glu Asp Ile Leu Pro
565 570 575
Ser Lys His Val Val Met His Asn Thr Leu Pro His Trp Arg Glu Gln
580 585 590
Leu Val Asp Phe Tyr Val Ser Ser Pro Phe Val Ser Val Thr Asp Leu
595 600 605
Ala Asn Asp Pro Val Glu Ala Gln Val Ser Pro Val Trp Ser Thr His
610 615 620
His Asp Thr Leu Thr Lys Thr Ile His Pro Glu Gly Ser Thr Thr Lys
625 630 635 640
Tyr Arg Ile Ile Phe Lys Ala Arg Val Pro Pro Met Gly Leu Ala Thr
645 650 655
Tyr Val Leu Thr Ile Ser Asp Ser Lys Pro Glu His Thr Ser Tyr Ala
660 665 670
Ser Asn Leu Leu Leu Arg Asn Pro Thr Ser Leu Pro Leu Gly Gln
675 680 685
Tyr Pro Glu Asp Val Lys Phe Gly Asp Pro Arg Glu Ile Ser Leu Arg
690 695 700
Val Gly Asn Gly Pro Thr Leu Ala Phe Ser Glu Gln Gly Leu Lys
705 710 715 720
Ser Ile Gln Leu Thr Gin Asp Ser Pro His Val Pro Val His Phe Lys
725 730 735
Phe Leu Lys Tyr Gly Val Arg Ser His Gly Asp Arg Ser Gly Ala Tyr
740 745 750
Leu Phe Leu Pro Asn Gly Pro Ala Ser Pro Val Glu Leu Gly Gln Pro
755 760 765
Val Val Leu Val Thr Lys Gly Leu Glu Ser Ser Val Ser Val Gly
770 775 780
Leu Pro Ser Val Val His Gin Thr Ile Met Arg Gly Gly Ala Pro Glu
785 790 795 800
Ile Arg Asn Leu Val Asp Ile Gly Ser Leu Asp Asn Thr Glu Ile Val
805 810 815
Met Arg Leu Glu Thr His Ile Asp Ser Gly Asp Ile Phe Tyr Thr Asp 820 825 830
Leu Asn Gly Leu Gln Phe Ile Lys Arg Arg Leu Asp Lys Leu Pro 835 840 845
Leu Gln Ala Asn Tyr Tyr Pro Ile Pro Ser Gly Met Phe Ile Glu Asp 850 855 860
Ala Asn Thr Arg Leu Thr Leu Thr Gly Gin Pro Leu Gly Gly Ser 865 870 875 880
Ser Leu Ala Ser Gly Glu Leu Glu Ile Met Gin Asp Arg Arg Leu Ala 885 890 895
Ser Asp Asp Glu Arg Gly Leu Gly Gin Gly Val Leu Asp Asn Lys Pro 900 905 910
Val Leu His Ile Tyr Arg Leu Val Leu Glu Lys Val Asn Asn Cys Val 915 920 925
Arg Pro Ser Lys Leu His Pro Ala Gly Tyr Leu Thr Ser Ala Ala His 930 935 940
Lys Ala Ser Glu Ser Leu Leu Asp Pro Leu Asp Lys Phe Ile Phe Ala 945 950 955 960
Glu Asn Glu Trp Ile Gly Ala Glu Gin Gly Phe Gly Gly Asp His Pro 965 970 975
Ser Ala Arg Glu Asp Leu Asp Val Ser Val Met Arg Arg Leu Thr Lys 980 985 990
Ser Ser Ala Lys Thr Gin Arg Val Gly Tyr Val Leu His Arg Thr Asn 995 1000 1005
Leu Met Gin Cys Gly Thr Pro Glu Glu His Thr Gin Lys Leu Asp Val 1010 1015 1020
Cys His Leu Leu Pro Asn Val Ala Arg Cys Glu Arg Thr Leu Thr 1025 1030 1035 1040
Phe Leu Gin Asn Leu Glu His Leu Asp Gin Met Val Ala Pro Glu Val 1045 1050 1055
Cys Pro Met Glu Thr Ala Tyr Val Ser Ser His Ser 1060 1065

<210> SEQ ID NO 8
<211> LENGTH: 3213
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:?

<400> SEQUENCE: 9
atgcgtgcg ccaagcggta ctctaaagctg ttcaagctga cttctcatgct gcgtgccttg 60
tggtggcttg tcgtgatat ccaacagta cttggagagc acacacgcc ggagtacccc 120
atccgaccc cctgaaagct gcacgaagcgc cccgacccg gccagtgcctgg 180
cagagctg ccaagcggta ctctaaagctg ttcaagctga cttctcatgct gcgtgccttg 240
gcatcgcag gcgcagcctg gcacagcggcg gccacatcgc ctgctgatcg 300
cagccaccc acacacgcag gcgtgttcgg gccactacct ctcaaaagc cccgctgtgg 360
atccagacct tcagagcaac ctacagcagc gacacagtcn ctacgcctctg 420
cacacagcgc acacacaccc ttgagtaag ttctactcgg gcagatcctc ttaccttcgc 480
cagttccct acacacgtcgg cggagacag agcgtgcag cttgatctctat cgtgaagagc 540
ggctacgctg agctgtgtgag cggcaggtcgg gctgatcgcgc acagggcaca cttctactgg 600
cggaaagtgc tgctgcagct gacocagggc cgaacctggc tgaagctagt ctgaagctgtg 660
cocccaccgc ccccctgcgg catgaccccc ttggccact ccccacact gecctacac 720
tcgcagagt ctggctccaa gaaactgctg atcccagcga aacactactc tggtaaggaag 780
gagctggcacc agcagcgaact tgtgagctgg ctggcggca aagactcgga caacaagggc 840
gacacgcggc ctgctcaacca catgagcggc ttctactctc acgacatcgc ccaacactctg 900
ggccccagcc caagaaggtgct ttcgtagctgc gaacctcagct gaccccgtcatc 960
tcggcctcct ggaaggtgcccc cccctgaaacc actcttgaccc agaagtgggcc cggctgatct 1020
agactgtagcttg tgaaccagact gaagagaaag gcgcagactgt acgccagcaca cgctgctgctg 1080
atcccccctg cagccacctgg cgacaccaag cgaacaccccg actggccagc cggagcgagt 1140
aatcagagcc gacgctgca gcaacccaaact ctcgacggcc acttcacaggt gcaggtgcatc 1200
tcggcactctg tggcagggct cttcagcagcc gcacacccgc cgcagcgcgc egccagccct 1260
ccgtgtcctgc cggattctgg cgaactcttct actacaacgc aagcagcttgaa caattaacttg 1320
tctgtactact cccacacttc gccctccacat accataacgc accagctacgct gactgactc 1380
ggtcgagcgc cgaactgctgg ttcctcttgag ccctctctgg aggactgctgg cgaactcagag 1440
gagcagtgtgg cgcagccagg cagagactctg ttcctctctgcc gcacatccagc egcctcaacc 1500
ggcaagccag gaccagcaagttgtggtgacc taagactcagc gaaatcaggg gcgccctgaag 1560
gcctgacac aagcagccagc cgactagctgc ttcactcaagct ccactacatc 1620
tccctccagact gctctcttcctcttcagcc ctctgatcagc ctggtactgctc 1680
gtgggaggact ctcagacacc caaactctgcg ggcagagacc cttggtgccctc taagatctgtg 1740
gttagctgaca cacccctgcgcc cccctgacgcg ggcagactgg cgctacctctca ggtggaactct 1800
ccctctgtgt ctcggtgacc cctggcccaac acccccgttg aggccagagc gcctccgctg 1860
tgcctgctgg gaccacagca cctggacacc acctocacac ccagaggtctct caaccctaaag 1920
tcgggactc accatggcct gccagcttccgcc ccctagggcgcc tgactagcct atgtctgac 1980
atccgtgct tcaacgccagc gcacactcttc actgcctctta actctcggtc gcggactaccc 2040
cctgggctgt cgtagatcacc gacttcggcg ggagacgttg aagttgggca ccccgagag 2100
atctgtgtgc ggaaagggcc cggcccccacc gctgtctcttg cctgacaggg cctgagctag 2160	tctactcagg tcgggcaggg ccctctccacag tcgctgccgtg ctcctcaagctc tctgaagctc 2220
ggtgctcgctg ctcagggcag cgaactgtgct cgtactagtct ttcgctccaa gccaagccgc 2280
tccctctggag aagtgggacca gcccggctgtg cttggctgcac ggcgctctgt ggtgactcctc 2340
gtgctctgag gcctgctcttc gttgtagcc acgacctacta tcggaaggcc agcccgcctag 2400
atcggaaacc tggtgctacagg cgaactctctg gcacatcctgc gacacactgc aqatctgtatc ggcgcgtgag 2460
acccacactc agtcgtggaag ccccttctcag acgacagctga acgcgctgca gtttaataag 2520
cggcagacac ccgacaagct gcgctccgcg cacaactact cccctccatcc ctctgaccatg 2580
ttcgctggca gacgatcactg gcagacggga cacgctgcttc gacgggtttct gggagatgt 2640
ttctctgtggct ctcggagacg ggaagctgtgg gacagcggac gacgctgtcc ctagcgcgctg 2700
cggagctgtg ggccgggctgt ggtggcaacg aacgctgtgc cggactccttc ggcagccttg 2760
tccgagaggg gttgggctcgt gcgagctcctgc ctcagctgctg cttgagctgcc accccgtcttg ctacgtcacc 2820
tctggccccc acaagggcgc ttcagctctg cttgcccctgc tggcaagcttg cattctgccc 2880
gagaacgagt ggtacggcgc ccaggcgcag ttccagggcg accaccctct gccgagag 2940
gacctggagc tgtttgtcct gcagcagctg accaagatcc tggccaagac cccgcagatg 3000
gagtcggtagc gacatgcggc gcagcagctg accaagatcc tggccaagac cccgcagatg 3060
aadcttccagc acgtgcagagc atggagtggc cccgaggttg tccagatgag 3120
acgccgct gcctgtgggct ccaatctctct tag 3180
acgccgct gcctgtgggct ccaatctctct tag 3213

<210> SEQ ID NO 9
<211> LENGTH: 763
<212> TYPE: PEPTIDE
<213> ORGANISM: Artificial Sequence
<222> FEATURE: OTHER INFORMATION: fusion of N-terminal portion of the S. cerevisiae Mnn2 protein, the S. pombe Gal10-like protein, and catalytic domain of human GALT I
<400> SEQUENCE: 9

Met Leu Leu Thr Lys Arg Phe Ser Lys Leu Phe Lys Leu Thr Phe Ile
Val Leu Ile Leu Cys Gly Leu Phe Val Ile Thr Asn Lys Tyr Met Asp
Glu Asn Thr Ser Val Lys Glu Tyr Lys Glu Tyr Leu Asp Arg Gly Arg
Ala Met Thr Gly Val His Gly Thr Val Leu Val Thr Gly Gly Ala
Gly Tyr Ile Gly Ser His Thr Cys Val Val Leu Glu Lys Gly Tyr
Asp Val Val Ile Val Asp Asn Leu Cys Asn Ser Arg Val Glu Ala Val
His Arg Ile Glu Lys Leu Thr Lys Lys Val Ile Phe His Glu Val
Asp Leu Leu Asp Glu Pro Ala Leu Asp Lys Val Phe Ala Asn Gln Asn
Ile Ser Ala Val Ile His Phe Ala Gly Leu Lys Ala Val Gly Glu Ser
Val Gln Val Pro Leu Ser Tyr Lys Asn Ile Ser Gly Thr Ile
Asn Leu Ile Glu Cys Met Lys Tyr Asn Val Arg Asp Phe Val Phe
Ser Ser Ser Ala Thr Val Tyr Glu Asp Pro Thr Arg Pro Gly Gly Thr
Ile Pro Ile Pro Glu Ser Cys Pro Arg Glu Gly Thr Ser Pro Tyr Gly
Arg Thr Lys Leu Phe Ile Glu Asn Ile Ile Glu Asp Glu Thr Lys Val
Asn Lys Ser Leu Asn Ala Ala Leu Leu Arg Tyr Phe Asn Pro Gly Gly
Ala His Pro Ser Gly Leu Gly Glu Asp Pro Leu Gly Ile Pro Asn
Asn Leu Leu Pro Tyr Ile Ala Gln Val Ala Val Gly Arg Leu Asp His
Leu Asn Val Phe Gly Asp Tyr Pro Thr Ser Asp Gly Thr Pro Ile
Arg Asp Tyr Ile His Val Cys Asp Leu Ala Glu Ala His Val Ala Ala
290 295 300
Leu Asp Tyr Leu Arg Gin His Phe Val Ser Cys Arg Pro Trp Asn Leu
305 310 315 320
Gly Ser Gly Thr Gly Ser Thr Val Phe Gin Val Leu Asn Ala Phe Ser
325 330 335
Lys Ala Val Gly Arg Asp Leu Pro Tyr Lys Val Thr Pro Arg Arg Ala
340 345 350
Gly Asp Val Val Arg Leu Thr Ala Asn Pro Thr Arg Ala Asn Glu Glu
355 360 365
Leu Lys Trp Lys Thr Ser Arg Ser Ile Tyr Glu Ile Cys Val Asp Thr
370 375 380
Trp Arg Trp Gln Gln Lys Tyr Pro Tyr Gly Phe Asp Leu Thr His Thr
385 390 395 400
Lys Thr Tyr Lys Gly Ser Gly Gly Arg Asp Leu Ser Arg Leu Pro
405 410 415
Gln Leu Val Gly Val Ser Thr Pro Leu Gin Gly Ser Arg Ser Ala
420 425 430
 Ala Ala Ile Gly Gin Ser Gly Glu Leu Arg Thr Gly Gly Ala Arg
435 440 445
Pro Pro Pro Pro Leu Gly Ala Ser Ser Gin Pro Arg Pro Gly Gly Asp
450 455 460
Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala Ser Asn Leu Thr
465 470 475 480
Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu Pro Ala Cys Pro
485 490 495
Glu Glu Ser Pro Leu Leu Gly Pro Met Leu Ile Glu Phe Asn Met
500 505 510
Pro Val Asp Leu Glu Leu Val Ala Lys Gin Asn Pro Asn Val Lys Met
515 520 525
Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro His Lys Val Ala
530 535 540
Ile Ile Ile Pro Phe Arg Asn Arg Gin Glu His Leu Tyr Thr Leu
545 550 555 560
Tyr Tyr Leu His Pro Val Leu Gin Arg Gin Gin Leu Asp Tyr Gly Ile
565 570 575
Tyr Val Ile Asn Gin Ala Gly Asp Thr Ile Phe Asn Arg Ala Lys Leu
580 585 590
Leu Asn Val Gly Phe Gin Glu Ala Leu Lys Asp Tyr Asp Tyr Thr Cys
595 600 605
Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn Asp His Asn Ala
610 615 620
Tyr Arg Cys Phe Ser Gin Pro Arg His Ile Ser Val Ala Met Asp Lys
625 630 635 640
Phe Gly Phe Ser Leu Pro Tyr Val Gin Tyr Phe Gly Gly Val Ser Ala
645 650 655
Leu Ser Lys Gin Gin Phe Leu Thr Ile Asn Gly Phe Pro Asn Asn Tyr
660 665 670
Trp Gly Trp Gly Glu Asp Asp Ile Phe Asn Arg Leu Val Phe
675 680 685
Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val Gly Arg Cys Arg
695
<210> SEQ ID NO 10
<211> LENGTH: 2292
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:9

<400> SEQUENCE: 10

atgctgtctga ccacgcgtat ccctaaacggt ttcgaagctga cttcctagctgt gctgatctgg ttg 60

tgtgctgttg cgctgtctac ccacagatc acagagtcac gacagacagac ctgctggtcct gctgatctcg 120

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 180

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 240

gacagctgatg ccagctgatacc ctctgagcag ccgagcgtac ccgagcgtcct gcctg 300

aatgtgatgtg ccagctgatacc ctctgagcag ccgagcgtac ccgagcgtcct gcctg 360

gacaagagcgt acctgaacgagc aacagtattg tcgtgctct ccggcgtccgg 420

gcctgacgac gccctgatccgg gctgagctct gcctgacgac gccctgatccgg 480

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 540

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 600

cgagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 660

GAATTC TCACTAGT AAGCTT GGCCTG AAGCTT 720

gcctgacgac gccctgatccgg gctgagctct gcctgacgac gccctgatccgg 780

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 840

cgagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 900

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 960

gcctgacgac gccctgatccgg gctgagctct gcctgacgac gccctgatccgg 1020

cgagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 1080

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 1140

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 1200

gcctgacgac gccctgatccgg gctgagctct gcctgacgac gccctgatccgg 1260

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 1320

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 1380

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 1440

gcctgacgac gccctgatccgg gctgagctct gcctgacgac gccctgatccgg 1500

aacgattct gagacgcgg ccacagccgtt acgacgtgtc acgagacgac ccagatccctg 1560

acggcgcagg ccgagcactt ccagctgttg gcctgatgag ccgagcgtac ccgagcgtcct gcctg 1620
cacaagttcg ccatcatcat ccctttcaga aaccgacagg agcacttgaa gtactggctg 1680
tactacgta acccgcgtcg gcagccagac cagetggtct acgcctatca ctggtacaaa 1740
caggcccgcc acacatcatc taacccggacc aacctgctga acgtggggctt ccaggggccc 1800
cctgaagggct acagactaca cttctctcgtg tttccgagc tggacactgat ccccatgaca 1860
gacacacacg ccctccccg ttccccccag cccggacaca tctctgtgcg cattggaacag 1920
tctctgctct gctctgccct aagtgcagtc tttgcggtgg tttcggccttt gctattagcag 1980
cagtcctgta ccatacaaggg ctctcccaca aactactggg gctggggggg agagagacac 2040
gacattttcca accggcagg tggcgaggg acgtctatct cttggcccaaa cgccggtttg 2100
gccgtatgtc gaagatcagc acactctctga cgcaagacgacaagcagaa cccccccgca 2160
tttgacagga ttgtccacac taaggaacc atcgtgtctg aagccctgaga ctctctgcac 2220
tacagggtgc tgaagttgca gcgataaccc cttgtacaccc agataaagcgt ggacatggcc 2280
acaccccttt ag 2292

<210> SEQ ID NO 11
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 11
ggatgatcgc acaatggycc tggttctg 28

<210> SEQ ID NO 12
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 12
tgtcttagac tgaattcag aagggttc 27

<210> SEQ ID NO 13
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 13
cgacgatga gcaggttctca cttgtggaaa tgcctg 35

<210> SEQ ID NO 14
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 14
tcaacagtgc gaagtggaaca tcccgcttgg gactg 35

<210> SEQ ID NO 15
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 15

ggatgatcac acaatggccc tggttctg

<210> SEQ ID NO 16
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 16

tgctctagac tagttccaa aggggtc

<210> SEQ ID NO 17
<211> LENGTH: 391
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: fusion of N-terminal portion of the S. cerevisiae Minn2 protein and catalytic domain of rat GlcNAc-transferase II

<400> SEQUENCE: 17

Met Leu Leu Thr Lys Arg Phe Ser Lys Leu Phe Lys Leu Thr Phe Ile 1 5 10 15
Val Leu Ile Leu Cys Gly Leu Phe Val Ile Thr Asn Lys Tyr Met Asp 20 25 30
Glu Asn Thr Ser Ser Leu Val Tyr Gln Leu Asn Phe Asp Gln Met Leu 35 40 45
Arg Asn Val Asp Lys Asp Gly Thr Trp Ser Pro Gly Glu Leu Val Leu 50 55 60
Val Val Glu Val His Asn Arg Pro Gly Tyr Leu Arg Leu Ile Asp 65 70 75 80
Ser Leu Arg Lys Ala Glu Gly Ile Arg Glu Val Leu Val Ile Phe Ser 95 100 105 110
His Asp Phe Trp Ser Ala Glu Ile Asn Ser Leu Ile Ser Ser Val Asp 115 120 125 130
Phe Cys Pro Val Leu Gln Val Phe Pro Phe Ser Ile Glu Tyr 135 140
Pro Ser Glu Phe Pro Gly Ser Asp Pro Arg Asp Cys Pro Arg Asp Leu 145 150 155 160
Lys Lys Asn Ala Ala Leu Lys Leu Gly Cys Ile Asn Ala Glu Tyr Pro 165 170 175
Asp Ser Phe Gly His Tyr Arg Ala Lys Phe Ser Gln Thr Lys His 180 185 190
His Trp Trp Lys Leu His Phe Val Trp Glu Arg Val Lys Val Leu 195 200 205
Gln Asp Tyr Thr Gly Leu Ile Leu Phe Leu Glu Glu Asp His Tyr Leu 210 215 220
Ala Pro Asp Phe Tyr His Val Phe Lys Met Trp Lys Leu Gln Glu 225 230 235 240
Gln Glu Cys Pro Gly Cys Asp Val Leu Ser Leu Gly Thr Tyr Thr Thr
<210> SEQ ID NO 18
<211> LENGTH: 1276
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:17

<400>SEQUENCE: 18
atgcgtgcgga ccagcagatt ctctaagctg ttcagctga ccctcatctg gtgtatcttg 60
tgccgacgcttggtcag ctgctgctac cccagctgct atgcggcgaa acacccgcct gctggtctac 120
cagctggact ggccgcaagct gctgcaagac ggaccaagac acgccacgct ggctccgccgc 180
gagctgagct ggctcgtgca ggccgcaagcc gcacccgagtc accctgagc ctgcgtggtcag 240
ttcctgcgga aggcccgaggt cactcagaggt gtcggtctgct ttccttcctc ccagtccttg 300
tccggcggaga tccacccctgt gatgtctctg gtcggtgcgtc gcacggttcc 360
tcagctctgta gcatccagct gctcctctgg aagtcctccg gctcctccgct ccagacactgt 420
ccccgagacc tgcagagaaag ggcgcccctg aagctggtccgt ctccgaacgc gcgatccccc 480
gacctttctgg gcaactacgcc agagccacgct ttcctccaga ccaagcaaca actgtggtgct 540
aagctgcaactgctgggcga ggcagttgaaag gtcggtcagct ccagctggtcc 600
ttcagtgagaaggagcagct gctggtcagct ctggccgccc gccttgccagcc gcaagactgg 660
aagctgagcagctggctgc ggctggtcagct cctgctgctac ccagctggtcc 720
atcgatcttg ctacagcactg gtcggtccctg cagctggtcagct ctggccgccc gccttgccagcc 780
acacactaaggcgtggccctg cgcagcaggt gctacccgca ctggtggtcagct ccagctggtcc 840
acctctggtcgttgagaaggagcagct gctggtcagct cctgctgctac ccagctggtcc 900
ttgcgtgctgctgctggccctg cgcagcaggt gctacccgca ctggtggtcagct ccagctggtcc 960
gacctggtcgttgagaaggagcagct gctggtcagct cctgctgctac ccagctggtcc 1020
ttcctgctgcagctggtcagct cctgctgctac ccagctggtcc 1080
ttcctgctgcagctggtcagct cctgctgctac ccagctggtcc 1140
caggaactg tgaagttcct acggacgtg cagtag

<210> SEQ ID NO 19
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 19
gtccccgaat tacctttccc

<210> SEQ ID NO 20
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 20
agttagaagtg tgaagsaggt g

<210> SEQ ID NO 21
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: protein tag
<400> SEQUENCE: 21
His Asp Glu Leu
1

<210> SEQ ID NO 22
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 22
gcctcaga cctttggac cgcttaccac c

<210> SEQ ID NO 23
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 23
gccaggtgac cgccctgtcg agaagaagat cg

<210> SEQ ID NO 24
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide
<400> SEQUENCE: 24
gctggaactct tctttcatcc
<210> SEQ ID NO 25
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 25

ggtctcttc agagacagg

<210> SEQ ID NO 26
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 26

ccaagtctca caaggacacc

<210> SEQ ID NO 27
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 27

ccctgacga ccttacagg

<210> SEQ ID NO 28
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 28

gaccagatgc tgcsaaacg

<210> SEQ ID NO 29
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 29

cttcsgctcc acctgtgctg

<210> SEQ ID NO 30
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 30

gcctgaacgg cacgatgca tcgtgcaat cc

<210> SEQ ID NO 31
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 31

caagaagct caggtgctgge gatctccat c 31

<210> SEQ ID NO 32
<211> LENGTH: 763
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO: 33

<400> SEQUENCE: 32

atgaagcttt caacacctt ctctacagcc tctgtagccct cgctgctcggc ccttcccttcctcc 60
cccacatcct cttctgagcc gccggttccct ccaagggcg gggcggggga ccacactaccott 120
acctagttct cctctctcct ctctctctct cttctggtgcc cactctgctca ctcctgtgctca 180
gtttcaggt gccgaagacat cccgttagtgg tggatcagc agagcactgt aaggttctct 240
aagctgctgg cctctctctcct ctcctctctct ctcctctctcct ctcctcctctcct ctcctctcctctcct 300
ttcagctgcc gacccgtgct gtcctctcct ctctcctctcct ctcctctctcct ctcctctcctctcct 360
acctacact gtcacagcagtt tccacacaccc ctctctctct ctcctggtgcc cactctgctca 420
gacagacttt gctggagctctt gctctctctct ctcctctctcct ctcctctctcct ctcctctcctctcct 480
cgtgctggct gtcctctctct ctcctctctct ctcctctctcct ctcctctctcct ctcctctcctctcct 540
aagctgctgg cctctctctcct ctcctctctcct ctcctctctcct ctcctctctcct ctcctctcctctcct 600
rgagactatt ctaagacttt ctctctctct ctcctctctcct ctcctctctcct ctcctctcctctcct 660
gacagacttt ctcctctctct ctcctctctcct ctcctctctcct ctcctctctcct ctcctctcctctcct 720
gtacagcagtt ctcctctctct ctcctctctcct ctcctctctcct ctcctctctcct ctcctctcctctcct 753

<210> SEQ ID NO 33
<211> LENGTH: 250
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: fusion of Lip2 protein prepro signal and anti-HER2 antibody light chain

<400> SEQUENCE: 33

Met Lys Leu Ser Thr Ile Leu Phe Thr Ala Cys Ala Thr Leu Ala Ala 1 5 10 15
Ala Leu Pro Ser Pro Ile Thr Pro Ser Glu Ala Ala Val Leu Gln Lys 20 25 30
Arg Gly Gly Asp Ile Gin Met Thr Gin Ser Pro Ser Ser Ser Leu Ser 35 40 45
Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gin Asp 50 55 60
Val Aem Thr Ala Val Ala Trp Tyr Gin Gin Lys Pro Gly Lys Ala Pro 65 70 75 80
Lys Leu Leu Ile Tyr Ser Ala Ser Phe Leu Tyr Ser Gin Val Pro Ser 90 95 100
Arg Phe Ser Gly Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser 105 110
Ser Leu Gin Pro Gin Asp Phe Ala Thr Tyr Tyr Cys Gin Gin His Tyr 115 120 125
Thr Thr Pro Pro Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg
130 135 140
Thr Val Ala Ala Pro Ser Val Phe Ile Phe Pro Pro Ser Asp Glu Gln
145 150 155 160
Leu Lys Ser Gly Thr Ala Ser Val Val Cys Leu Leu Asn Asn Phe Tyr
165 170 175
Pro Arg Glu Ala Lys Val Gln Trp Lys Val Asp Ala Leu Gln Ser
180 185 190
Gly Asn Ser Gln Glu Ser Val Thr Glu Gln Asp Ser Thr
195 200 205
Tyr Ser Leu Ser Ser Thr Leu Ser Ser Thr Leu Ser Lys Ala Asp Tyr Glu Lys
210 215 220
His Lys Val Tyr Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro
225 230 235 240
Val Thr Lys Ser Phe Asn Arg Arg Gly Lys Cys
245 250

<210> SEQ ID NO 34
<211> LENGTH: 1461
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:36

<400> SEQUENCE: 34
atgaagcttt cccacatctt ttcocagcc tgggtctacc cttggtgcgc accoccttcc 60
cctccatctt cttctgaggt ccgaattctt cagaagcgag gcggggcggc ggttcgagctg 120
gtgcgctgg ctggagagct gcgtctgctcc gcgctgactt ggtggctgct 180
tggctgacct acatacaggg cacatcactt cattgggtcc gacaggtctc gcggaggag 240
cggagtgtgg ttggtcagtg tcatatcctt cacgggtcc tgcgtggtgctgct 300
aagggacgt gaattgggtct cggccgacgg ttatgacata ctctgctacat ctgggtgagc 360
tctcttgagagct tgggtcagct gcgtctggtg tcatatcctt cggccgaccgtgctgctt 420
tcccctgtcc gcccccgtcc tccatgttcct ctcgctgctgg tcatatcctt ctcgctgctgct 480
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 540
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 600
gtgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 660
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 720
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 780
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 840
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 900
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 960
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1020
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1080
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1140
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1200
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1260
cgggtctttgg gtgggtcagct gattgcttc ccggcggtgcc ttcgctgctgg tcatatcctt 1320
<table>
<thead>
<tr>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Lys Leu Ser Thr Ile Leu Phe Thr Ala Cys Ala Thr Leu Ala Ala</td>
</tr>
<tr>
<td>Ala Leu Pro Ser Pro Ile Thr Pro Ser Glu Ala Ala Val Leu Glu Lys</td>
</tr>
<tr>
<td>Arg Gly Gly Glu Val Glu Leu Val Glu Ser Gly Gly Gly Leu Val</td>
</tr>
<tr>
<td>Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn</td>
</tr>
<tr>
<td>Ile Lys Asp Thr Tyr Ile His Trp Val Arg Glu Ala Pro Gly Lys Gly</td>
</tr>
<tr>
<td>Leu Glu Trp Val Ala Arg Ile Tyr Pro Thr Asn Gly Tyr Thr Arg Tyr</td>
</tr>
<tr>
<td>Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Ala Asp Thr Ser Lys</td>
</tr>
<tr>
<td>Asn Thr Ala Tyr Leu Glu Met Asn Ser Leu Arg Ala Glu Asp Thr Ala</td>
</tr>
<tr>
<td>Val Tyr Tyr Cys Ser Arg Trp Gly Asp Gly Phe Tyr Ala Met Asp</td>
</tr>
<tr>
<td>Tyr Trp Gly Glu Gly Thr Leu Val Thr Val Ser Ala Ser Thr Lys</td>
</tr>
<tr>
<td>Gly Pro Ser Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly</td>
</tr>
<tr>
<td>Gly Thr Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro</td>
</tr>
<tr>
<td>Val Thr Val Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr</td>
</tr>
<tr>
<td>Phe Pro Ala Val Leu Gln Ser Gly Leu Tyr Ser Leu Ser Val</td>
</tr>
<tr>
<td>Val Thr Val Pro Ser Ser Leu Gly Thr Glu Thr Tyr Ile Cys Asn</td>
</tr>
<tr>
<td>Val Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys Val Glu Pro</td>
</tr>
<tr>
<td>Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu</td>
</tr>
<tr>
<td>Leu Leu Gly Gln Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp</td>
</tr>
<tr>
<td>Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Asp</td>
</tr>
<tr>
<td>Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Thr Tyr Val Asp Gly</td>
</tr>
</tbody>
</table>
1. A method of producing a fungal cell capable of producing proteins comprising GlcNAcMan, GlcNAc, N-glycans, said method comprising:  
a. providing a fungal cell genetically engineered to produce proteins comprising Man, GlcNAc, N-glycans;  
b. introducing into said cell a nucleic acid encoding a GlcNAc-transferase I, wherein said nucleic acid comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment, wherein expression of said GlcNAc-transferase I in said fungal cell produces proteins comprising GlcNAcMan, GlcNAc, N-glycans.  
2. The method of claim 1, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GlcNAcMan, GlcNAc, N-glycans.  
3. The method of claim 2, wherein said target protein binds to an Fc receptor.  
4. The method of claim 3, wherein said target protein is an antibody or fragment thereof.  
5. The method of claim 2, wherein said target protein is a therapeutic glycoprotein.  
6. The method of claim 2, wherein said target protein is Interferon-β, GM-CSF, Interferon γ, or erythropoietin.  
7. The method of claim 1, wherein said intracellular compartment is the Golgi apparatus.  
8. The method of claim 1, wherein said fungal cell genetically engineered to produce proteins comprising Man, GlcNAc, N-glycans is deficient in OCH1 activity and comprises a nucleic acid encoding an α-1,2-mannosidase, wherein said nucleic acid encoding said α-1,2-mannosidase comprises a nucleotide sequence encoding a targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum.  
9. The method of claim 8, wherein said targeting sequence is an HDEL sequence.  
10. The method of claim 1, wherein said fungal cell is Yarrowia lipolytica or Arxula adeninivorans.  
11. The method of claim 1, said method further comprising introducing into said cell a nucleic acid encoding a mannosidase II, wherein said nucleic acid encoding said mannosidase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded mannosidase II to the Golgi apparatus, wherein expression of said mannosidase II in said fungal cell produces proteins comprising GlcNAcMan, GlcNAc, N-glycans.  
12. The method of claim 1 or claim 11, said method further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising GalGlcNAcMan, GlcNAc, or GalGlcNAcMan, GlcNAc, N-glycans.  
13. The method of claim 12, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GalGlcNAcMan, GlcNAc, or GalGlcNAcMan, GlcNAc, N-glycans.  
14. The method of claim 12, wherein said galactosyltransferase is a fusion of a UDP-Glc-4-epimerase and the catalytic domain of a β-1,4-galactosyltransferase I.  
15. The method of claim 13, further comprising isolating said target protein modified to comprise said GalGlcNAcMan, GlcNAc, or GalGlcNAcMan, GlcNAc, N-glycans.
16. A method of producing a target protein comprising GlcNAcMan$_5$GlcNAc$_2$ N-glycans, said method comprising: a) providing a fungal cell genetically engineered to comprise a nucleic acid encoding a GlcNAc-transferase I, an α-1,2-mannosidase, and a mannosidase II, wherein said nucleic acid comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein said fungal cell is deficient in OCH1 activity; and b) introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said GlcNAcMan$_5$GlcNAc$_2$ N-glycans.

17. The method of claim 16, wherein said fungal cell is Yarrowia lipolytica or Arxula adeninivorans.

18. The method of claim 16, wherein said nucleic acid encoding said α-1,2-mannosidase comprises an endoplasmic reticulum targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum.

19. The method of claim 18, wherein said targeting sequence is an HDEL sequence.

20. The method of claim 16, wherein said nucleic acid encoding said GlcNAc-transferase I and said mannosidase II comprises nucleotide sequences encoding Golgi targeting sequences to target the encoded GlcNAc-transferase I and mannosidase II to the Golgi apparatus.

21. The method of claim 16, wherein said target protein binds to an Fe receptor.

22. The method of claim 16, wherein said target protein is an antibody or fragment thereof.

23. The method of claim 16, wherein said target protein is a therapeutic glycoprotein.

24. The method of claim 16, wherein said target protein is Interferon-β, GM-CSF, Interferon γ, or erythropoietin.

25. The method of claim 16, further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces said target protein modified to comprise GalGlcNAcMan$_5$GlcNAc$_2$N-glycans.

26. The method of claim 25, further comprising isolating said target protein modified to comprise said GalGlcNAcMan$_5$GlcNAc$_2$N-glycans.

27. A method of making a fungal cell capable of producing proteins comprising GlcNAcMan$_5$GlcNAc$_2$N-glycans, said method comprising: a) providing a fungal cell genetically engineered to produce proteins comprising Man$_5$GlcNAc$_2$N-glycans; b) introducing into said cell a nucleic acid encoding a GlcNAc-transferase I, wherein said nucleic acid comprises a nucleotide sequence encoding a targeting sequence to target said encoded GlcNAc-transferase I to an intracellular compartment, wherein expression of said GlcNAc-transferase I in said fungal cell produces proteins comprising GlcNAcMan$_5$GlcNAc$_2$N-glycans.

28. The method of claim 27, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GlcNAcMan$_5$GlcNAc$_2$N-glycans.

29. The method of claim 27, wherein said target protein binds to an Fe receptor.

30. The method of claim 27, wherein said target protein is an antibody or fragment thereof.

31. The method of claim 27, wherein said target protein is a therapeutic glycoprotein.

32. The method of claim 27, wherein said target protein is Interferon-β, GM-CSF, Interferon γ, or erythropoietin.

33. The method of claim 27, wherein said intracellular compartment is the Golgi apparatus.

34. The method of claim 27, wherein said fungal cell genetically engineered to produce proteins comprising Man$_5$GlcNAc$_2$N-glycans is deficient in ALG3 activity, and comprises a nucleic acid encoding an α-1,2-mannosidase, said nucleic acid comprising a nucleotide sequence encoding a targeting sequence to target the encoded α-1,2-mannosidase to the endoplasmic reticulum.

35. The method of claim 34, wherein said fungal cell genetically engineered to produce proteins comprising Man$_5$GlcNAc$_2$N-glycans further is deficient in OCH1 activity.

36. The method of claim 34 or 35, wherein said fungal cell genetically engineered to produce proteins comprising Man$_5$GlcNAc$_2$N-glycans further comprises a nucleic acid encoding an α-1,3-glucosyltransferase.

37. The method of claim 36, wherein said α-1,3-glucosyltransferase is ALG6.

38. The method of claim 27, wherein said fungal cell is Yarrowia lipolytica or Arxula adeninivorans.

39. The method of claim 27, further comprising introducing into said cell a nucleic acid encoding a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of said GlcNAc-transferase II in said fungal cell produces proteins comprising GlcNAcMan$_5$GlcNAc$_2$N-glycans.

40. The method of claim 27 or claim 39, said method further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising GalGlcNAcMan$_5$GlcNAc$_2$ or GalGlcNAcMan$_5$GlcNAc$_2$N-glycans.

41. The method of claim 40, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GalGlcNAcMan$_5$GlcNAc$_2$ or GalGlcNAcMan$_5$GlcNAc$_2$N-glycans.

42. The method of claim 40, wherein said galactosyltransferase is a fusion of a UDP-Glc-4-epimerase and catalytic domain of a β-1,4-galactosyltransferase I.

43. The method of claim 40 said method further comprising introducing into said cell a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces proteins comprising GalGlcNAcMan$_5$GlcNAc$_2$ or GalGlcNAcMan$_5$GlcNAc$_2$N-glycans.

44. A method of producing a target protein comprising Gal$_5$GlcNAc$_2$Man$_5$GlcNAc$_2$N-glycans, said method comprising: a) providing a fungal cell genetically engineered to be deficient in ALG3 activity and comprising a nucleic acid...
encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment;

b) introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

44. The method of claim 44, wherein said fungal cell further is deficient in OCH1 activity.

46. The method of claim 44 or 45, wherein said fungal cell further comprises a nucleic acid encoding an α-1,3-galactosyltransferase.

47. The method of claim 46, wherein said nucleic acid encoding said α-1,3-galactosyltransferase is ALG6.

48. The method of claim 46, wherein said fungal cell further comprises a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said target protein comprising said Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

49. An isolated fungal cell genetically engineered to produce proteins comprising GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans, wherein said fungal cell is deficient in OCH1 activity and comprises a nucleic acid encoding an α-1,2-mannosidase, a GlcNAc-transferase I, and a mannosidase II, wherein said nucleic acid encoding said α-1,2-mannosidase, said GlcNAc-transferase I, and said mannosidase II comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said α-1,2-mannosidase, said GlcNAc-transferase I, and said mannosidase II in said fungal cell produces proteins comprising GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

50. The fungal cell of claim 49, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

51. The fungal cell of claim 49 or claim 50, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of said GlcNAc-transferase II in said fungal cell produces proteins comprising GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

52. The fungal cell of claim 51, said fungal cell further comprising a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

53. An isolated fungal cell genetically engineered to produce proteins comprising GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans, wherein said fungal cell is genetically engineered to be deficient in ALG3 activity and comprises a nucleic acid encoding a GlcNAc-transferase I and a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase I and said GlcNAc-transferase II comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said GlcNAc-transferase I and said GlcNAc-transferase II in said fungal cell produces proteins comprising GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

54. The fungal cell of claim 53, wherein said genetically engineered fungal cell further is deficient in OCH1 activity.

55. The fungal cell of claim 53 or claim 54, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding an α-1,3-galactosyltransferase.

56. The fungal cell of claim 53, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

57. The fungal cell of claim 53, said fungal cell further comprising a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said protein comprising said GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

58. The fungal cell of claim 57, said fungal cell further comprising a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

59. A substantially pure culture of Yarrowia lipolytica cells, a substantial number of which are genetically engineered to produce glycoproteins comprising Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans, wherein said cells are genetically engineered to be deficient in ALG3 activity and comprises a nucleic acid encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase in said cell produces proteins comprising Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

60. The substantially pure culture of claim 59, wherein said cells further are deficient in OCH1 activity.

61. The substantially pure culture of claim 59 or claim 60, wherein said cells further comprise a nucleic acid encoding an α-1,3-galactosyltransferase.

62. The substantially pure culture of claim 61, wherein said cells further comprise a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said target protein comprising said Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans.

63. A substantially pure culture of Yarrowia lipolytica cells, a substantial number of which are genetically engineered to produce glycoproteins comprising Gal\textsubscript{1}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{3} N-glycans, wherein said cells are genetically engineered to be deficient in OCH1 activity and comprise a nucleic acid encoding an α-1,2-mannosidase,
a GlcNAc-transferase I, a mannosidase II, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said α-1,2-mannosidase, said GlcNAc-transferase I, said mannosidase II, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said α-1,2-mannosidase, said GlcNAc-transferase I, said mannosidase II, said GlcNAc-transferase II, and a galactosyltransferase in said cells produces proteins comprising Gal\textsubscript{2}GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{2}N-glycans.

64. A composition comprising a glycoprotein, wherein at least 50% of the N-glycans on said glycoprotein are GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{2}N-glycans.

65. The composition of claim 64, wherein at least 70% of the N-glycans on said glycoprotein are GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{2}N-glycans.

66. The composition of claim 64, wherein at least 85% of the N-glycans on said glycoprotein are GlcNAc\textsubscript{2}Man\textsubscript{3}GlcNAc\textsubscript{2}N-glycans.