(19) IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

(11) Publication number: **SG 190321 A1**
(43) Publication date: **28.06.2013**
(51) Int. Cl: **C12N 9/10, C12P 21/00, C12N 5/10, C12N 1/14, C12N 15/80;**

(12) Patent Application

(21) Application number: **2013038088**
(22) Date of filing: **24.11.2011**
(30) Priority: **US 61/417,144 24.11.2010**

(71) Applicant: **NOVARTIS INTERNATIONAL PHARMACEUTICAL LTD. 131 FRONT STREET HAMILTON BM**
GLYKOS FINLAND OY VIIKINKAARI 6 FI-00790 HELSINKI FI

(72) Inventor: **NATUNEN, JARI OOLANNINTIE 10 E 18 FI-01520 VANTAA FI**
KANERVA, ANNE RIIHIKALLIONKUJA 8 FI-00890 HELSINKI FI
HILTUNEN, JUKKA HAKANIEMENRANTA 28 B 72 FI-00530 HELSINKI FI
SALOHEIMO, MARKKU KYNTÄJÄNTIE 18B FI-00390 HELSINKI FI
VISKARI, HELI PENKASKUJA 3 FI-03100 NUMMELA FI
HUUSKONEN, ANNE MECHELININKATU 47 A 29 FI-00250 HELSINKI FI

(54) **Title:**
FUSION ENZYMES HAVING N-ACETYLGLUCOSAMINYLTRANSFERASE ACTIVITY

(57) **Abstract:**
The present disclosure relates to recombinant proteins having N-acetylglucosaminyltransferase activity. The present disclosure further relates to methods for producing complex N-glycans including the steps of providing host cells containing such recombinant proteins and culturing the host cells such that the recombinant proteins are expressed.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2012 (31.05.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/069593 A2

(51) International Patent Classification:
C07K 14/435 (2006.01)

(21) International Application Number:
PCT/EP2011/070956

(22) International Filing Date:
24 November 2011 (24.11.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/417,144 24 November 2010 (24.11.2010) US

(71) Applicants *(for all designated States except US)*: **NOVARTIS INTERNATIONAL PHARMACEUTICAL LTD.**; 131 Front Street, Hamilton (BM). **GLYKOS FINLAND Oy** [FI/FI]; Viikinkaari 6, FI-00790 Helsinki (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **NATUNEN, Jari** [FI/FI]; Oolannintie 10 E 18, FI-01520 Vantaa (FI). **KANERVA, Anne** [FI/FI]; Riihikallionkuja 8, FI-00890 Helsinki (FI). **HILTUNEN, Jukka** [FI/FI]; Hakaniemenranta 28 B 72, FI-00530 Helsinki (FI). **SALOHEIMO, Markku** [FI/FI]; Kyntäjäntie 18b, FI-00390 Helsinki (FI). **VISKARI, Heli** [FI/FI]; Penkaskuja 3, FI-03100 Nummela (FI). **HUUSKONEN, Anne** [FI/FI]; Mechelininkatu 47 A 29, FI-00250 Helsinki (FI).

(74) Agent: **CABINET PLASSERAUD**; Group No. 280, 235 cours Lafayette, F-69006 Lyon (FR).

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

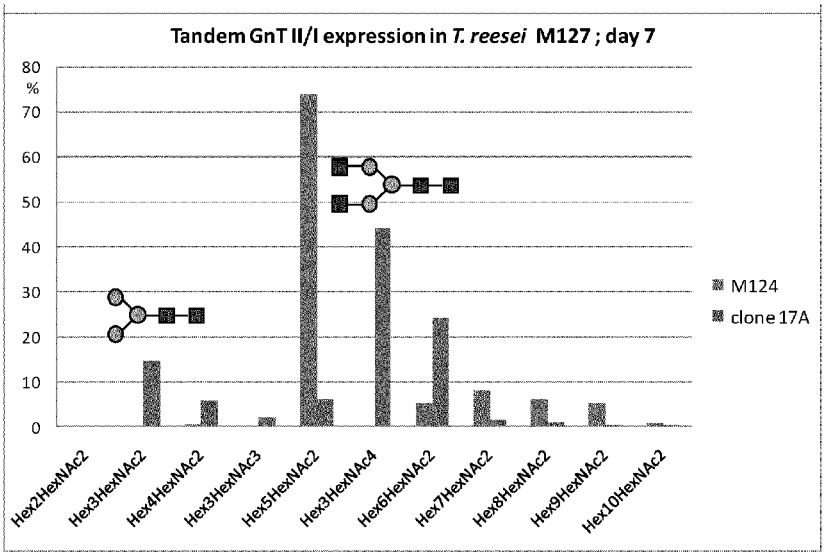
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: FUSION ENZYMES

Figure 27

D.



(57) Abstract: The present disclosure relates to recombinant proteins having N-acetylglucosaminyltransferase activity. The present disclosure further relates to methods for producing complex N-glycans including the steps of providing host cells containing such recombinant proteins and culturing the host cells such that the recombinant proteins are expressed.

WO 2012/069593 A2

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

Published:

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

FUSION ENZYMES

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of U.S. Provisional Application No. 61/417,144, filed November 24, 2010, which is hereby incorporated by reference in its entirety.

SUBMISSION OF SEQUENCE LISTING ON ASCII TEXT FILE

[0002] The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 619672001040SEQLIST.txt, date recorded: November 22, 2011, size: 305 KB).

FIELD OF THE INVENTION

[0003] The present disclosure relates to compositions and methods useful for the production of N-glycans.

BACKGROUND

[0004] Posttranslational modification of proteins is often necessary for proper protein folding and function. A common protein modification is the addition of oligosaccharides (glycans) to nascent polypeptides in the endoplasmic reticulum to form glycoproteins, a process known as glycosylation. N-glycosylation is of particular importance in the production of recombinant proteins used for therapeutic purposes. Because standard prokaryotic expression systems lack the proper machinery necessary for such modifications, alternative expression systems have to be used in production of these therapeutic proteins. Yeast and fungi are attractive options for expressing proteins as they can be easily grown at a large scale in simple media, which allows low production costs. Moreover, tools are available to manipulate the relatively simple genetic makeup of yeast and fungal cells as well as more complex eukaryotic cells such as mammalian or insect cells (De Pourcq et al., *Appl Microbiol Biotechnol*, 87(5):1617-31).

[0005] Fungal cells and mammalian cells share common steps in the early stages of glycosylation that result in the formation of mannose(8)N-acetylglucosamine(2) (Man8GlcNAc2). However, significant differences exist in the later stages of the process.

For example, in yeast, additional mannose subunits are added to Man8GlcNAc2 by mannosyltransferases and mannan polymerases to yield high-mannose type N-glycans. In contrast, mannose sugars are removed from the human Man8GlcNAc2 to yield Man5GlcNAc2, followed by three sequential reactions involving the enzymes N-acetylglucosaminyltransferase I (GnTI), mannosidase II (Mns II), and N-acetylglucosaminyltransferase II (GnTII), to convert Man5GlcNAc2 into GlcNAc2Man3GlcNAc2.

[0006] The differences between the glycosylation process in mammalian and fungal cells pose a challenge to the expression of glycosylated mammalian proteins in fungal cells since glycoproteins with high-mannose type N-glycans are not suitable for therapeutic use in humans (De Pourcq et al., 2010; Wildt and Gerngross, *Nature Reviews Microbiology*, 3: 119-128). Consequently, studies have been conducted to re-engineer the glycosylation pathways in yeast and fungal species to enable them to express recombinant human proteins. The general approach in glycoengineering of yeast or fungal cells has been to disrupt endogenous genes that are involved in formation of high-mannose type N-glycans. These gene disruptions can be combined with over-expression of endogenous mannosidases and/or glycosyltransferases and glycosidases from different species (Chiba et al., 1998, *J Biol Chem* 273: 26298-304; Kainz et al., 2008, *Appl Environ Microbiol* 74: 1076-86; Maras et al., 1997, *Euro J Biochem* 249: 701-07; Maras et al., 1999, *Febs Letters* 452: 365-70; Hamilton et al., 2003, *Science* 301: 1244-6; De Pourcq et al., 2010). However, the production of glycosylated mammalian proteins in non-mammalian cells still requires complicated and time-consuming genetic engineering and can be inefficient at producing a desired glycoprotein.

[0007] Thus, a need remains in the art for a simpler and more efficient system to express complex N-glycans in non-mammalian cells.

SUMMARY

[0008] Described herein are compositions including recombinant proteins having N-acetylglucosaminyltransferase activity. Further described herein are methods of producing complex N-glycans and methods of producing Man3GlcNAc2 glycans.

[0009] Thus one aspect includes recombinant proteins having N-acetylglucosaminyltransferase activity, where the recombinant proteins catalyze the

transfer of N-acetylglucosamine to a terminal Manα3 residue and catalyze the transfer of N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan, and where the recombinant protein contains catalytic domains from at least two different enzymes. In certain embodiments, the acceptor glycan is attached to a molecule selected from an amino acid, a peptide, or a polypeptide. In certain embodiments, the molecule is a heterologous polypeptide. In certain embodiments that may be combined with the preceding embodiments, the acceptor glycan is Man3. In certain embodiments that may be combined with the preceding embodiments, the recombinant protein is a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain are from human enzymes. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain includes a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues 105-445 of SEQ ID NO: 1. In certain embodiments that may be combined with the previous embodiments, the N-acetylglucosaminyltransferase II catalytic domain includes a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical amino acid residues 30-447 of SEQ ID NO: 21. In certain embodiments that may be combined with the preceding embodiments, the N-acetylglucosaminyltransferase I catalytic domain is N-terminal to the N-acetylglucosaminyltransferase II catalytic domain. In certain embodiments that may be combined with the preceding embodiments, the N-acetylglucosaminyltransferase II catalytic domain is N-terminal to the N-acetylglucosaminyltransferase I catalytic domain.

[0010] In certain embodiments that may be combined with the preceding embodiments, the recombinant proteins further contain a spacer in between the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain. In certain embodiments, the spacer contains sequence from a stem domain. In certain embodiments that may be combined with the preceding embodiments, the spacer is at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, or at least 50 amino acids in length. In certain embodiments that may be combined with the preceding embodiments, the spacer contains a sequence

that is selected from SEQ ID NO: 118, SEQ ID NO: 120, SEQ ID NO: 122, and SEQ ID NO: 124. In certain embodiments, the spacer contains a sequence that is selected from SEQ ID NO: 118, SEQ ID NO: 120, and SEQ ID NO: 124. In certain embodiments, the spacer contains the sequence of SEQ ID NO: 120 or SEQ ID NO: 124. In certain embodiments, the spacer contains the sequence of SEQ ID NO: 124.

[0011] In certain embodiments that may be combined with the preceding embodiments, the recombinant proteins further contain a targeting peptide linked to the N-terminal end of the catalytic domains. In certain embodiments, the targeting peptide contains a stem domain. In certain embodiments, the stem domain is from an N-acetylglucosaminyltransferase I enzyme or an N-acetylglucosaminyltransferase II enzyme. In certain embodiments, the N-acetylglucosaminyltransferase I enzyme and the N-acetylglucosaminyltransferase II enzyme are human enzymes. In certain embodiments that may be combined with the preceding embodiments, the stem domain is from a protein selected from a mannosidase, a mannosyltransferase, a glycosyltransferase, a Type 2 Golgi protein, MNN2, MNN4, MNN6, MNN9, MNN10, MNS1, KRE2, VAN1, or OCH1. In certain embodiments, the protein is from an organism selected from *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, or *Trichoderma*. In certain embodiments that may be combined with the preceding embodiments, the targeting peptide is a Kre2 targeting peptide. In certain embodiments, the targeting peptide contains a transmembrane domain. In certain embodiments that may be combined with the preceding embodiments, the targeting peptide further contains a transmembrane domain linked to the N-terminal end of the stem domain. In certain embodiments that may be combined with the preceding embodiments, the transmembrane domain is from an N-acetylglucosaminyltransferase I enzyme or an N-acetylglucosaminyltransferase II enzyme. In certain embodiments, the N-acetylglucosaminyltransferase I enzyme and the N-acetylglucosaminyltransferase II enzyme are human enzymes. In certain embodiments that may be combined with the preceding embodiments, the transmembrane domain is from a protein selected from a mannosidase, a mannosyltransferase, a glycosyltransferase, a Type 2 Golgi protein, MNN2, MNN4, MNN6, MNN9, MNN10, MNS1, KRE2, VAN1, or OCH1. In certain

embodiments, the protein is from an organism selected from *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, or *Trichoderma*. In certain embodiments, the targeting peptide contains a cytoplasmic domain. In certain embodiments that may be combined with the preceding embodiments, the targeting peptide further contains a cytoplasmic domain linked to the N-terminal end of the stem domain. In certain embodiments that may be combined with the preceding embodiments, the targeting peptide further contains a cytoplasmic domain linked to the N-terminal end of the transmembrane domain. In certain embodiments that may be combined with the preceding embodiments, the cytoplasmic domain is from an N-acetylglucosaminyltransferase I enzyme or an N-acetylglucosaminyltransferase II enzyme. In certain embodiments, the N-acetylglucosaminyltransferase I enzyme and the N-acetylglucosaminyltransferase II enzyme are human enzymes. In certain embodiments that may be combined with the preceding embodiments, the cytoplasmic domain is from a protein selected from a mannosidase, a mannosyltransferase, a glycosyltransferase, a Type 2 Golgi protein, MNN2, MNN4, MNN6, MNN9, MNN10, MNS1, KRE2, VAN1, or OCH1. In certain embodiments, the protein is from an organism selected from *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, or *Trichoderma*.

[0012] Another aspect includes recombinant proteins containing a human N-acetylglucosaminyltransferase II catalytic domain and a human N-acetylglucosaminyltransferase I catalytic domain where the N-acetylglucosaminyltransferase II catalytic domain is located N-terminal to the N-acetylglucosaminyltransferase I catalytic domain, a spacer sequence containing sequence from a human N-acetylglucosaminyltransferase I stem domain located in between the catalytic domains, and a targeting peptide located N-terminal to the N-acetylglucosaminyltransferase II catalytic domain where the targeting peptide contains a cytoplasmic domain, a transmembrane domain, and a stem domain from human N-

acetylglucosaminyltransferase II. Another aspect includes a recombinant protein containing a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 95.

[0013] Another aspect includes recombinant proteins containing N-acetylglucosaminyltransferase II catalytic domain and a N-acetylglucosaminyltransferase I catalytic domain, where the N-acetylglucosaminyltransferase II catalytic domain is located N-terminal to the N-acetylglucosaminyltransferase I catalytic domain; a spacer located in between the catalytic domains, where the spacer contains a sequence selected from SEQ ID NO: 118, SEQ ID NO: 120, SEQ ID NO: 122, and SEQ ID NO: 124; and a targeting peptide located N-terminal to the N-acetylglucosaminyltransferase II catalytic domain where the targeting peptide contains a cytoplasmic domain, a transmembrane domain, and a stem domain from human N-acetylglucosaminyltransferase II. In certain embodiments, the spacer contains a sequence that is selected from SEQ ID NO: 118, SEQ ID NO: 120, and SEQ ID NO: 124. In certain embodiments, the spacer contains the sequence of SEQ ID NO: 120 or SEQ ID NO: 124. In certain embodiments, the spacer contains the sequence of SEQ ID NO: 124.

[0014] Another aspect includes isolated polynucleotides encoding the recombinant protein of any of the preceding embodiments. Another aspect includes expression vectors containing the isolated polynucleotide of the preceding embodiment operably linked to a promoter. In certain embodiments, the promoter is a constitutive promoter. In certain embodiments, the promoter is an inducible promoter. In certain embodiments, the promoter is from a gene selected from gpdA, cbh1, *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* glucoamylase (glaA), *Aspergillus awamori* glaA, *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Aspergillus oryzae* acetamidase, *Fusarium oxysporum* trypsin-like protease, fungal endo α-L-arabinase (abnA), fungal α-L-arabinofuranosidase A (abfA), fungal α-L-arabinofuranosidase B (abfB), fungal xylanase (xlnA), fungal phytase, fungal ATP-synthetase, fungal subunit 9 (oliC), fungal triose phosphate isomerase (tpi), fungal alcohol dehydrogenase (adhA), fungal α-amylase (amy), fungal amyloglucosidase (glaA), fungal

acetamidase (amdS), fungal glyceraldehyde-3-phosphate dehydrogenase (gpd), yeast alcohol dehydrogenase, yeast alcohol oxidase, yeast lactase, yeast 3-phosphoglycerate kinase, yeast triosephosphate isomerase, bacterial α-amylase, bacterial Spo2, or SSO. Another aspect includes host cells containing the expression vector of any of the preceding embodiments.

[0015] Another aspect includes methods of producing the recombinant protein of any the preceding embodiments, including the steps of introducing an isolated polynucleotide that encodes the recombinant protein into a host cell, and culturing the host cell such that the recombinant protein is expressed. In certain embodiments, the methods further include a step of purifying the recombinant protein from the host cell. In certain embodiments that may be combined with the preceding embodiments, the host cell is a fungal cell. In certain embodiments, the fungal cell is selected from yeast or filamentous fungus.

[0016] Another aspect includes methods of producing a complex N-glycan including the steps of providing a host cell, where the host cell contains a polynucleotide encoding a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, and culturing the host cell such that the fusion protein is expressed, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan. In certain embodiments, the complex N-glycan is attached to a molecule selected from an amino acid, a peptide, or a polypeptide. In certain embodiments, the molecule is a heterologous polypeptide. In certain embodiments that may be combined with the preceding embodiments, the acceptor glycan is Man3. In certain embodiments that may be combined with the preceding embodiments, the complex N-glycan is GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4GlcNAc. In certain embodiments that may be combined with the preceding embodiments, the host cell is a eukaryotic cell. In certain embodiments that may be combined with the preceding embodiments, the host cell is a fungal cell. In certain embodiments, the fungal cell is a yeast cell selected from *S. cerevisiae*, *K. lactis*, *P. pastoris*, *H. polymorpha, C. albicans, Schizosaccharomyces*, or *Yarrowia*. In certain embodiments that may be combined with the preceding embodiments, the fungal cell is a filamentous fungal cell selected from *Trichoderma* sp., *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*,

Chrysosporium, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, or *Tolypocladium*. In certain embodiments that may be combined with the preceding embodiments, the host cell further contains a polynucleotide encoding a UDP-GlcNAc transporter. In certain embodiments that may be combined with the preceding embodiments, the host cell has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase compared to the level of activity in a wild-type host cell. In certain embodiments, the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *alg3* gene is deleted from the host cell. In certain embodiments that may be combined with the preceding embodiments, the host cell has a reduced level of activity of an α-1,6-mannosyltransferase compared to the level of activity in a wild-type host cell. In certain embodiments, the host cell has a reduced level of expression of an *och1* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *och1* gene is deleted from the host cell. In certain embodiments that may be combined with the preceding embodiments, the host cell further contains a polynucleotide encoding an α-1,2-mannosidase. In certain embodiments that may be combined with the preceding embodiments, the host cell further contains a polynucleotide encoding a β-1,4-galactosyltransferase. In certain embodiments that may be combined with the preceding embodiments, the host cell further contains a polynucleotide encoding a sialyltransferase. In certain embodiments that may be combined with the preceding embodiments, the host cell is a *Trichoderma* cell that has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase compared to the level of activity in a wild-type *Trichoderma* cell. In certain embodiments that may be combined with the preceding embodiments, the host cell is a yeast or fungal cell that has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase and a reduced level of activity of an alpha-1,6-mannosyltransferase compared to the levels of activity in a wild-type yeast cell and further contains a polynucleotide encoding an α-1,2-mannosidase.

[0017] Another aspect includes methods of producing a complex N-glycan including the steps of providing a *Trichoderma* host cell, where the host cell has a reduced level of

expression of an *alg3* gene compared to the level of expression in a wild-type host cell and contains a first polynucleotide encoding an N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding an N-acetylglucosaminyltransferase II catalytic domain, and culturing the host cell to produce a complex N-glycan.

[0018] Another aspect includes methods of producing a complex N-glycan including the steps of incubating a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, an acceptor glycan, and an N-acetylglucosamine donor together in a buffer, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan. In certain embodiments, the acceptor glycan is attached to a molecule selected from an amino acid, a peptide, or a polypeptide. In certain embodiments, the molecule is a heterologous polypeptide. In certain embodiments, the acceptor glycan is Man3. In certain embodiments that may be combined with the preceding embodiments, the N-acetylglucosamine donor is a UDP-GlcNAc transporter.

[0019] Another aspect includes filamentous fungal cells containing a mutation of *alg3* and Man3GlcNAc2, where the Man3GlcNAc2 includes at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% (mol %) of neutral N-glycans secreted by the cells. The neutral N-glycans may be attached to a molecule selected from the group consisting of an amino acid, a peptide, and a polypeptide. In certain embodiments, the mutation of *alg3* is a deletion of *alg3*. In certain embodiments that may be combined with the preceding embodiments, the cell is a *Trichoderma reesei* cell. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a first polynucleotide encoding an N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding an N-acetylglucosaminyltransferase II catalytic domain. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a polynucleotide encoding a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain.

[0020] Another aspect includes methods of producing a Man3GlcNAc2 glycan in a host cell including the steps of providing a host cell with a reduced level of activity of a mannosyltransferase compared to the level of activity in a wild-type host cell and culturing the host cell to produce a Man3GlcNAc2 glycan, where the Man3GlcNAc2 glycan includes at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or100% (mol %) of the neutral N-glycans secreted by the host cell. The neutral N-glycans may be attached to a molecule selected from an amino acid, a peptide, and a polypeptide. In certain embodiments, the Man3GlcNAc2 glycan is attached to a heterologous polypeptide. In certain embodiments that may be combined with the preceding embodiments, the mannosyltransferase is a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase. In certain embodiments that may be combined with the preceding embodiments, the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *alg3* gene is deleted from the host cell. In certain embodiments that may be combined with the preceding embodiments, the host cell is a *Trichoderma* cell. In certain embodiments that may be combined with the preceding embodiments, the level of activity of alpha-1,6-mannosyltransferase in the host cell is reduced compared to the level of activity in a wild-type host cell. In certain embodiments that may be combined with the preceding embodiments, the host cell contains an endogenous polynucleotide encoding an α-1,2-mannosidase.

[0021] Another aspect includes a filamentous fungal cell having a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type filamentous fungal cell, where the filamentous fungal cell contains a recombinant protein of any of the preceding embodiments. In certain embodiments, the *alg3* gene contains a mutation. Preferably, the recombinant protein has N-acetylglucosaminyltransferase activity, where the recombinant protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and catalyzes the transfer of N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan, and where the recombinant protein is a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain. In certain embodiments, the mutation of the *alg3* gene is a deletion of the *alg3* gene. In certain embodiments that may be combined with the preceding embodiments, the fusion protein is encoded by a polynucleotide operably linked to a promoter. In certain embodiments, the promoter is an

inducible promoter. In certain embodiments, the inducible promoter is the *cbh1* promoter. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a polynucleotide encoding a UDP-GlcNAc transporter. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal has a reduced level of activity of an α-1,6-mannosyltransferase compared to the level of activity in a wild-type filamentous fungal cell. In certain embodiments, the filamentous fungal has a reduced level of expression of an och1 gene compared to the level of expression in a wild-type filamentous fungal cell. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a polynucleotide encoding an α-1,2-mannosidase. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a polynucleotide encoding a β-1,4-galactosyltransferase. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell further contains a polynucleotide encoding a sialyltransferase. In certain embodiments that may be combined with the preceding embodiments, the filamentous fungal cell is selected from *Trichoderma* sp., *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, and *Tolypocladium*.

DESCRIPTION OF THE FIGURES

[0022] Fig. 1 shows mass spectrometric neutral N-glycan profiles of average glycosylation on *T. reesei* strains M44, M81, M84, M109, M110, M131, M132, M133, M134, and M124.

[0023] Fig. 2 shows fragmentation analysis of monophosphorylated Man7Gn2. Only one example structure of monophosphorylated Man7Gn2 is shown.

[0024] Fig. 3 shows mass spectrometric acidic glycan profiles of *T. reesei* strains M44, M81, M84, M109, M110, M131, M132, M133, M134, and M124.

[0025] Fig. 4 shows neutral (a) and acidic (b) N-glycan profiles of *T. reesei* strain M44 cultured in a fermentor for 131.4 hours (fed batch).

[0026] Fig. 5 shows mass spectrometric neutral (a) and acidic (b) N-glycan profiles of *T. reesei* culture medium.

[0027] Fig. 6 shows a membrane blot of *T. reesei* M44 secreted proteins.

[0028] Fig. 7 shows an example of analyzed protein bands of *T. reesei* M44 cultivated in a fermentor. The glycosylation of proteins did not differ significantly from average glycosylation in *T. reesei*. The spectrum was focused to the minor base line signals, and the major signal of the spectrum was not quantitative in comparison to other signals.

[0029] Fig. 8 shows a Southern blot of DNA from the parental strain and from Alg3 knockout strains with an *alg3* probe.

[0030] Fig. 9A shows a restriction enzyme map of a section of the pTTv38 construct with sizes of predicted restriction products. Figure 9B shows a Southern blot of genomic DNA from the parental strain and the Alg3 knockout strains digested with EcoRI + PvuI (E + P) or KpnI + NheI (K + N). The control DNA was pTTv38 plasmid DNA digested with NotI. The blot was probed with an AmdS probe.

[0031] Fig. 10 shows MALDI analysis of neutral N-glycans. Part A shows the parental strain M124. Part B shows the Alg3 knockout 4A. Squares represent N-acetylglucosamine, and circles represent mannose, except for the one labeled glucose.

[0032] Fig. 11 shows fragmentation analysis of Man3Gn2 from the 4A Alg3 knockout strain.

[0033] Fig. 12 shows fragmentation analysis of Hex5Gn2 from Alg3 knockout strain 4A (part A) and parental strain M124 (part B). The signal marked with a box exists only as an isomer from the Alg3 knockout strain.

[0034] Fig. 13 shows neutral N-glycans from Alg3 knockout strain 4A after α-mannosidase digestion.

[0035] Fig. 14 shows the separation of two major glycans from the Alg3 knockout strain by liquid chromatography.

[0036] Fig. 15 shows proton NMR spectra of Hex3HexNAc2 (part A) and Hex6HexNAc2 (part B) fractions. Spectra were collected at 40°C using a Varian Unity INOVA 600 MHz spectrometer equipped with a cryoprobe.

[0037] Fig. 16 shows the acidic fraction of parental strain M124 (part A) and Alg3 knockout strain 4A (B). N-glycans with two phosphate units are marked with an asterisk.

[0038] Fig. 17 shows neutral N-glycans from supernatant of *T. reesei* Alg3 knockout strain 4A that was cultured in a flask for 5 days.

[0039] Fig. 18 shows neutral N-glycans from supernatant of *T. reesei* Alg3 knockout strain 4A that was cultured in a fermentor for 10 days.

[0040] Fig. 19 shows a MALDI spectrum of GnTI reaction mixture. GnTI has converted 54% of the acceptor to the product with one additional HexNAc.

[0041] Fig. 20 shows Western blot analysis of GnTII expression. Samples were run in 12% SDS-PAGE gel and blotted on nitrocellulose membrane. Histidine-tagged GnTII was detected on the membrane using mouse α-HIS monoclonal antibodies. Numbers shown on the left are the sizes of molecular weight marker proteins (kDa).

[0042] Fig. 21 shows a MALDI spectrum of GnTII reaction mixture. 83% of the acceptor (m/z 913.340) was converted to product (m/z 1136.433).

[0043] Fig. 22 shows GnTI activity observed for the GnTI/GnTII fusion protein.

[0044] Fig. 23 shows the N-glycans present in GnTI/GnTII *T. reesei* transformants obtained by targeting to the *alg3* locus.

[0045] Fig. 24 shows a MALDI spectrum of the purified reaction mixture from the enzyme activity test of the GnTII/GnTI fusion protein.

[0046] Fig. 25 shows a spectrum of the β1-2,3,4,6-N-acetylglucosaminidase reaction mixture.

[0047] Fig. 26 shows a MALDI spectrum of β1-4GalT reaction mixture.

[0048] Fig. 27 shows diagrams of observed N-glycans from supernatant proteins of *T. reesei* M127 pTTv110 transformants (gnt II/I in alg3 locus) on days 3 (A), 5 (B) and 7 (C

and D). The clone 17A produced the most G0 on day 7. (E) Mass spectrum of neutral N-glycans of supernatant proteins from *T. reesei* strain M127 GnT II/I transformant clone 17A cultivated for 7 days in shake flasks. Signals marked with asterisks originated from the culture medium.

[0049] Fig. 28 shows neutral N-glycans of rituximab from *T. reesei* M202 GnT II/I transformant clones (A) 9A-1 and (B) 31A-1, both cultivated with soybean trypsin inhibitor, and (C) mass spectrum of neutral N-glycans of rituximab purified from *T. reesei* strain M202 GnT II/I transformant clone 9A-1 cultivated for 5 days in shake flasks in the presence of soybean trypsin inhibitor.

[0050] Fig. 29 shows MALDI spectra of spacer modified GnTII/GnTI fusion reaction mixtures. Part (A) shows a reaction mixture of GnTII/GnTI with 3xG4S spacer modification. 36% of the acceptor has been converted to product with two additional HexNAcs. Part (B) shows a reaction mixture of GnTII/GnTI with 2xG4S spacer modification. 38% of the acceptor has been converted to product with two additional HexNAcs. Calculated m/z values for [M+Na]+ -signals of GnTI product, Hex3HexNAc2 (calc. m/z 933.318), was not detected in either spectra because all of the GnTI product was converted directly to Hex3HexNAc3, (calc. m/z 1136.318).

[0051] Fig. 30 shows Western blots of GnTII/I spacer variant cell pellets(A), and supernatants (B). Lanes 1.GnTII positive control, 2 GY3 mock strain, 3.GY7-2 wild type GnTII/I, 4.GY32-5 3xG4S spacer, 5. GY32-9 3xG4S spacer, 6.GY33-7 2xG4S spacer, 7.GY33-8 2xG4S spacer, 8.GY49-3 CBHI spacer and 9.GY50-10 EGIV spacer.

[0052] Fig. 31 shows GnT activities of wild-type GnTII/I and spacer variants from supernatants and expressed in the presence of protease inhibitors after day 3 (A) expression phases and day 4 (B) expression phases. The x-axis depicts sample identity (wt = wild-type, _1, _2 = parallel clones of the spacer variants), and the y-axis depicts percentage of products formed (GnTI and GnTII reaction products added together).

[0053] Fig. 32 shows GnT activities of GnTII/I fusion protein (with wild type spacer) in supernatant, cells and lysate. GnTI and GnTII products have been added together

[0054] Fig. 33 shows GnT activities of GnTII/I wild-type and spacer variants in (A) supernatants, (B) cells, and (C) lysates.

[0055] Fig. 34 shows example spectra of neutral N-glycans of parental strain M124 and GnT1 transformants on day 5. Signal with Gn addition (m/z 1460) is marked with an arrow. (pTTv11 with *cbh1* promoter, pTTv13 with *gpdA* promoter).

[0056] Fig. 35 shows the amounts of Man5 and Gn1Man5 in four positive GNT1 transformants on days 3 and 5. Quantitation was carried out against internal calibrant (Hex2HexNAc4, 2 pmol).

[0057] Fig. 36 shows example spectra of phosphorylated N-glycans of parental M124 strain and GnT1 transformants with internal calibrant (NeuAcHex4HexNAc2, 0.5 pmol.). GnT1 products are marked with an arrow.

[0058] Fig. 37 shows diagrams of neutral N-glycans of different GnTII strains/clones from day 5. Part (A) show the pTTv140 clone. Part (B) shows the pTTv142 clone. Part (C) shows the pTTv143 clone. Part (D) shows the pTTv141 clone.

[0059] Fig. 38 shows an example of neutral N-glycans of different GnTII strains/clones and the parental strain M198 from days 3, 5, and 7. Part (A) shows clone 1-117A. Part (B) shows clone
3-11A. Part (C) shows clone 30A. Part (D) shows parental stain M198.

[0060] Fig. 39 shows the membrane of separated proteins of *T. reesei* strain M198 and GnTII clone 3-17A. The 50 kDA protein is marked with an arrow.

[0061] Fig. 40 shows column diagrams of total secreted proteins versus individual secreted protein(s) of parental strain M198 (A) and the GnTII clone 3-17A (B).

[0062] Fig. 41 shows a column diagram of fermentor cultured GnTII strain M329 from day 3 to day 7, and shake flask culture of strain M329 from day 5.

[0063] Fig. 42 shows a multiple amino acid sequence alignment of *T. reesei* ALG3 and ALG3 homologs.

DETAILED DESCRIPTION

[0064] The present invention relates to recombinant proteins having N-acetylglucosaminyltransferase activity where the recombinant protein catalyzes the transfer of N-acetylglucosamine (GlcNAc) to a terminal Manα3 residue and catalyzes the

transfer of N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan, and where the recombinant protein contains catalytic domains from at least two different enzymes.

[0065] In some embodiments, the recombinant proteins of the invention include two catalytic domains, where one catalytic domain has N-acetylglucosaminyltransferase I (GnTI) activity (e.g., reacts with a terminal Manα3 residue), and the other catalytic domain has N-acetylglucosaminyltransferase II (GnTII) activity (e.g., reacts with a terminal Manα6 residue).

[0066] In some embodiments, the recombinant proteins of the present invention catalyze reactions that occur essentially sequentially. For example, the recombinant proteins of the present invention may catalyze the transfer of GlcNAc to a terminal Manα3-residue, first, and then catalyze the transfer of GlcNAc to a terminal Manα6-residueof an acceptor glycan. In one embodiment, the essentially sequential reactions are at least 10 fold, at least 20 fold, at least 30 fold, at least 40 fold, at least 50 fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, or at least 100 fold, more effective than the two reactions in the reversed order. In certain embodiments, a sequential reaction means that essentially or absolutely no GlcNAc can be transferred to the terminal Manα6-residue if GlcNAc has not yet been transferred to the terminal Manα3-residue. In a specific embodiment, the acceptor glycan contains a GlcNAcβ2Manα3-branch.

[0067] In some embodiments, the recombinant proteins react specifically with both Manα3 and Manα6 residues, optionally in branched acceptor glycans but not substantially or absolutely with other Manα-structures, e.g. Manα-monosaccharide conjugates, with Manαbenzyl and/or ManαSer/Thr-peptide. The non-substantial reactivity is preferably below 10%, below 8%, below 6%, below 4%, below 2%, below 1 %, or below 0.1 % of the Vmax with 0.1 mM acceptor glycan concentrations of reactions with terminal Manα3 and Manα6 residues. In a specific embodiment, the recombinant proteins have substantially similar reactivities with the terminal Manα3 (preferably as GnTI reaction) and the terminal Manα6 residue (preferably as GnTII reaction) of the acceptor glycan. Preferably neither catalytic activity has more than a 10-fold, 5-fold, 3-fold or 2-fold difference in reaction effectiveness compared to the other catalytic activity under the same conditions.

[0068] In a specific embodiment, the transfer of GlcNAc to the terminal Manα3 and Manα6 cause a conversion of at least 10 %, at least 25 %, at least 50 %, at least 70 %, at least 90%, or at least 95 % of Man3 glycan to a glycan with two terminal GlcNAcs. The effectiveness of the reaction can be measured by *in vitro* or *in vivo* assays as described in the examples disclosed herein. The effectiveness of the GlcNAc transfer reactions can be measured essentially as described in the Examples or as maximal reaction rate Vmax with 0.1 mM acceptor concentrations and saturating donor concentrations. In a specific embodiment, the effectiveness of the reaction is measured with a Man3 acceptor glycan attached to an amino acid, a peptide, or a polypeptide.

[0069] The present disclosure further relates to methods of producing a complex N-glycan, including the steps of providing a host cell, where the host cell contains a nucleic acid encoding a fusion protein containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, and culturing the host cell such that the fusion protein is expressed, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan.

[0070] The present invention also relates to a filamentous fungal cell having a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type filamentous fungal cell, where the filamentous fungal cell contains a recombinant protein of the invention.

Definitions

[0071] As used herein, "recombinant protein" refers to any protein that has been produced from a recombinant nucleic acid. "Recombinant nucleic acid" as used herein refers to a polymer of nucleic acids where at least one of the following is true: (a) the sequence of nucleic acids is foreign to (*i.e.*, not naturally found in) a given host cell; (b) the sequence may be naturally found in a given host cell, but is present in an unnatural (*e.g.*, greater than expected) amount or expressed at a level that is more or less than the natural level of expression; or (c) the sequence of nucleic acids includes two or more sub-sequences that are not found in the same relationship to each other in nature. For example, regarding instance (c), a recombinant nucleic acid sequence will have two or more sequences from unrelated genes arranged to make a new functional nucleic acid. In

another example, a recombinant nucleic acid sequence will contain a promoter sequence and a gene-encoding sequence that are not naturally found adjacent to one another.

[0072] As used herein, "N-acetylglucosaminyltransferase activity" refers to the activity of an enzyme that transfers an N-acetylglucosaminyl residue (GlcNAc) to an acceptor glycan. Typically, enzymes having this activity are N-acetylglucosaminyltransferases (GlcNAc transferases). In certain embodiments, GlcNAc transferases are eukaryotic. In certain embodiments, the GlcNAc transferases are mammalian enzymes forming a β-linkage from the 1-position of a GlcNAc-residue to the terminal mannose residues. In certain embodiments, the GlcNAc transferases are β2-N-acetylglucosaminyltransferases transferring β2-linked GlcNAc-residue(s) to the 2-position terminal mannose residues of glycans, in particular to an N-linked glycan. In certain embodiments, the β2-GlcNAc transferases are enzymes having GnTI activity and GnTII activity. GnTI activity transfers a GlcNAc residue to a Manα3 branch. The Manα3 branch may be a Manα3(R-Manα6)Manβ-branch of on N-linked glycan core structure, such as Man3GlcNAc2 or Man3 or Man5GlcNAc2 or Man5. GnTI enzymes may be mammalian enzymes, plant enzymes, or lower eukaryotic enzymes. GnTII activity transfers a GlcNAc residue to a Manα6-branch such as a Manα6(GlcNAcβ2Manα3)Manβ-branch of an N-linked glycan core structure. An example of such a Manα6-branch is GlcNAc1Man3GlcNAc2.

[0073] As used herein, "N-acetylglucosamine" refers to an N-acetylglucosamine residue (GlcNAc). GlcNAc may be part of a glycan structure. The amine group is on position 2, has a D-configuration, and has a pyranose structure as a residue. It may be alternatively named 2-acetamido-2-deoxy-D-glucopyranose (D-GlcpNAc). GlcNAc may also be a free reducing monosaccharide (i.e., not part of glycan).

[0074] As used herein, "Man" refers to a mannose residue. A "terminal Manα3" or a "terminal Manα6" refers to a mannose that is not substituted to the non-reducing end terminal residue by another monosaccharide residue or residues.

[0075] As used herein, "glycan" refers to an oligosaccharide chain that can be linked to a carrier such as an amino acid, peptide, polypeptide, lipid or a reducing end conjugate. In certain embodiments, the invention relates to N-linked glycans conjugated to a polypeptide N-glycosylation site such as -Asn-Xxx-Ser/Thr- by N-linkage to side-chain amide nitrogen of asparagine residue (Asn), where Xxx is any amino acid residue except

Pro. The invention may further relate to glycans as part of dolichol-phospho-oligosaccharide (Dol-P-P-OS) precursor lipid structures, which are precursors of N-linked glycans in the endoplasmic reticulum of eukaryotic cells. The precursor oligosaccharides are linked from their reducing end to two phosphate residues on the dolichol lipid. For example, α3-mannosyltransferase Alg3 modifies the Dol-P-P-oligosaccharide precursor of N-glycans. Generally, the glycan structures described herein are terminal glycan structures, where the non-reducing residues are not modified by other monosaccharide residue or residues.

[0076] As used herein, "glycoprotein" refers to a peptide or polypeptide attached to a glycan. The glycan may be attached to the peptide or polypeptide in a cotranslational or posttranslational modification.

[0077] As used herein, "glycolipid" refers to a lipid attached to a glycan and includes glyceroglycolipids, glycosphingolipids, and glycosylphosphatidylinositols.

[0078] As used throughout the present disclosure, glycolipid and carbohydrate nomenclature is essentially according to recommendations by the IUPAC-IUB Commission on Biochemical Nomenclature (e.g. Carbohydrate Res. 1998, 312, 167; Carbohydrate Res. 1997, 297, 1; Eur. J. Biochem. 1998, 257, 29). It is assumed that Gal (galactose), Glc (glucose), GlcNAc (N-acetylglucosamine), GalNAc (N-acetylgalactosamine), Man (mannose), and Neu5Ac are of the D-configuration, Fuc of the L-configuration, and all the monosaccharide units in the pyranose form (D-Galp, D-Glcp, D-GlcpNAc, D-GalpNAc, D-Manp, L-Fucp, D-Neup5Ac). The amine group is as defined for natural galactose and glucosamines on the 2-position of GalNAc or GlcNAc. Glycosidic linkages are shown partly in shorter and partly in longer nomenclature, the linkages of the sialic acid SA/Neu5X-residues α3 and α6 mean the same as α2-3 and α2-6, respectively, and for hexose monosaccharide residues α1-3, α1-6, β1-2, β1-3, β1-4, and β1-6 can be shortened as α3, α6, β2, β3, β4, and β6, respectively. Lactosamine refers to type II N-acetyllactosamine, Galβ4GlcNAc, and/or type I N-acetyllactosamine. Galβ3GlcNAc and sialic acid (SA) refer to N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc), or any other natural sialic acid including derivatives of Neu5X. Sialic acid is referred to as NeuNX or Neu5X, where preferably X is Ac or Gc. Occasionally Neu5Ac/Gc/X may be referred to as NeuNAc/NeuNGc/NeuNX.

Recombinant Proteins of the Invention

[0079] The invention herein relates to recombinant proteins having N-acetylglucosaminyltransferase activity, where the recombinant proteins catalyze the transfer of N-acetylglucosamine to a terminal Manα3 residue and catalyze the transfer of N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan. Recombinant proteins of the invention may include, without limitation, full length proteins having N-acetylglucosaminyltransferase activity, fragments of proteins having N-acetylglucosaminyltransferase activity, catalytic domains having N-acetylglucosaminyltransferase activity, and fusion proteins having N-acetylglucosaminyltransferase activity. A single recombinant protein of the invention has the capability to catalyze both transfers of N-acetylglucosamines. The transfer of N-acetylglucosamine to a terminal Manα3 residue may occur before or after the transfer of N-acetylglucosamine to a terminal Manα6 residue. Alternatively, the transfers may occur simultaneously.

[0080] The acceptor glycan may be attached to a molecule such as an amino acid, a peptide, or a polypeptide. In certain embodiments, the amino acid is an asparagine residue. The asparagine residue may be in aminoglycosidic linkage from the side-chain amide (a biologic mammalian polypeptide N-glycan linkage structure) and may be part of a peptide chain such as a dipeptide, an oligopeptide, or a polypeptide. The glycan may be a reducing end derivative such as an N-, O-, or C-linked, preferably glycosidic, derivative of the reducing GlcNAc or Man, such as a spacer or terminal organic residue with a certain glycan linked structure selected from the group of an amino acid, alkyl, heteroalkyl, acyl, alkyloxy, aryl, arylalkyl, or heteroarylalkyl. The spacer may be further linked to a polyvalent carrier or a solid phase. In certain embodiments, alkyl-containing structures include methyl, ethyl, propyl, and C4-C26 alkyls, lipids such as glycerolipids, phospholipids, dolichol-phospholipids and ceramides and derivatives. The reducing end may also be derivatized by reductive amination to a secondary amine linkage or a derivative structure. Certain carriers include biopoly- or oligomers such as (poly)peptides, poly(saccharides) such as dextran, cellulose, amylose, or glycosaminoglycans, and other organic polymers or oligomers such as plastics including polyethylene, polypropylene, polyamides (e.g., nylon or polystyrene), polyacrylamide, and polylactic acids, dendrimers such as PAMAM, Starburst or Starfish dendrimers, or polylysine, and polyalkylglycols

such as polyethylene glycol (PEG). Solid phases may include microtiter wells, silica particles, glass, metal (including steel, gold, and silver), polymer beads such as polystyrene or resin beads, polylactic acid beads, polysaccharide beads or organic spacers containing magnetic beads.

[0081] In certain embodiments, the acceptor glycan is attached to a heterologous polypeptide. As used herein, a “peptide” and a "polypeptide" are amino acid sequences including a plurality of consecutive polymerized amino acid residues. For purpose of this invention, typically, peptides are those molecules including up to 50 amino acid residues, and polypeptides include more than 50 amino acid residues. The peptide or polypeptide may include modified amino acid residues, naturally occurring amino acid residues not encoded by a codon, and non-naturally occurring amino acid residues. As used herein, “protein” may refer to a peptide or a polypeptide of any size. The term “heterologous polypeptide” refers to a polypeptide that is not naturally found in a given host cell or is not endogenous to a given host cell. In certain embodiments, the heterologous polypeptide is a therapeutic protein. Therapeutic proteins, for example, may include monoclonal antibodies, erythropoietins, interferons, growth hormones, enzymes, or blood-clotting factors. For example, the acceptor glycan may be attached to a therapeutic protein such as rituximab.

Acceptor Glycans

[0082] In certain embodiments, the structure of the acceptor glycan has the following formula, $[R_1]_y$Manα3($[R_2]_z$Manα6)Man$\{\beta$4GlcNAc(Fucαx$)_n[\beta$4GlcNAc$]_m\}_q$, where q, y, z, n and m are 0 or 1; x is linkage position 3 or 6, of optional fucose residue; R1 is GlcNAc, preferably GlcNAcβ2; and R2 is a branched structure Manα3(Manα6), with the provision that when z is 1, then y is 0, and when z is 0, then y is 0 or 1. ( ) defines a branch in the regular N-glycan core structure, either present or absent. [ ] and { } define a part of the glycan structure either present or absent in a linear sequence. When z is 0 and y is 0 then the structure is a Man3 glycan, and when z is 0 and y is 1, the structure is a GlcNAcMan3 glycan. When y is 0 and z is 1, the glycan is a Man5 glycan. The acceptor glycan may be beta-glycosidically linked to an Asn residue, preferably from the reducing end GlcNAc. In one embodiment, the acceptor glycan is a polypeptide linked N-glycan, where m and q are 1, and the acceptor structure contains a derivative of

$[R_1]_y$Manα3($[R_2]_z$Manα6)Manβ4GlcNAc(Fucαx)$_n$β4GlcNAc. Optional derivatives include substitutions by monosaccharide residues such as GlcNAc or xylose.

[0083] The acceptor glycan may be Man3, GlcNAcMan3, or Man5. In certain embodiments, the acceptor glycan is Man3 or GlcNAcMan3. Man3 is a trimannosyl glycan comprising at least one of Manα3 or Manα6 residues and is preferably a branched oligosaccharide, such as Manα3(Manα6)Man. Other certain Man3 oligosaccharides are Manα3(Manα6)Manβ, Manα3(Manα6)Manβ4GlcNAc, and polypeptide-linked Manα3(Manα6)Manβ4GlcNAcβ4GlcNAc. In addition, depending on the host cell, the glycan can contain Fuc, Xyl or GlcNAc in Manβ and/or GlcNAc residues, such as Manα3(Manα6)Manβ4GlcNAcβ4(Fucαx)$_n$GlcNAc, where x is 3 or 6 and n is 0 or 1, also described by a monosaccharide composition formula indicating the terminal mannose structure and reducing end composition as Man3GlcNAc2 (n is 0) and Man3GlcNAc2Fuc (n is 1). In certain embodiments, especially those with a polypeptide-linked structure, the Man3 structure is a Manα3(Manα6)Manβ4GlcNAcβ4(Fucα6)$_n$GlcNAc. In certain embodiments, the polypeptide-linked GlcNAcMan3 structure is GlcNAcβ2Manα3(Manα6)Manβ4GlcNAcβ4(Fucα6)$_n$GlcNAc, also described by a monosaccharide composition formula GlcNAcMan3GlcNAc2 (n is 0) and GlcNAcMan3GlcNAc2Fuc (n is 1). . In certain embodiments, the polypeptide-linked Man5 structure is Manα3{Manα3(Manα6)Manα6}Manβ4GlcNAcβ4(Fucα6)$_n$GlcNAc, where { } and ( ) indicate a branch and n is 0 or 1, also described by a monosaccharide composition formula Man5GlcNAc2 (n is 0) and Man5GlcNAc2Fuc (n is 1).

[0084] Accordingly, the certain Man3 glycans have structures according to the following formula, Manα3(Manα6)Manβ4GlcNAc(Fucαx)$_n$β4GlcNAc, where n is 0 or 1, indicating presence or absence of part of the molecule, where x is 3 or 6, and where ( ) defines a branch in the structure. In embodiments of the invention where the acceptor glycan is Man3, the recombinant protein catalyzes the transfer of N-acetylglucosamine to the terminal Manα3 and Manα6 of Man3, thus resulting in GlcNAc2Man3, GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4(Fucαx)$_n$GlcNAc, where n is 0 or 1, also described by a monosaccharide composition formula GlcNAc2Man3GlcNAc2 (n is 0) and GlcNAc2Man3GlcNAc2Fuc (n is 1).

[0085] In embodiments of the invention where the acceptor glycan is Man5, the recombinant protein catalyzes the transfer of N-acetylglucosamine to the terminal Manα3 of Man5. After 2 mannoses have been removed from GlcNAcMan5 (for example, by mannosidase II) to form GlcNAcMan3, the recombinant protein catalyzes the transfer of N-acetylglucosamine to the terminal Manα6, thus resulting in GlcNAc2Man3 (which has the structure GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4(Fucαx)$_n$GlcNAc, where n is 0 or 1, also referred to as G0 if attached to an antibody).

Fusion Proteins Containing N-acetylglucosaminyltransferase Catalytic Domains

[0086] In certain embodiments, the recombinant proteins of the invention are fusion proteins containing an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain. The term "fusion protein" refers to any protein or polypeptide containing a protein or polypeptide linked to heterologous amino acids.

[0087] N-acetylglucosaminyltransferase I (GlcNAc-TI; GnTI; EC 2.4.1.101) catalyzes the reaction UDP-N-acetyl-D-glucosamine + 3-(alpha-D-mannosyl)-beta-D-mannosyl-R <=> UDP + 3-(2-(N-acetyl-beta-D-glucosaminyl)-alpha-D-mannosyl)-beta-D-mannosyl-R, where R represents the remainder of the N-linked oligosaccharide in the glycan acceptor. An N-acetylglucosaminyltransferase I catalytic domain is any portion of an N-acetylglucosaminyltransferase I enzyme that is capable of catalyzing this reaction. Amino acid sequences for N-acetylglucosaminyltransferase I enzymes from various organisms are listed in SEQ ID NOs: 1-19. Additional GnTI enzymes are listed in the CAZy database in the glycosyltransferase family 13 (cazy.org/GT13_all). Enzymatically characterized species includes *A. thaliana* AAR78757.1 (US6 653 459), *C. elegans* AAD03023.1 (Chen S. et al J. Biol.Chem 1999;274(1):288-97), *D. melanogaster* AAF57454.1 (Sarkar & Schachter Biol Chem. 2001 Feb;382(2):209-17); *C. griseus* AAC52872.1 (Puthalakath H. et al J. Biol.Chem 1996 271(44):27818-22); *H. sapiens* AAA52563.1 (Kumar R. et al Proc Natl Acad Sci U S A. 1990 Dec;87(24):9948-52); *M. auratus* AAD04130.1 (Opat As et al Biochem J. 1998 Dec 15;336 (Pt 3):593-8), (including an example of deactivating mutant), Rabbit, *O. cuniculus* AAA31493.1 (Sarkar M et al. Proc Natl Acad Sci U S A. 1991 Jan 1;88(1):234-8). Additional examples of characterized active enzymes can be found at cazy.org/GT13_characterized. The 3D structure of the catalytic domain of rabbit GnTI was defined by X-ray crystallography in

Unligil UM et al. EMBO J. 2000 Oct 16;19(20):5269-80. The Protein Data Bank (PDB) structures for GnTI are 1FO8, 1FO9, 1FOA, 2AM3 , 2AM4, 2AM5, and 2APC. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain is from the human N-acetylglucosaminyltransferase I enzyme (SEQ ID NO: 1), or variants thereof. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain contains a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues 84-445 of SEQ ID NO: 1. In some embodiments, a shorter sequence can be used as a catalytic domain (e.g. amino acid residues 105-445 of the human enzyme or amino acid residues 107-447 of the rabbit enzyme; Sarkar et al. (1998) Glycoconjugate J 15:193-197). Additional sequences that can be used as the GnTI catalytic domain include amino acid residues from about amino acid 30 to 445 of the human enzyme or any C-terminal stem domain starting between amino acid residue 30 to 105 and continuing to about amino acid 445 of the human enzyme, or corresponding homologous sequence of another GnTI or a catalytically active variant or mutant thereof. The catalytic domain may include N-terminal parts of the enzyme such as all or part of the stem domain, the transmembrane domain, or the cytoplasmic domain.

[0088] As used herein, "cytoplasmic" is used to refer to a part of a protein that interacts with the cytoplasm of a cell.

[0089] N-acetylglucosaminyltransferase II (GlcNAc-TII; GnTII; EC 2.4.1.143) catalyzes the reaction UDP-N-acetyl-D-glucosamine + 6-(alpha-D-mannosyl)-beta-D-mannosyl-R <=> UDP + 6-(2-(N-acetyl-beta-D-glucosaminyl)-alpha-D-mannosyl)-beta-D-mannosyl-R, where R represents the remainder of the N-linked oligosaccharide in the glycan acceptor. An N-acetylglucosaminyltransferase II catalytic domain is any portion of an N-acetylglucosaminyltransferase II enzyme that is capable of catalyzing this reaction. Amino acid sequences for N-acetylglucosaminyltransferase II enzymes from various organisms are listed in SEQ ID NOs: 20-33. In certain embodiments, the N-acetylglucosaminyltransferase II catalytic domain is from the human N-acetylglucosaminyltransferase II enzyme (SEQ ID NO: 20). Additional GnTII species are listed in the CAZy database in the glycosyltransferase family 16 (cazy.org/GT16_all). Enzymatically characterized species include GnTII of *C. elegans*, *D. melanogaster, Homo sapiens, Rattus norvegigus, Sus scrofa* (cazy.org/GT16_characterized). In certain

embodiments, the N-acetylglucosaminyltransferase II catalytic domain contains a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues from about 30 to about 447 of SEQ ID NO: 21. The catalytic domain may include N-terminal parts of the enzyme such as all or part of the stem domain, the transmembrane domain, or the cytoplasmic domain.

[0090] In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain is N-terminal to the N-acetylglucosaminyltransferase II catalytic domain. In other embodiments, the N-acetylglucosaminyltransferase II catalytic domain is N-terminal to the N-acetylglucosaminyltransferase I catalytic domain. The term "N-terminal" refers to the positioning of a set of amino acid residues closer to the end of a polypeptide that is terminated by an amino acid with a free amine group ($-NH_2$) compared to a reference set of amino acid residues.

Spacers

[0091] In certain embodiments of the invention, the recombinant protein contains a spacer in between the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain. The term "spacer" refers to any number of consecutive amino acids of any sequence separating the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain such that the spacer has no effect on the enzymatic function of the catalytic domains. Typically, the spacer is at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, or at least 50 amino acids in length. In certain embodiments, the spacer contains sequence from a stem domain. "Stem domain" refers to a protein domain, or a fragment thereof, which is located adjacent to the transmembrane domain of a native enzyme, such as a glycosyltransferase or a glycosyl hydrolase, and optionally targets the enzyme to or assists in retention of the enzyme in the ER/Golgi. Stem domains generally start with the first amino acid following the hydrophobic transmembrane domain and end at the catalytic domain. Exemplary stem domains include, but are not limited to, the stem domain of human GnTI, amino acid residues from about 30 to about 83 or from about 30 to about 105 for the human GnTII, or amino acid residues from about 26 to about 106 or from about 26 to about 83 for the *T. reesei* KRE2. In certain embodiments where the spacer contains sequence from a stem

domain, the spacer includes amino acids 30-83 of the human GnTI sequence (SEQ ID NO: 34). In other embodiments, the spacer may include any of the sequences listed in SEQ ID NOs: 35-38.

[0092] Further examples of suitable spacers include, without limitation, the flexible spacer 3XG4S (SEQ ID NO: 118), the flexible spacer 2XG4S (SEQ ID NO: 120), the spacer for the *T. reesei* CBHI (SEQ ID NO: 122); and the spacer for the *T. reesei* EGIV cellulase (SEQ ID NO: 124).

[0093] In certain embodiments, the length of the spacer is about the same as the length of a stem domain of GnT1. In certain embodiments, the length is about 74 amino acid residues, plus or minus about 37 amino acids. For example, the spacer length is about 30 amino acids to about 110 amino acids, or from about 35 amino acids to about 100 amino acids, or as exemplified in the examples described herein, plus or minus 2, 3, 4, or 5 amino acids. In one embodiment, the spacer length corresponds to a truncated stem domain of GnT1, for example, start from amino acid 25 to amino acid 104, or between amino acid 30 to amino acid 101, to the end of the GnT1 stem domain. In certain embodiments, the spacer may include a part of the stem domain of human GnT1, which may start from an amino acid positioned between amino acid 70 to amino acid 87 (according to numbering in SEQ ID NO: 34), or between amino acid 76 and amino acid 104, or beginning from amino acid 30, 35, 40, 45, 50, 60, 70, 73, 74, 75, 76, 80, 83, 84, 85, 86, 87, 100, 101, 102, 103, or 104, to the end of the human GnT1 stem domain. In other embodiments, the spacer may include a heterologous spacer peptide, which may include a fungal spacer peptide and/or a repetitive oligomer spacer peptide.

[0094] Typically, the spacer is an elongated peptide without specific conformation and contains amino acid residues allowing high flexibility (*e.g.*, Gly and Ala), hydroplicity (e.g., Ser and Thr), and optionally Pro to prevent conformation. The spacer may be glycosylated. In certain embodiments the spacer is O-glycosylated including fungal O-mannosylation. In certain embodiments the spacer is an endogenous fungal, filamentous fungal, or *Trichoderma* spacer peptide, such as a spacer that naturally separates protein domains. The spacer may be derived from a secreted or cellulolytic enzyme of a fungus such as a filamentous fungus (e.g., *T. reesei*), a fragment thereof, or a multimer of the spacer and/or its fragment or mutated analog or equivalent thereof. The natural fungal spacer may contain dimeric or oligomeric proline and/or glycine and/or serine and/or

threonine, and/or multiple amino acid residues selected from Ser, Thr, Gly, Pro or Ala or any combinations thereof. In certain embodiments, the spacer is a repeating oligomer containing a monomer with 1-10 or 1-5 amino acid residues selected from Ser, Thr, Gly, Pro or Ala and optionally a charged amino acid residue selected from negatively charged residues Glu or Asp or positively charged residues Lys or Arg. In certain embodiments the charged residue is negatively charged. In certain embodiments the monomer contains dimeric or oligomeric amino acid residues, and/or multiple single amino acid residues selected from Ser, Thr, Gly, Pro and Ala. In certain embodiments the oligomer contains a monomer of a dimer or oligomer of glycine and a single residue selected from the Ser, Thr, Gly, Pro and Ala. In certain embodiments the single residue is Ser or Thr. In certain embodiments the residue is Ser. In certain embodiments, the sequence of the repeating spacer is $\{(Yyy)_nXxx\}_m$ where n is 2 to 10, m is 2 to 10, and Xxx and Yyy are selected from Ser, Thr, Gly, Pro and Ala, with the proviso that Xxx and Yyy are not the same amino acid residue. In certain embodiments the repeating spacer is $\{(Gly)_nXxx\}_m$ where n is 2 to 10, m is 2 to 10, and Xxx is selected from Ser, Thr, Gly, Pro and Ala. In certain embodiments Xxx is Ser or Thr. In certain embodiments Xxx is Ser.

Targeting Peptides

[0095] In certain embodiments, recombinant proteins of the invention include a targeting peptide linked to the catalytic domains. The term "linked" as used herein means that two polymers of amino acid residues in the case of a polypeptide or two polymers of nucleotides in the case of a polynucleotide are either coupled directly adjacent to each other or are within the same polypeptide or polynucleotide but are separated by intervening amino acid residues or nucleotides. A "targeting peptide", as used herein, refers to any number of consecutive amino acid residues of the recombinant protein that are capable of localizing the recombinant protein to the endoplasmic reticulum (ER) or Golgi apparatus (Golgi) within the host cell. The targeting peptide may be N-terminal or C-terminal to the catalytic domains. In certain embodiments, the targeting peptide is N-terminal to the catalytic domains. In certain embodiments, the targeting peptide provides binding to an ER or Golgi component, such as to a mannosidase II enzyme. In other embodiments, the targeting peptide provides direct binding to the ER or Golgi membrane.

[0096] Components of the targeting peptide may come from any enzyme that normally resides in the ER or Golgi apparatus. Such enzymes include mannosidases,

mannosyltransferases, glycosyltransferases, Type 2 Golgi proteins, and MNN2, MNN4, MNN6, MNN9, MNN10, MNS1, KRE2, VAN1, and OCH1 enzymes. Such enzymes may come from a yeast or fungal species such as those of *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filobasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, and *Trichoderma*. Sequences for such enzymes can be found in the GenBank sequence database.

[0097] In certain embodiments the targeting peptide comes from the same enzyme and organism as one of the catalytic domains of the recombinant protein. For example, if the recombinant protein includes a human GnTII catalytic domain, the targeting peptide of the recombinant protein is from the human GnTII enzyme. In other embodiments, the targeting peptide may come from a different enzyme and/or organism as the catalytic domains of the recombinant protein.

[0098] Examples of various targeting peptides for use in targeting proteins to the ER or Golgi that may be used for targeting recombinant proteins of the invention include: Kre2/Mnt1 N-terminal peptide fused to galactosyltransferase (Schwientek, JBC 1996, 3398), HDEL for localization of mannosidase to ER of yeast cells to produce Man5 (Chiba, JBC 1998, 26298-304; Callewaert, FEBS Lett 2001, 173-178), OCH1 targeting peptide fused to GnTI catalytic domain (Yoshida et al, Glycobiology 1999, 53-8), yeast N-terminal peptide of Mns1 fused to α2-mannosidase (Martinet et al, Biotech Lett 1998, 1171), N-terminal portion of Kre2 linked to catalytic domain of GnTI or β4GalT (Vervecken, Appl. Environ Microb 2004, 2639-46), various approaches reviewed in Wildt and Gerngross (Nature Rev Biotech 2005, 119), full-length GnTI in *Aspergillus nidulans* (Kalsner et al, Glycocon. J 1995, 360-370), full-length GnTI in *Aspergillus oryzae* (Kasajima et al, Biosci Biotech Biochem 2006, 2662-8), portion of yeast Sec12 localization structure fused to *C. elegans* GnTI in *Aspergillus* (Kainz et al 2008), N-terminal portion of yeast Mnn9 fused to human GnTI in *Aspergillus* (Kainz et al 2008), N-terminal portion of *Aspergillus* Mnn10 fused to human GnTI (Kainz et al, Appl. Environ Microb 2008, 1076-86), and full-length human GnTI in *T. reesei* (Maras et al, FEBS Lett 1999, 365-70).

[0099] In certain embodiments the targeting peptide is the Kre2/Mnt1(i.e., Kre2) targeting peptide having the amino acid sequence of SEQ ID NO: 115 or SEQ ID NO: 116.

[00100] Further examples of sequences that may be used for targeting peptides include the sequences listed in Table 1 below.

Table 1: Targeting peptides. Putative transmembrane domains are underlined. In KRE2, the stem domain enabling Golgi localization is underlined and double-underlined. Other1 and Other02 are putative mannosylation-related proteins.

Homologous to	Cytoplasmic	Transmembrane	Luminal
KRE2 estExt_fgenesh1_pm.C_30039	MASTNARYVR SEQ ID NO: 39	YLLIAFFTILVFYFVSN SEQ ID NO: 40	SKYEGVDLNKGTFTAPDSTKTTPKPPATGDAKDFPLALTPNDP GFNDLVGIAPGPRMNATFVTLARNSDVWDIARSIRQVEDRFNRRYNY DWVFLNDKPFDNTFKKVTTSLVSGKTHYGEIAPEHWSFPDWIDQDKA KKVREDMAERKIIYGDSVSYRHMCRFESGFFFRQPLMMNYEYYWRV EPSIELYCDIHYDPFRLMVEQGKKYSFVISLYEYPATIATLWESTKKFM KNHPEHIAPDNSMRFLSDDGGETYNNCHFWSNFEIGSLEWLRSKQYI DFFESLDKDGGFFYERWGDAPVHSIAAGLMLNRSEIHFFNDIAYWHV PFTHCPTGEKTRLDLKCHCDPKENFDWKGYSCTSRFFEMNGMDKPE GWENQQD SEQ ID NO: 41
KRE2 alternative1 e_gw1.28.231.1	MAIARPVR SEQ ID NO: 42	ALGGLAAILWCFFLY SEQ ID NO: 43	QLLRPSSSYNSPGDRYINFERDPNLDPTGEPEGILVRTSDRYAPDAK DTDRASATLLALVRNEEVDDMVASMVDLERTWNSKFNYPWTFFNDK PFSEEFKKKTSAVTNATCNYELIPKEHWDAPSWIDPAIFEESAAVLKK NGVQYANMMSYHQMCRWNSGMFYKHPALKDVRYYWRVEPKVHFF CDVDYDVFRYMQDNNKTYGFTINLYDDPHTLPTLWPQTAKFLADHPN YLHEHSAIKWVIDDARRPQHNREAQGFSTCHFWSNFEVADMEFWRS KVYEDYFEHLDRAGGFFYERWGDAPVHSIALGLFEDSSKIHWFRDIG YQHIPFFNCPNSPKCKGCVTGRLTDGEPFLHREDCRPNWFKYAGMG SEQ ID NO: 44
OCH1 e_gw1.16.371.1	MLNPRR SEQ ID NO: 45	ALIAAAFILTVFFLI SEQ ID NO: 46	SRSHNSESASTSEPKDAEAEALSAANAQQRAAPPPPPQKPMIDMSG MSTYDKLAYAYEYDIESKFPAYIWQTWRKTPSEGDFEFREQEASWSI EHPGFIHEVITDSVADTLLQLLYGSIPEVLEAYHALPLPVLKADLFRYLIL YARGGIYSDIDTYAIRSALEWIPPQIPKETVGLVIGIEADPDRPDWADW YSRRIQFCQWTIQSKPGHPVLRDIISRITNQTLEMKKSGKLSAFQGNR VVDLTGPAVWTDTIMDYFNDERYFDMENSKGRIDYRNFTGMETSKRV GDVVVLPITSFSPGVGQMGAKDYDDPMAFVKHDFEGTWKPESERHI GEIVQELGEGQGEAPKEQ SEQ ID NO: 47
OCH1 alternative1 fgenesh1_pm.C_scaffold_13000080	MGMGQCQWSPFRNKVPTQMRRC SEQ ID NO: 48	LPLYITVVCVFLVIV SEQ ID NO: 49	NFDWILAIPNPASVLRREPKAPPLPGSTFPQKIWQTWKVDPLNFDERD LVTARTWTTINPGMRYEVVTDANEMAYIEDRYGPNGFDRPDIVEFYK MINLPIIKADLLRYMIMYAEGGIYADIDVETMKPFHRFIPDRYDEKDIDIII GVEIDQPDFKDHPILGKKSMSFCQWTFVARPQQPVMMRLIENIMKWF KTVARDQGVPLGEVQLDFDQVISGTGPSAFTKAMLEEMNRKTKGPKV TWDAFHNLDESKLVGGVLVLTVEAFCAGQGHSDSGNHNARNALVKH HFHASNWPSRHPRYKHPAYGQVEDCNWVPECVRKWDEDTSNWDK YSENEQKKILQDIENARLERERQQQALAALP

Homologous to	Cytoplasmic	Transmembrane	Luminal
			SEQ ID NO: 50
MNN9 e_gw1.5.262.1	MARPMGSVRLKK ANPST SEQ ID NO: 51	LILGAVLCIFIIIFLV SEQ ID NO: 52	SPSSPASASRLSIVSAQHHLSPPTSPYQSPRSGAVQGPPPVTRYNLN KVTVTSDPVRNQEHILILTPMARFYQEYWDNLLRLNYPHELITLGFILP KTKEGNQATSMLQKQIQKTQNYGPEKDRFKSIIILRQDFDPAVVSQDE SERHKLANQKARREVMAKARNSLLFTTLGPSTSWVLWLDADITETAP TLIQDLASHDKPIIVANCFQKYYDPESKKMAERPYDFNSWQDSETALK MAEQMGPDDILLEGYAEMATYRTLLAYMSTPGGSKDLVVPLDGVGG TALLVKADVHRDGAMFPPFAFYHLIESEGFAKMAKRLGWQPYGLPNY KVYHYNE SEQ ID NO: 53
MNN9 alternative1 estExt_GeneWise Plus.C_230146	MLLPKGGLDWRS ARAQIPPTRAL WNAVTRTR SEQ ID NO: 54	FILLVGITGLILLLW SEQ ID NO: 55	RGVSTSASEMQSFYCWGPAKPPMEMSPNEHNRWNGHLQTPVIFNH HAPVEVNSSTIEHVDLNPINSTKQAVTKEERILILTPLKDAAPYLSKYFE LLAELTYPHRLIDLAFLVSDSTDDTLAVLASELDRIQKRPDQIPFHSATV IEKDFGFKLSQNVEERHSFEAQGPRRKAMGRARNYLLYTALKPEHSW VYWRDVDIVDSPTGILEDFIAHDRDILVPNIWFHRYRDGVDIEGRFDYN SWVESDKGRKLANSLDKDVVLAEGYKQYDTGRTYMAKMGDWRENK DVELELDGIGGVNILVKADVHRSGINFPCYAFENQAETEGFAKMAKRA GYEVYGLPNYVVWHIDTEEKGGNA SEQ ID NO: 56
MNN9 alternative2 estExt_GeneWise Plus.C_400029	MMPRHHSSGFSN GYPRADTFEI SPHRFQPRATLPP HRKRKRTAIR SEQ ID NO: 57	VGIAVVVILVLVL WFG SEQ ID NO: 58	QPRSVASLISLGILSGYDDLKLETVRYYDLSNVQGTARGWEREERILL CVPLRDAEQHLPMFFSHLKNFTYPHNLIDLAFLVSDSKDHTLESLTEH LEAIQADPDPKQPYGEISIIEKDFGQKVNQDVESRHGFAAQASRRKLM AQARNWLLSAALRPYHSWVYWRDVDVETAPFTILEDLMRHNKDVIVP NVWRPLPDWLGGEQPYDLNSWQESETALALADTLDEDAVIVEGYAE YATWRPHLAYLRDPYGDPDMEMEIDGVGGVSILAKAKVFRAGVHFPA FSFEKHAETEGFGKMAKRMHFSVVGLPHYTIWHLYEPSVDDIKHMEE MERERIAREKEEEERKKKEAQIKEEFGDANSQWEQDKQQMQDLKLQ DRGGDKEAAAAGVNQGAAAKAAGAMEGQKN SEQ ID NO: 59
MNN10 fgenesh5_pg.C_scaffold_5000342	MSLSRSPSPVPG GGWSSPGLNINS GRSSPSNAAGSS VSWESAKMRKQG ANGYPSFSTQNQ GFFTRHMRRI SSSLPRFAAGPG NTYAEREKYERG GHSPHAGGGRLR AFLARIGRRLKWR SEQ ID NO: 60	ILLPLIIICTIVAYY SEQ ID NO: 61	GTHEAPGFVHWWRRISMGGGGEKFVIILGANVGGGVMEWKGAREW AIERDSVRNKRKYATRWGYDLEIVDMKTKKRYAHEWRESWEKVDFIR AAMRKYPKAEWFWWLDLNTYVMEPSYSLQRHLFNHLDRHVYRDINV FNPLNITHPPTEEYLDAEARSPVGDGNINSVNLMLTQDCSGFNLGSFF IRRSAWTEQLLDIWWDPVLYEQKHMEWEHKEQDALEQLYRTQPWIR QHTGFLPQRLINSFPPAACADESGLNNTRIHYNEKDRDFVVNMAGCE WGRDCWGEMYHYREFSYWLNRNPWELFKEEIVAVIWYKLTGQRVKL SEQ ID NO: 62
MNN10 alternative1 estExt_GeneWise Plus.C_150339	MHFAYPSRKSSN PPPFRPRSTRLPG LRRSRIKT SEQ ID NO: 63	IGIVLFLVLATLWF F SEQ ID NO: 64	SNPRVPRPDPERVPSGRPPVVLVTVIDPTQYPNAYLKTIKENREQYAA KHGYEAFIVKAYDYDTQGAPQSWSKLMAMRHALTKFPECRFVWYLD QDAYIMDMSKSLEEQLLNRQKLESLMIKNYPVVPPDSIIKTFSHLRPDE VDLIVSQDSSGLVAGSVVVRNSQWSKFLLETWMDPLYRSYNFQKAE RHALEHIVQWHPTILSKLALVPQRTLGPYTRTDQGDAYQDGDFVVMF TGCTKSGEQSCETVSASYYQKWSSSL SEQ ID NO: 65
MNS1 fgenesh1_pm.C_s	MIRDPFGIHSKNA FKATALRAARDIK	VLGMIAAAVMFVL YVTGFF	SSGQTEEAKKKASGSAFSWLGLSQERGGVDWDERRKSVVEAFEVW DAYERYAWGKDEFHPISKNGRNMAPKGLGWIIIDSLDTMMLMNQTTR

Homologous to	Cytoplasmic	Transmembrane	Luminal
caffold_3000175	EAATQAGANALE MSFSLPKHVPDF GDPSRALEDRAW AALLPMYKDKPYA YAPSMRLRPWWR RRK SEQ ID NO: 66	SEQ ID NO: 67	LQHAREWISTSLTWDQDQDVNTFETTIRMLGGLLSAHYLSTEFPELAP LTEDDEGAPGEDLYLEKAKDLADRLLSAFESESGIPYASVNIGEYKGP SHSDNGASSTAEATTLQLEFKYLAKLTGEKNFWDKVEKVMEVVDDN QPEDGLVPIYIYATTGEFRGQNIRLGSRGDSYYEYLIKQYLQTNKQEPI YEEMWDEALAGVRKHLVTYTEPSEFTIIAERPDGLEHPMSPKMDHLV CFMPGTIALAATGGLTEAEARKLSTWNKKKDDDMQLARELMHTCWG MYKYMKTGLAPEIMYFNIPNPPPESSAPHQAPAAFDEDPHAEWRKDF VVHSNDVHNLQRPETVESLFYMWRITGDVKYREWGWDMFKSFVNYT AVEDQGGFTSLLDANSIPPTPKDNMESFWLAETLKYMYLLFSPNDVLP LHKIVLNTEAHPFPRFDMGPLFSTGWKRKPRDGSAKKKATTAATTDA E SEQ ID NO: 68
MNS1 alternative1 estExt_fgenesh1_ pm.C_80182	MARRRYR SEQ ID NO: 69	LFMICAAVILFLLY R SEQ ID NO: 70	VSQNTWDDSAHYATLRHPPASNPPAAGGESPLKPAAKPEHEHEHEN GYAPESKPKPQSEPKPESKPAPEHAAGGQKSQGKPSYEDDEETGKN PPKSAVIPSDTRLPPDNKVHWRPVKEHFPVPSESVISLPTGKPLKVPR VQHEFGVESPEAKSRRVARQERVGKEIERAWSGYKKFAWMHDELSP VSAKHRDPFCGWAATLVDSLDTLWIAGLKEQFDEAARAVEQIDFTTTP RNNIPVFETTIRYLGGLLGAFDVSGGHDGGYPMLLTKAVELAEILMGIF DTPNRMPILYYQWQPEYASQPHRAGSVGIAELGTLSMEFTRLAQLTS QYKYYDAVDRITDALIELQKQGTSIPGLFPENLDASGCNHTATALRSSL SEAAQKQMDEDLSNKPENYRPGKNSKADPQTVEKQPAKKQNEPVEK AKQVPTQQTAKRGKPPFGANGFTANWDCVPQGLVVGGYGFQQY HMGGGQDSAYEYFPKEYLLLGGLESKYQKLYVDAVEAINEWLLYRPM TDGDWDILFPAKVSTAGNPSQDLVATFEVTHLTCFIGGMYGLGGKIFG REKDLETAKRLTDGCVWAYQSTVSGIMPEGSQVLACPTLEKCDFN ETLWWEKLDPAKDWRDKQVADDKDKATVGEALKETANSHDAAGGS KAVHKRAAVPLPKPGADDDVGSELPQSLKDKIGFKNGEQKKPTGSSV GIQRDPDAPVDSVLEAHRLPPQEPEEQQVILPDKPQTHEEFVKQRIAE MGFAPGVVHIQSRQYILRPEAIESVWYMYRITGDPIWMEKGWKMFEA TIRATRTEIANSAIDDVNSEEPGLKDEMESFWLAETLKYYYLLFSEPSVI SLDEWVLNTEAHPFKRPGGSVIGHSI SEQ ID NO: 71
MNS1 alternative2 estExt_GeneWise Plus.C_120298	MLNQLQGRVPRR Y SEQ ID NO: 72	IALVAFAFFVAFLL W SEQ ID NO: 73	SGYDFVPRTATVGRFKYVPSSYDWSKAKVYYPVKDMKTLPQGTPVT FPRLQLRNQSEAQDDTTKARKQAVKDAFVKSWEAYKTYAWTKDQLQ PLSLSGKETFSGWSAQLVDALDTLWIMDLKDDFFLAVKEVAVIDWSKT KDNKVINLFEVTIRYLGGLIAAYDLSQEPVLRAKAIELGDTLYATFDTPN RLPSHWLDYSKAKKGTQRADDSMSGAAGGTLCMEFTRLSQITGDPK YYDATERIKQFFYRFQNETTLPGMWPVMMNYREETMVESRYSMGGS ADSLYEYLVKMPALLGGLDPQYPEMAIRALDTARDNLLFRPMTEKGD NILALGNALVDHGNVQRITEMQHLTCFAGGMYAMAGKLFKRDDYVDL GSRISSGCVWAYDSFPSGIMPESADMAACAKLDGPCPYDEVKAPVD PDGRRPHGFIHVKSRHYLLRPEAIESVFYMWRITGDQVWRDTAWRM WENIVREAETEHAFAIVEDVTRTASKLTNNYLLQTFWLAETLKYFYLIF DDESAIDLDKWVFNTEAHPFKRPAV SEQ ID NO: 74
MNS1 alternative3 estExt_GeneWise Plus.C_160228	MLVVGRPRLVRN S SEQ ID NO: 75	IILTLAILSIWHLGL L SEQ ID NO: 76	SRTPTSASALVSASVSASSEWSRLERLMNRGAPLTPYPDSNSSFDW SAIPFRYPPHNTTHLPPRHKQPPLPRIQHRFGPESPAAAKERIKRLKA VKQVFLRAWQAYKGYAWKQDALLPISGGGREQFSGWAATLVDALDT LWIMGLREEFDEAVAAVAEIDFGSSTSSRVNIFETNIRYLGGLLAAYDL

Homologous to	Cytoplasmic	Transmembrane	Luminal
			SGREVLLKKAVELGDLIYAGFNTENGMPVDFLNFYSAKSGEGLVVES SVVSASPGTLSLELAHLSQVTGDDKYYSAVSQVMDVFYQGQNKTRLP GVWPIDVNMRAKDVVSGSRFTLGGCADSLYEYLPKMHQLLGGGEPK YETMSRTFLQAADRHFVFRPMLPGAEEDVLMPGNVNVDEDSGEAVL DPETEHLACFVGGMFGLAGRLFSRPDDVETGVRLTNGCVYAYRAFP TGMMPERLDLAPCRDRSSRCPWDEEHWLEERAKRPEWEPHLPRGF TSAKDPRYLLRPEAIESVFYSYRITGRQEFQTAAWDMFTAVEKGTRT QFANAAVLDVTRAADELPQEDYMESFWLAETLKYFYLMFTTPDIISLD DYVLNTEAHPFKLVG SEQ ID NO: 77
MNS1 alternative4 e_gw1.13.279.1	-	MVMLVAIALAWL GCSLL SEQ ID NO: 78	RPVDAMRADYLAQLRQETVDMFYHGYSNYMEHAFPEDELRPISCTPL TRDRDNPGRISLNDALGNYSLTLIDSLSTLAILAGGPQNGPYTGPQAL SDFQDGVAEFVRHYGDGRSGPSGAGIRARGFDLDSKVQVFETVIRG VGGLLSAHLFAIGELPITGYVPRPEGVAGDDPLELAPIPWPNGFRYDG QLLRLALDLSERLLPAFYTPTGIPYPRVNLRSGIPFYVNSPLHQNLGEA VEEQSGRPEITETCSAGAGSLVLEFTVLSRLTGDARFEQAAKRAFWE VWHRRSEIGLIGNGIDAERGLWIGPHAGIGAGMDSFFEYALKSHILLS GLGMPNASTSRRQSTTSWLDPNSLHPPLPPEMHTSDAFLQAWHQAH ASVKRYLYTDRSHFPYYSNNHRATGQPYAMWIDSLGAFYPGLLALAG EVEEAIEANLVYTALWTRYSALPERWSVREGNVEAGIGWWPGRPEFI ESTYHIYRATRDPWYLHVGEMVLRDIRRRCYAECGWAGLQDVQTGE KQDRMESFFLGETAKYMYLLFDPDHPLNKLDAAYVFTTEGHPLIIPKS KRGSGSHNRQDRARKAKKSRDVAVYTYYDESFTNSCPAPRPPSEHH LIGSATAARPDLFSVSRFTDLYRTPNVHGPLEKVEMRDKKKGRVVRY RATSNHTIFPWTLPPAMLPENGTCAAPPERIISLIEFPANDITSGITSRF GNHLSWQTHLGPTVNILEGLRLQLEQVSDPATGEDKWRITHIG NTQLGRHETVFFHAEHVRHLKDEVFSCRRRRDAVEIELLVDKPSDTN NNNTLASSDDDVVVDAKAEEQDGMLADDDGDTLNAETLSSNSLFQSL LRAVSSVFEPVYTAIPESDPSAGTAKVYSFDAYTSTGPGAYPMPSI SDTPIPGNPFYNFRNPASNFPWSTVFLAGQACEGPLPASAPREHQVI VMLRGGCSFSRKLDNIPSFSPHDRALQLVVVLDEPPPPPPPPPANDR RDVTRPLLDTEQTTPKGMKRLHGIPMVLVRAARGDYELFGHAIGVG MRRKYRVESQGLVVENAVVL SEQ ID NO: 79
VAN1 estExt_GeneWise Plus.C_400029	MMPRHHSSGFSN GYPRADTFEISPH RFQPRATLPPHRK RKRTAIR SEQ ID NO: 80	VGIAVVVILVLVL WFG SEQ ID NO: 81	QPRSVASLISLGILSGYDDLKLETVRYYDLSNVQGTARGWEREERILL CVPLRDAEQHLPMFFSHLKNFTYPHNLIDLAFLVSDSKDHTLESLTEH LEAIQADPDPKQPYGEISIIEKDFGQKVNQDVESRHGFAAQASRRKLM AQARNWLLSAALRPYHSWVYWRDVDVETAPFTILEDLMRHNKDVIVP NVWRPLPDWLGGEQPYDLNSWQESETALALADTLDEDAVIVEGYAE YATWRPHLAYLRDPYGDPDMEMEIDGVGGVSILAKAKVFRAGVHFPA FSFEKHAETEGFGKMAKRMHFSVVGLPHYTIWHLYEPSVDDIKHMEE MERERIAREKEEEERKKKEAQIKEEFGDANSQWEQDKQQMQDLKLQ DRGGDKEAAAAGVNQGAAAKAAGAMEGQKN SEQ ID NO: 82
VAN1 alternative1 estExt_GeneWise Plus.C_230146	MLLPKGGLDWRS ARAQIPPTR ALWNAVTRTR SEQ ID NO: 83	FILLVGITGLILLLW SEQ ID NO: 84	RGVSTSASEMQSFYCWGPAKPPMEMSPNEHNRWNGHLQTPVIFNH HAPVEVNSSTIEHVDLNPINSTKQAVTKEERILILTPLKDAAPYLSKYF ELLAELTYPHRLIDLAFLVSDSTDDTLAVLASELDRIQKRPDQIPFHSAT VIEKDFGFKLSQNVEERHSFEAQGPRRKAMGRARNYLLYTALKPEHS WVYWRDVDIVDSPTGILEDFIAHDRDILVPNIWFHRYRDGVDIEGRFD

Homologous to	Cytoplasmic	Transmembrane	Luminal
			YNSWVESDKGRKLANSLDKDVVLAEGYKQYDTGRTYMAKMGDWRE NKDVELELDGIGGVNILVKADVHRSGINFPCYAFENQAETEGFAKMAK RAGYEVYGLPNYVVWHIDTEEKGGNA SEQ ID NO: 85
VAN1 alternative2 e_gw1.5.262.1	MARPMGSVRLKK ANPST SEQ ID NO: 86	LILGAVLCIFIIIFLV SEQ ID NO: 87	SPSSPASASRLSIVSAQHHLSPPTSPYQSPRSGAVQGPPPVTRYNLN KVTVTSDPVRNQEHILILTPMARFYQEYWDNLLRLNYPHELITLGFILP KTKEGNQATSMLQKQIQKTQNYGPEKDRFKSIIILRQDFDPAVVSQDE SERHKLANQKARREVMAKARNSLLFTTLGPSTSWVLWLDADITETAP TLIQDLASHDKPIIVANCFQKYYDPESKKMAERPYDFNSWQDSETALK MAEQMGPDDILLEGYAEMATYRTLLAYMSTPGGSKDLVVPLDGVGG TALLVKADVHRDGAMFPPFAFYHLIESEGFAKMAKRLGWQPYGLPNY KVYHYNE SEQ ID NO: 88
Other01 estExt_GeneWise Plus.C_150339	MHFAYPSRKSSN PPPFRPRSTRLPG LRRSRIKT SEQ ID NO: 89	IGIVLFLVLATLWF F SEQ ID NO: 90	SNPRVPRPDPERVPSGRPPVVLVTVIDPTQYPNAYLKTIKENREQYAA KHGYEAFIVKAYDYDTQGAPQSWSKLMAMRHALTKFPECRFVWYLD QDAYIMDMSKSLEEQLLNRQKLESLMIKNYPVVPPDSIIKTFSHLRPDE VDLIVSQDSSGLVAGSVVVRNSQWSKFLLETWMDPLYRSYNFQKAE RHALEHIVQWHPTILSKLALVPQRTLGPYTRTDQGDAYQDGDFVVMF TGCTKSGEQSCETVSASYYQKWSSSL SEQ ID NO: 91
Other02 fgenesh5_pg.C_s caffold_5000342	MSLSRSPSPVPG GGWSSPGLNINS GRSSPSNAAGSS VSWESAKMRKQG ANGYPSFSTQNQ GFFTRHMRRISSS LPRFAAGPGNTYA EREKYERGGHSP HAGGGRLRAFLA RIGRRLKWR SEQ ID NO: 92	ILLPLIIICTIVAYYG SEQ ID NO: 93	THEAPGFVHWWRRISMGGGGEKFVIILGANVGGGVMEWKGAREWAI ERDSVRNKRKYATRWGYDLEIVDMKTKKRYAHEWRESWEKVDFIRA AMRKYPKAEWFWWLDLNTYVMEPSYSLQRHLFNHLDRHVYRDINVF NPLNITHPPTEEYLDAEARSPVGDGNINSVNLMLTQDCSGFNLGSFFI RRSAWTEQLLDIWWDPVLYEQKHMEWEHKEQDALEQLYRTQPWIR QHTGFLPQRLINSFPPAACADESGLNNTRIHYNEKDRDFVVNMAGCE WGRDCWGEMYHYREFSYWLNRNPWELFKEEIVAVIWYKLTGQRVKL SEQ ID NO: 94

[00101] Uncharacterized sequences may be tested for use as targeting peptides by expressing proteins in the glycosylation pathway in a host cell, where one of the proteins contains the uncharacterized sequence as the sole targeting peptide, and measuring the glycans produced in view of the cytoplasmic localization of glycan biosynthesis (e.g. as in Schwientek JBC 1996 3398), or by expressing a fluorescent reporter protein fused with the targeting peptide, and analyzing the localization of the protein in the Golgi by immunofluorescence or by fractionating the cytoplasmic membranes of the Golgi and measuring the location of the protein.

[00102] The targeting peptide may include a stem domain. In certain embodiments, the stem domain is from an N-acetylglucosaminyltransferase I enzyme or an N-

acetylglucosaminyltransferase II enzyme. In especially certain embodiments, the stem domain is from a human N-acetylglucosaminyltransferase I enzyme or a human N-acetylglucosaminyltransferase II enzyme. The sequence corresponding to the stem domain from human N-acetylglucosaminyltransferase I enzyme is SEQ ID NO: 34. The sequence corresponding to the stem domain from human N-acetylglucosaminyltransferase II enzyme is residues 30-85 of SEQ ID NO: 20.

[00103] The targeting peptide may include a transmembrane domain. A "transmembrane domain" refers to any sequence of amino acid residues that is thermodynamically stable in a membrane as a three-dimensional structure. In embodiments where the targeting peptide also includes a stem domain, the transmembrane domain is N-terminal to the stem domain. In certain embodiments, the transmembrane domain is from an N-acetylglucosaminyltransferase I enzyme or an N-acetylglucosaminyltransferase II enzyme. In especially certain embodiments, the transmembrane domain is from a human N-acetylglucosaminyltransferase I enzyme or a human N-acetylglucosaminyltransferase II enzyme. The sequence corresponding to the transmembrane domain from human N-acetylglucosaminyltransferase I enzyme is residues 7-29 of SEQ ID NO: 1. The sequence corresponding to the transmembrane domain from human N-acetylglucosaminyltransferase II enzyme is residues 10-29 of SEQ ID NO: 20.

[00104] The targeting peptide may include a cytoplasmic domain. The term "cytoplasmic domain" refers to an amino acid sequence that is thermodynamically stable in a cytoplasmic environment as a three-dimensional structure. In embodiments where the targeting peptide also includes a stem domain, the cytoplasmic domain is N-terminal to the stem domain. In embodiments where the targeting peptide also includes a transmembrane domain, the cytoplasmic domain is N-terminal to the transmembrane domain. In certain embodiments, the cytoplasmic domain is from an N-acetylglucosaminyltransferase I enzyme or an N-acetylglucosaminyltransferase II enzyme. In especially certain embodiments, the cytoplasmic domain is from a human N-acetylglucosaminyltransferase I enzyme or a human N-acetylglucosaminyltransferase II enzyme. The sequence corresponding to the cytoplasmic domain from human N-acetylglucosaminyltransferase I enzyme is residues 1-6 of SEQ ID NO: 1. The sequence

corresponding to the cytoplasmic domain from human N-acetylglucosaminyltransferase II enzyme is residues 1-9 of SEQ ID NO: 20.

[00105] In certain embodiments, the recombinant protein contains a human GnTII catalytic domain N-terminal to a human GnTI catalytic domain with a spacer sequence containing human GnTI stem domain sequence in between the catalytic domains. In this embodiment, the recombinant protein also includes a targeting peptide N-terminal to the GnTII catalytic domain with cytoplasmic, transmembrane, and stem domains from human GnTII. The sequence of the recombinant protein in this embodiment is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 95, and the sequence of a possible cDNA encoding the recombinant protein of this embodiment is SEQ ID NO: 96.

[00106] In other embodiments, the recombinant protein contains a human GnTII catalytic domain N-terminal to a human GnTI catalytic domain with a spacer sequence. The spacer sequence may include, without limitation, a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NOs: 118, 120, 122, or 124. In this embodiment, the recombinant protein also includes a targeting peptide N-terminal to the GnTII catalytic domain with cytoplasmic, transmembrane, and stem domains from human GnTII. Accordingly, in certain embodiments, the sequence of the recombinant protein is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to a sequence selected from SEQ ID NOs: 119, 121, 123, and 125. In certain embodiments, the sequence of a possible cDNA encoding the recombinant protein of SEQ ID NO: 119 is SEQ ID NO: 141. In other embodiments, the sequence of a possible cDNA encoding the recombinant protein of SEQ ID NO: 121 is SEQ ID NO: 139. In still other embodiments, the sequence of a possible cDNA encoding the recombinant protein of SEQ ID NO: 123 is SEQ ID NO: 143. In further embodiments, the sequence of a possible cDNA encoding the recombinant protein of SEQ ID NO: 125 is SEQ ID NO: 145.

Production of Recombinant Proteins of the Invention

[00107] Another aspect of the invention includes isolated polynucleotides encoding the recombinant proteins of the invention. As used herein, the terms "polynucleotide," "nucleic acid sequence," "sequence of nucleic acids," and variations thereof shall be generic to polydeoxyribonucleotides (containing 2-deoxy-D-ribose), to polyribonucleotides (containing D-ribose), to any other type of polynucleotide that is an N-glycoside of a purine or pyrimidine base, and to other polymers containing non-nucleotidic backbones, provided that the polymers contain nucleobases in a configuration that allows for base pairing and base stacking, as found in DNA and RNA. Thus, these terms include known types of nucleic acid sequence modifications, for example, substitution of one or more of the naturally-occurring nucleotides with an analog; inter-nucleotide modifications, such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), with negatively charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), and with positively charged linkages (e.g., aminoalkylphosphoramidates, aminoalkylphosphotriesters); those containing pendant moieties, such as, for example, proteins (including nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.); those with intercalators (e.g., acridine, psoralen, etc.); and those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.). As used herein, the symbols for nucleotides and polynucleotides are those recommended by the IUPAC-IUB Commission of Biochemical Nomenclature (Biochem. 9:4022, 1970).

[00108] Sequences of the isolated polynucleotides are prepared by any suitable method known to those of ordinary skill in the art, including, for example, direct chemical synthesis or cloning. For direct chemical synthesis, formation of a polymer of nucleic acids typically involves sequential addition of 3'-blocked and 5'-blocked nucleotide monomers to the terminal 5'-hydroxyl group of a growing nucleotide chain, where each addition is effected by nucleophilic attack of the terminal 5'-hydroxyl group of the growing chain on the 3'-position of the added monomer, which is typically a phosphorus derivative, such as a phosphotriester, phosphoramidite, or the like. Such methodology is known to those of ordinary skill in the art and is described in the pertinent texts and literature [e.g., in Matteucci et al., (1980) Tetrahedron Lett 21:719-722; U.S. Pat. Nos. 4,500,707; 5,436,327; and 5,700,637]. In addition, the desired sequences may be isolated

from natural sources by splitting DNA using appropriate restriction enzymes, separating the fragments using gel electrophoresis, and thereafter, recovering the desired nucleic acid sequence from the gel via techniques known to those of ordinary skill in the art, such as utilization of polymerase chain reactions (PCR; e.g., U.S. Pat. No. 4,683,195).

[00109] Each polynucleotide of the invention can be incorporated into an expression vector. "Expression vector" or "vector" refers to a compound and/or composition that transduces, transforms, or infects a host cell, thereby causing the cell to express nucleic acids and/or proteins other than those native to the cell, or in a manner not native to the cell. An "expression vector" contains a sequence of nucleic acids (ordinarily RNA or DNA) to be expressed by the host cell. Optionally, the expression vector also comprises materials to aid in achieving entry of the nucleic acid into the host cell, such as a virus, liposome, protein coating, or the like. The expression vectors contemplated for use in the present invention include those into which a nucleic acid sequence can be inserted, along with any certain or required operational elements. Further, the expression vector must be one that can be transferred into a host cell and replicated therein. Certain expression vectors are plasmids, particularly those with restriction sites that have been well documented and that contain the operational elements certain or required for transcription of the nucleic acid sequence. Such plasmids, as well as other expression vectors, are well known to those of ordinary skill in the art.

[00110] Incorporation of the individual polynucleotides may be accomplished through known methods that include, for example, the use of restriction enzymes (such as BamHI, EcoRI, HhaI, XhoI, XmaI, and so forth) to cleave specific sites in the expression vector, e.g., plasmid. The restriction enzyme produces single-stranded ends that may be annealed to a polynucleotide having, or synthesized to have, a terminus with a sequence complementary to the ends of the cleaved expression vector. Annealing is performed using an appropriate enzyme, e.g., DNA ligase. As will be appreciated by those of ordinary skill in the art, both the expression vector and the desired polynucleotide are often cleaved with the same restriction enzyme, thereby assuring that the ends of the expression vector and the ends of the polynucleotide are complementary to each other. In addition, DNA linkers may be used to facilitate linking of nucleic acids sequences into an expression vector.

[00111] A series of individual polynucleotides can also be combined by utilizing methods that are known to those having ordinary skill in the art (e.g., U.S. Pat. No. 4,683,195).

[00112] For example, each of the desired polynucleotides can be initially generated in a separate PCR. Thereafter, specific primers are designed such that the ends of the PCR products contain complementary sequences. When the PCR products are mixed, denatured, and reannealed, the strands having the matching sequences at their 3' ends overlap and can act as primers for each other. Extension of this overlap by DNA polymerase produces a molecule in which the original sequences are "spliced" together. In this way, a series of individual polynucleotides may be "spliced" together and subsequently transduced into a host cell simultaneously. Thus, expression of each of the plurality of polynucleotides is affected.

[00113] Individual polynucleotides, or "spliced" polynucleotides, are then incorporated into an expression vector. The invention is not limited with respect to the process by which the polynucleotide is incorporated into the expression vector. Those of ordinary skill in the art are familiar with the necessary steps for incorporating a polynucleotide into an expression vector. A typical expression vector contains the desired polynucleotide preceded by one or more regulatory regions, along with a ribosome binding site, e.g., a nucleotide sequence that is 3-9 nucleotides in length and located 3-11 nucleotides upstream of the initiation codon in E. coli. See Shine and Dalgarno (1975) Nature 254(5495):34-38 and Steitz (1979) Biological Regulation and Development (ed. Goldberger, R. F.), 1:349−399 (Plenum, New York).

[00114] The term “operably linked” as used herein refers to a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of a DNA sequence or polynucleotide such that the control sequence directs the expression of a polypeptide.

[00115] Regulatory regions include, for example, those regions that contain a promoter and an operator. A promoter is operably linked to the desired polynucleotide or portion of a polynucleotide encoding a polypeptide, thereby initiating transcription of the polynucleotide, or portion of the polynucleotide encoding a polypeptide, via an RNA polymerase enzyme. An operator is a sequence of nucleic acids adjacent to the promoter,

which contains a protein-binding domain where a repressor protein can bind. In the absence of a repressor protein, transcription initiates through the promoter. When present, the repressor protein specific to the protein-binding domain of the operator binds to the operator, thereby inhibiting transcription. In this way, control of transcription is accomplished, based upon the particular regulatory regions used and the presence or absence of the corresponding repressor protein. Examples include lactose promoters (Lad repressor protein changes conformation when contacted with lactose, thereby preventing the Lad repressor protein from binding to the operator) and tryptophan promoters (when complexed with tryptophan, TrpR repressor protein has a conformation that binds the operator; in the absence of tryptophan, the TrpR repressor protein has a conformation that does not bind to the operator). Another example is the tac promoter (see de Boer et al., (1983) Proc Natl Acad Sci USA 80(1):21-25). As will be appreciated by those of ordinary skill in the art, these and other regulatory regions may be used in the present invention, and the invention is not limited in this respect.

[00116] Examples of certain promoters for linkage to the isolated polynucleotides encoding the recombinant proteins of the invention include promoters from the following genes: gpdA, cbh1, *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* glucoamylase (glaA), *Aspergillus awamori* glaA, *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Aspergillus oryzae* acetamidase, *Fusarium oxysporum* trypsin-like protease, fungal endo α-L-arabinase (abnA), fungal α-L-arabinofuranosidase A (abfA), fungal α-L-arabinofuranosidase B (abfB), fungal xylanase (xlnA), fungal phytase, fungal ATP-synthetase, fungal subunit 9 (oliC), fungal triose phosphate isomerase (tpi), fungal alcohol dehydrogenase (adhA), fungal α-amylase (amy), fungal amyloglucosidase (glaA), fungal acetamidase (amdS), fungal glyceraldehyde-3-phosphate dehydrogenase (gpd), yeast alcohol dehydrogenase, yeast lactase, yeast 3-phosphoglycerate kinase, yeast triosephosphate isomerase, bacterial α-amylase, bacterial Spo2, and SSO. In certain embodiments, isolated polynucleotides encoding the recombinant proteins of the invention are operably linked to a constitutive promoter. In other embodiments, isolated polynucleotides encoding the recombinant proteins of the invention are operably linked to an inducible promoter. In certain preferred embodiments, the inducible promoter is from a cbh1 gene.

[00117] Although any suitable expression vector may be used to incorporate the desired sequences, readily available expression vectors include, without limitation: plasmids, such as pSClOl, pBR322, pBBRlMCS-3, pUR, pEX, pMRlOO, pCR4, pBAD24, pUC19; bacteriophages, such as Ml 3 phage and λ phage. Of course, such expression vectors may only be suitable for particular host cells. One of ordinary skill in the art, however, can readily determine through routine experimentation whether any particular expression vector is suited for any given host cell. For example, the expression vector can be introduced into the host cell, which is then monitored for viability and expression of the sequences contained in the vector. In addition, reference may be made to the relevant texts and literature, which describe expression vectors and their suitability to any particular host cell.

[00118] Another aspect of the invention includes host cells containing expression vectors containing isolated polynucleotides that encode the recombinant proteins of the invention. "Host cell" as used herein refers to a living biological cell that can be transformed via insertion of recombinant DNA or RNA. Such recombinant DNA or RNA can be in an expression vector. Thus, a host cell as described herein may be a prokaryotic organism (e.g., an organism of the kingdom eubacteria) or a eukaryotic cell. As will be appreciated by one of ordinary skill in the art, a prokaryotic cell lacks a membrane-bound nucleus, while a eukaryotic cell has a membrane-bound nucleus. In certain embodiments, host cells used for production of the recombinant proteins of the invention are fungal cells such as yeast or filamentous fungi. In other embodiments, the host cells are mammalian cells. Such cells may be human or non-human.

[00119] Another aspect of the invention includes methods of producing the recombinant proteins of the invention. The method includes the steps of introducing an isolated polynucleotide that encodes the recombinant protein into a host cell, and culturing the host cell such that the recombinant protein is expressed. The method may also include a step of purifying the recombinant protein from the host cell.

[00120] Methods of producing the recombinant proteins of the invention may include the introduction or transfer of expression vectors containing the recombinant polynucleotides of the invention into the host cell. Such methods for transferring expression vectors into host cells are well known to those of ordinary skill in the art. For example, one method for transforming *E. coli* with an expression vector involves a

calcium chloride treatment where the expression vector is introduced via a calcium precipitate. Other salts, e.g., calcium phosphate, may also be used following a similar procedure. In addition, electroporation (i.e., the application of current to increase the permeability of cells to nucleic acid sequences) may be used to transfect the host cell. Also, microinjection of the nucleic acid sequences provides the ability to transfect host cells. Other means, such as lipid complexes, liposomes, and dendrimers, may also be employed. Those of ordinary skill in the art can transfect a host cell with a desired sequence using these or other methods.

[00121] The vector may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host, or a transposon may be used.

[00122] The vectors may contain one or more selectable markers which permit easy selection of transformed hosts. A selectable marker is a gene, the product of which provides, for example, biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Selection of bacterial cells may be based upon antimicrobial resistance that has been conferred by genes such as the amp, gpt, neo, and hyg genes.

[00123] Suitable markers for yeast hosts are, for example, ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine 5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Certain for use in *Aspergillus* are the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar gene of *Streptomyces hygroscopicus*. Certain for use in *Trichoderma* are bar, pyr4, and amdS.

[00124] The vectors may contain an element(s) that permits integration of the vector into the host's genome or autonomous replication of the vector in the cell independent of the genome.

[00125] For integration into the host genome, the vector may rely on the gene's sequence or any other element of the vector for integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleotide sequences for directing integration by homologous recombination into the genome of the host. The additional nucleotide sequences enable the vector to be integrated into the host genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements may contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host. Furthermore, the integrational elements may be non-encoding or encoding nucleotide sequences. On the other hand, the vector may be integrated into the genome of the host by non-homologous recombination.

[00126] For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host in question. The origin of replication may be any plasmid replicator mediating autonomous replication which functions in a cell. The term "origin of replication" or "plasmid replicator" is defined herein as a sequence that enables a plasmid or vector to replicate in vivo. Examples of origins of replication for use in a yeast host are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. Examples of origins of replication useful in a filamentous fungal cell are AMA1 and ANS1 (Gems et al., 1991; Cullen et al., 1987; WO 00/24883). Isolation of the AMA1 gene and construction of plasmids or vectors comprising the gene can be accomplished according to the methods disclosed in WO 00/24883.

[00127] For other hosts, transformation procedures may be found, for example, in Jeremiah D. Read, et al., Applied and Environmental Microbiology, Aug. 2007, p. 5088–5096, for *Kluyveromyces*, in Osvaldo Delgado, et al., FEMS Microbiology Letters 132,

1995, 23-26, for *Zymomonas*, in US 7,501,275 for *Pichia stipitis*, and in WO 2008/040387 for *Clostridium*.

[00128] More than one copy of a gene may be inserted into the host to increase production of the gene product. An increase in the copy number of the gene can be obtained by integrating at least one additional copy of the gene into the host genome or by including an amplifiable selectable marker gene with the nucleotide sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the gene, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[00129] The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well-known to one skilled in the art (see, e.g., Sambrook et al., 1989, supra).

[00130] The host cell is transformed with at least one expression vector. When only a single expression vector is used (without the addition of an intermediate), the vector will contain all of the nucleic acid sequences necessary.

[00131] Once the host cell has been transformed with the expression vector, the host cell is allowed to grow. Methods of the invention may include culturing the host cell such that recombinant nucleic acids in the cell are expressed. For microbial hosts, this process entails culturing the cells in a suitable medium. Typically, cells are grown at 35°C in appropriate media. Certain growth media in the present invention include, for example, common commercially-prepared media such as Luria-Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular host cell will be known by someone skilled in the art of microbiology or fermentation science. Temperature ranges and other conditions suitable for growth are known in the art (see, e.g., Bailey and Ollis 1986).

[00132] Methods for purifying recombinant proteins of the invention from the host cell are well known in the art (see E. L. V. Harris and S. Angel, Eds. (1989) Protein Purification Methods: A Practical Approach, IRL Press, Oxford, England). Such methods include, without limitation, preparative disc-gel electrophoresis, isoelectric focusing, high-performance liquid chromatography (HPLC), reversed-phase HPLC, gel filtration,

ion exchange and partition chromatography, and countercurrent distribution, and combinations thereof. In certain embodiments, the recombinant proteins carry additional sequence tags to facilitate purification. Such markers include epitope tags and protein tags. Non-limiting examples of epitope tags include c-myc, hemagglutinin (HA), polyhistidine (6x-HIS), GLU-GLU, and DYKDDDDK (FLAG) (SEQ ID NO: 117) epitope tags. Epitope tags can be added to peptides by a number of established methods. DNA sequences of epitope tags can be inserted into recombinant protein coding sequences as oligonucleotides or through primers used in PCR amplification. As an alternative, peptide-coding sequences can be cloned into specific vectors that create fusions with epitope tags; for example, pRSET vectors (Invitrogen Corp., San Diego, Calif.). Non-limiting examples of protein tags include glutathione-S-transferase (GST), green fluorescent protein (GFP), and maltose binding protein (MBP). Protein tags are attached to peptides or polypeptides by several well-known methods. In one approach, the coding sequence of a polypeptide or peptide can be cloned into a vector that creates a fusion between the polypeptide or peptide and a protein tag of interest. Suitable vectors include, without limitation, the exemplary plasmids, pGEX (Amersham Pharmacia Biotech, Inc., Piscataway, N.J.), pEGFP (CLONTECH Laboratories, Inc., Palo Alto, Calif.), and pMAL.TM. (New England BioLabs, Inc., Beverly, Mass.). Following expression, the epitope or protein-tagged polypeptide or peptide can be purified from a crude lysate of the host cell by chromatography on an appropriate solid-phase matrix. In some cases, it may be preferable to remove the epitope or protein tag (i.e., via protease cleavage) following purification.

Methods of Producing Complex Glycans

[00133] Another aspect of the invention includes methods of producing a complex N-glycan, including the steps of providing a host cell, where the host cell contains a polynucleotide encoding a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain and culturing the host cell such that the fusion protein is expressed, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan. In certain embodiments, this aspect includes methods of producing human-like N-glycans in a *Trichoderma* cell.

[00134] As used herein, the term "complex N-glycan" refers to an N-glycan comprising a terminal $GlcNAc_2Man_3$ structure.

[00135] The complex N-glycan includes any glycan having the formula $[GlcNAc\beta2]_zMan\alpha3([GlcNAc\beta2]_wMan\alpha6)Man\{\beta4GlcNAc\beta(Fuc\alpha x)_n[4GlcNAc]_m\}_p$, where n, m, and p are 0 or 1, indicating presence or absence of part of the molecule, with the provision that when m is 0, then n is 0 (fucose is a branch linked to the GlcNAc), where x is 3 or 6, where ( ) defines a branch in the structure, where [ ] defines a part of the glycan structure either present or absent in a linear sequence, and where z and w are 0 or 1. Preferably w and z are 1. In certain embodiments, the complex N-glycan includes GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4GlcNAc, GlcNAcβ2Manα3(Manα6)Manβ4GlcNAcβ4GlcNAc, GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4(Fucα6)GlcNAc, GlcNAcβ2Manα3(Manα6)Manβ4GlcNAcβ4(Fucα6)GlcNAc, and Manα3(Manα6)Manβ4GlcNAcβ4GlcNAc. In certain embodiments, the complex N-glycans are fungal non-fucosylated GlcNAcMan3, GlcNAc2Man3, and or Man3

[00136] In certain embodiments, the method of producing a complex N-glycan will generate a mixture of different glycans. The complex N-glycan may constitute at least 1%, at least 3%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, or at least 75% or more of such a glycan mixture.

[00137] The acceptor glycan, and thus the complex N-glycan, may be attached to a molecule such as an amino acid, a peptide, or a polypeptide. In certain embodiments, the amino acid derivative is an asparagine residue. The asparagine residue may be in aminoglycosidic linkage from the side-chain amide (a biologic mammalian polypeptide N-glycan linkage structure) and may be part of a peptide chain such as a dipeptide, an oligopeptide, or a polypeptide. The glycan may be a reducing end derivative such as an N-, O-, or C-linked, preferably glycosidic, derivative of the reducing GlcNAc or Man, such as a spacer or terminal organic residue with a certain glycan linked structure selected from the group of an amino acid, alkyl, heteroalkyl, acyl, alkyloxy, aryl, arylalkyl, and heteroarylalkyl. The spacer may be further linked to a polyvalent carrier or a solid phase. In certain embodiments, alkyl-containing structures include methyl, ethyl, propyl, and C4-C26 alkyls, lipids such as glycerolipids, phospholipids, dolichol-phospholipids and

ceramides and derivatives. The reducing end may also be derivatized by reductive amination to a secondary amine linkage or a derivative structure. Certain carriers include biopoly- or oligomers such as (poly)peptides, poly(saccharides) such as dextran, cellulose, amylose, or glycosaminoglycans, and other organic polymers or oligomers such as plastics including polyethylene, polypropylene, polyamides (e.g., nylon or polystyrene), polyacrylamide, and polylactic acids, dendrimers such as PAMAM, Starburst or Starfish dendrimers, or polylysine, and polyalkylglycols such as polyethylene glycol (PEG). Solid phases may include microtiter wells, silica particles, glass, metal (including steel, gold and silver), polymer beads such as polystyrene or resin beads, polylactic acid beads, polysaccharide beads or organic spacers containing magnetic beads.

[00138] In certain embodiments, the acceptor glycan is attached to a heterologous polypeptide. In certain embodiments, the heterologous polypeptide is a therapeutic protein. Therapeutic proteins may include monoclonal antibodies, erythropoietins, interferons, growth hormones, enzymes, or blood-clotting factors and may be useful in the treatment of humans or animals. For example, the acceptor glycan may be attached to a therapeutic protein such as rituximab.

[00139] The acceptor glycan may be any of the acceptor glycans described in the section entitled, "Recombinant Proteins of the Invention."

[00140] In certain embodiments, the acceptor glycan may be Man5. In such embodiments, a Man5 expressing T. reesei strain is transformed with a GnTII/GnTI fusion enzyme using random integration or by targeted integration to a known site known not to affect Man5 glycosylation. Strains that produce GlcNAcMan5 are selected. The selected strains are further transformed with a catalytic domain of a mannosidase II-type mannosidase capable of cleaving Man5 structures to generate GlcNAcMan3. In certain embodiments mannosidase II-type enzymes belong to glycoside hydrolase family 38 (cazy.org/GH38_all.html). Characterized enzymes include enzymes listed in cazy.org/GH38_characterized.html. Especially useful enzymes are Golgi-type enzymes that cleaving glycoproteins, such as those of subfamily α-mannosidase II (Man2A1;ManA2). Examples of such enzymes include human enzyme AAC50302, *D. melanogaster* enzyme (Van den Elsen J.M. et al (2001) EMBO J. 20**:** 3008-3017), those with the 3D structure according to PDB-reference 1HTY, and others referenced with the catalytic domain in PDB. For cytoplasmic expression, the catalytic domain of the

mannosidase is typically fused with an N-terminal targeting peptide or expressed with endogenous animal or plant Golgi targeting structures of animal or plant mannosidase II enzymes. After transformation with the catalytic domain of a mannosidase II-type mannosidase, a strain effectively producing GlcNAc2Man3 is selected.

Host Cells

[00141] The methods of producing a complex N-glycan include a first step of providing a host cell. Any prokaryotic or eukaryotic host cell may be used in the present invention so long as it remains viable after being transformed with a sequence of nucleic acids. Preferably, the host cell is not adversely affected by the transduction of the necessary nucleic acid sequences, the subsequent expression of recombinant proteins, or the resulting intermediates. Suitable eukaryotic cells include, but are not limited to, fungal, plant, insect or mammalian cells.

[00142] In certain embodiments, the host is a fungal strain. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al.*, In, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth *et al.*, 1995, supra, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, supra).

[00143] In particular embodiments, the fungal host is a yeast strain. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in Biology and Activities of Yeast (Skinner, F. A., Passmore, S. M., and Davenport, R. R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

[00144] In certain embodiments, the yeast host is a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* strain.

[00145] In certain embodiments, the yeast host is *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Pichia pastoris*, *Candida albicans*, *Hansenula polymorpha*, *Schizosaccharomyces*, or *Yarrowia*.

[00146] In another particular embodiment, the fungal host cell is a filamentous fungal strain. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, supra). The filamentous fungi are generally characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative.

[00147] The filamentous fungal host cell may be, for example, an *Acremonium, Aspergillus, Fusarium, Humicola, Mucor, Myceliophthora, Neurospora, Penicillium, Scytalidium, Thielavia, Tolypocladium*, or *Trichoderma* strain.

[00148] In certain embodiments, the filamentous fungal host cell is a *Trichoderma* sp., *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, or *Tolypocladium* strain.

[00149] In certain embodiments, the host cell is a mammalian cell. Such cells may be human or non-human.

[00150] In other certain embodiments, the host cell is prokaryotic, and in certain embodiments, the prokaryotes are *E. coli*, *Bacillus subtilis*, *Zymomonas mobilis*, *Clostridium* sp., *Clostridium phytofermentans*, *Clostridium thermocellum*, *Clostridium beijerinckii*, *Clostridium acetobutylicum* (*Moorella thermoacetica*), *Thermoanaerobacterium saccharolyticum*, or *Klebsiella oxytoca*. In other embodiments, the prokaryotic host cells are *Carboxydocella* sp., *Corynebacterium glutamicum*, *Enterobacteriaceae*, *Erwinia chrysanthemi*, *Lactobacillus* sp., *Pediococcus acidilactici*, *Rhodopseudomonas capsulata*, *Streptococcus lactis*, *Vibrio furnissii*, *Vibrio furnissii* M1, *Caldicellulosiruptor saccharolyticus*, or *Xanthomonas campestris*. In other embodiments, the host cells are cyanobacteria. Additional examples of bacterial host cells include, without limitation, those species assigned to the *Escherichia, Enterobacter, Azotobacter, Erwinia, Bacillus, Pseudomonas, Klebsiella, Proteus, Salmonella, Serratia,*

Shigella, Rhizobia, Vitreoscilla, Synechococcus, Synechocystis, and *Paracoccus* taxonomical classes.

[00151] In methods of the invention for producing a complex N-glycan, the methods include a step of culturing the host cell such that the fusion protein is expressed. For microbial hosts, this process entails culturing the cells in a suitable medium. Typically, cells are grown at 35°C in appropriate media. Certain growth media in the present invention include, for example, common commercially-prepared media such as Luria-Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular host cell will be known by someone skilled in the art of microbiology or fermentation science. Temperature ranges and other conditions suitable for growth are known in the art (see, e.g., Bailey and Ollis 1986). In certain embodiments the pH of cell culture is between 3.5 and 7.5, between 4.0 and 7.0, between 4.5 and 6.5, between 5 and 5.5, or at 5.5.

[00152] The host cells used in the methods of producing a complex N-glycan contain a polynucleotide encoding any of the recombinant proteins of the invention as described in the section entitled "Recombinant Proteins of the Invention." In certain embodiments, the host cell contains a polynucleotide encoding a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan.

[00153] In certain embodiments, the host cell contains a polynucleotide encoding a UDP-GlcNAc transporter. The polynucleotide encoding the UDP-GlcNAc transporter may be endogenous (i.e., naturally present) in the host cell, or it may be heterologous to the host cell.

[00154] In certain embodiments, the host cell contains a polynucleotide encoding a α-1,2-mannosidase. The polynucleotide encoding the α-1,2-mannosidase may be endogenous in the host cell, or it may be heterologous to the host cell. These polynucleotides are especially useful for a host cell expressing high-mannose glycans transferred from the Golgi to the ER without effective exo-α-2-mannosidase cleavage.

The α-1,2-mannosidase may be a mannosidase I type enzyme belonging to the glycoside hydrolase family 47 (cazy.org/GH47_all.html). In certain embodiments the α-1,2-mannosidase is an enzyme listed at cazy.org/GH47_characterized.html. In particular, the α-1,2-mannosidase may be an ER-type enzyme that cleaves glycoproteins such as enzymes in the subfamily of ER α-mannosidase I EC 3.2.1.113 enzymes. Examples of such enzymes include human α-2-mannosidase 1B (AAC26169), a combination of mammalian ER mannosidases, or a filamentous fungal enzyme such as α-1,2-mannosidase (MDS1) (*T. reesei* AAF34579; Maras M et al J Biotech. 77, 2000, 255). For cytoplasmic expression the catalytic domain of the mannosidase is typically fused with a targeting peptide, such as HDEL, KDEL, or part of an ER or early Golgi protein, or expressed with an endogenous ER targeting structures of an animal or plant mannosidase I enzyme.

[00155] In certain embodiments, the host cell contains a polynucleotide encoding a galactosyltransferase. Galactosyltransferases transfer β-linked galactosyl residues to terminal N-acetylglucosaminyl residue. In certain embodiments the galactosyltransferase is a β-4-galactosyltransferase. Generally, β–4-galactosyltransferases belong to the CAZy glycosyltransferase family 7 (cazy.org/GT7_all.html) and include β-N-acetylglucosaminyl-glycopeptide β-1,4-galactosyltransferase (EC 2.4.1.38), which is also known as N-acetylactosamine synthase (EC 2.4.1.90). Useful subfamilies include β4-GalT1, β4-GalT-II, -III, -IV, -V, and -VI, such as mammalian or human β4-GalTI or β4GalT-II, -III, -IV, -V, and -VI or any combinations thereof. β4-GalT1, β4-GalTII, or β4-GalTIII are especially useful for galactosylation of terminal GlcNAcβ2-structures on N-glycans such as GlcNAcMan3, GlcNAc2Man3, or GlcNAcMan5 (Guo S. et al. Glycobiology 2001, 11:813-20). The three-dimensional structure of the catalytic region is known (e.g. (2006) J.Mol.Biol. 357: 1619-1633), and the structure has been represented in the PDB database with code 2FYD. The CAZy database includes examples of certain enzymes. Characterized enzymes are also listed in the CAZy database at cazy.org/GT7_characterized.html. Examples of useful β4GalT enzymes include β4GalT1, e.g. bovine *Bos taurus* enzyme AAA30534.1 (Shaper N.L. et al Proc. Natl. Acad. Sci. U.S.A. 83 (6), 1573-1577 (1986)), human enzyme (Guo S. et al. Glycobiology 2001, 11:813-20), and *Mus musculus* enzyme AAA37297 (Shaper, N.L. et al. 1998 J. Biol. Chem. 263 (21), 10420-10428); β4GalTII enzymes such as human β4GalTII BAA75819.1, Chinese hamster *Cricetulus griseus* AAM77195, *Mus musculus* enzyme

BAA34385, and Japanese Medaka fish *Oryzias latipes* BAH36754; and β4GalTIII enzymes such as human β4GalTIII BAA75820.1, Chinese hamster *Cricetulus griseus* AAM77196 and *Mus musculus* enzyme AAF22221.

[00156] The galactosyltransferase may be expressed in the cytoplasm of the host cell. A heterologous targeting peptide, such as a Kre2 peptide described in Schwientek J.Biol. Chem 1996 3398, may be used. Promoters that may be used for expression of the galactosyltransferase include constitutive promoters such as *gpd,* promoters of endogenous glycosylation enzymes and glycosyltransferases such as mannosyltransferases that synthesize N-glycans in the Golgi or ER, and inducible promoters of high-yield endogenous proteins such as the *cbh1* promoter.

[00157] In certain embodiments of the invention where the host cell contains a polynucleotide encoding a galactosyltransferase, the host cell also contains a polynucleotide encoding a UDP-Gal and/or UDP-Gal transporter. In certain embodiments of the invention where the host cell contains a polynucleotide encoding a galactosyltransferase, lactose may be used as the carbon source instead of glucose when culturing the host cell. The culture medium may be between pH 4.5 and 7.0 or between 5.0 and 6.5. In certain embodiments of the invention where the host cell contains a polynucleotide encoding a galactosyltransferase and a polynucleotide encoding a UDP-Gal and/or UDP-Gal transporter, a divalent cation such as Mn2+, Ca2+ or Mg2+ may be added to the cell culture medium.

[00158] In certain embodiments, the host cell contains a polynucleotide encoding a sialyltransferase. A sialyltransferase transfers α3- or α6-linked sialic acid, such as Neu5Ac, to the terminal Gal of galactosylated complex glycans. Examples of suitable sialyltransferases can be found in the glycosylation protein family 29 (cazy.org/GT29.html). Useful α3- or α6-sialyltransferases include β-galactoside α-2,6-sialyltransferase (EC 2.4.99.1) with a certain subfamily ST6Gal-I, and N-acetylactosaminide α-2,3-sialyltransferase (EC 2.4.99.6) with possible cross-reactivity with β-galactoside α-2,3-sialyltransferase (EC 2.4.99.4). Useful subtypes of α3-sialyltransferases include ST3Gal-III and ST3Gal-IV. Certain enzymatically characterized species of these are listed as characterized in the CAZy database of glycosylation enzymes (cazy.org/GT29_characterized.html). The polynucleotide

encoding the α3- or α6-linked sialyltransferase may be endogenous to the host cell, or it may be heterologous to the host cell. Sialylation in the host cell may require expression of enzymes synthesizing the donor CMP-sialic acid such as CMP-Neu5Ac, especially in fungal, plant, nematode/parasite, or insect cells.

[00159] The host cell may have increased or reduced levels of activity of various endogenous enzymes. A reduced level of activity may be provided by inhibiting the activity of the endogenous enzyme with an inhibitor, an antibody, or the like. In certain embodiments, the host cell is genetically modified in ways to increase or reduce activity of various endogenous enzymes. "Genetically modified" refers to any recombinant DNA or RNA method used to create a prokaryotic or eukaryotic host cell that expresses a polypeptide at elevated levels, at lowered levels, or in a mutated form. In other words, the host cell has been transfected, transformed, or transduced with a recombinant polynucleotide molecule, and thereby been altered so as to cause the cell to alter expression of a desired protein.

[00160] Genetic modifications which result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e., the protein encoded by the gene) can be referred to as inactivation (complete or partial), deletion, interruption, blockage, silencing, or down-regulation, or attenuation of expression of a gene. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such gene, can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein does not exist), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no enzymatic activity or action). More specifically, reference to decreasing the action of proteins discussed herein generally refers to any genetic modification in the host cell in question, which results in decreased expression and/or functionality (biological activity) of the proteins and includes decreased activity of the proteins (e.g., decreased catalysis), increased inhibition or degradation of the proteins as well as a reduction or elimination of expression of the proteins. For example, the action or activity of a protein of the present invention can be decreased by blocking or reducing the production of the protein, reducing protein action, or inhibiting the action of the protein. Combinations of some of these modifications are

also possible. Blocking or reducing the production of a protein can include placing the gene encoding the protein under the control of a promoter that requires the presence of an inducing compound in the growth medium. By establishing conditions such that the inducer becomes depleted from the medium, the expression of the gene encoding the protein (and therefore, of protein synthesis) could be turned off. Blocking or reducing the action of a protein could also include using an excision technology approach similar to that described in U.S. Pat. No. 4,743,546. To use this approach, the gene encoding the protein of interest is cloned between specific genetic sequences that allow specific, controlled excision of the gene from the genome. Excision could be prompted by, for example, a shift in the cultivation temperature of the culture, as in U.S. Pat. No. 4,743,546, or by some other physical or nutritional signal.

[00161] In general, according to the present invention, an increase or a decrease in a given characteristic of a mutant or modified protein (e.g., enzyme activity) is made with reference to the same characteristic of a wild-type (i.e., normal, not modified) protein that is derived from the same organism (from the same source or parent sequence), which is measured or established under the same or equivalent conditions. Similarly, an increase or decrease in a characteristic of a genetically modified host cell (e.g., expression and/or biological activity of a protein, or production of a product) is made with reference to the same characteristic of a wild-type host cell of the same species, and preferably the same strain, under the same or equivalent conditions. Such conditions include the assay or culture conditions (e.g., medium components, temperature, pH, etc.) under which the activity of the protein (e.g., expression or biological activity) or other characteristic of the host cell is measured, as well as the type of assay used, the host cell that is evaluated, etc. As discussed above, equivalent conditions are conditions (e.g., culture conditions) which are similar, but not necessarily identical (e.g., some conservative changes in conditions can be tolerated), and which do not substantially change the effect on cell growth or enzyme expression or biological activity as compared to a comparison made under the same conditions.

[00162] Preferably, a genetically modified host cell that has a genetic modification that increases or decreases the activity of a given protein (e.g., an enzyme) has an increase or decrease, respectively, in the activity or action (e.g., expression, production and/or biological activity) of the protein, as compared to the activity of the wild-type protein in a

wild-type host cell, of at least about 5%, and more preferably at least about 10%, and more preferably at least about 15%, and more preferably at least about 20%, and more preferably at least about 25%, and more preferably at least about 30%, and more preferably at least about 35%, and more preferably at least about 40%, and more preferably at least about 45%, and more preferably at least about 50%, and more preferably at least about 55%, and more preferably at least about 60%, and more preferably at least about 65%, and more preferably at least about 70%, and more preferably at least about 75%, and more preferably at least about 80%, and more preferably at least about 85%, and more preferably at least about 90%, and more preferably at least about 95%, or any percentage, in whole integers between 5% and 100% (e.g., 6%, 7%, 8%, etc.). The same differences are certain when comparing an isolated modified nucleic acid molecule or protein directly to the isolated wild-type nucleic acid molecule or protein (e.g., if the comparison is done *in vitro* as compared to *in vivo*).

[00163] In another aspect of the invention, a genetically modified host cell that has a genetic modification that increases or decreases the activity of a given protein (e.g., an enzyme) has an increase or decrease, respectively, in the activity or action (e.g., expression, production and/or biological activity) of the protein, as compared to the activity of the wild-type protein in a wild-type host cell, of at least about 2-fold, and more preferably at least about 5-fold, and more preferably at least about 10-fold, and more preferably about 20-fold, and more preferably at least about 30-fold, and more preferably at least about 40-fold, and more preferably at least about 50-fold, and more preferably at least about 75-fold, and more preferably at least about 100-fold, and more preferably at least about 125-fold, and more preferably at least about 150-fold, or any whole integer increment starting from at least about 2-fold (e.g., 3-fold, 4-fold, 5-fold, 6-fold, etc.).

[00164] In certain embodiments, the host cell has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase compared to the level of activity in a wild-type host cell. Dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase (EC 2.4.1.130) transfers an alpha-D-mannosyl residue from dolichyl-phosphate D-mannose into a membrane lipid-linked oligosaccharide. Typically, the dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase enzyme is encoded by an *alg3* gene. In certain embodiments, the host cell has a reduced level of

expression of an *alg3* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *alg3* gene is deleted from the host cell.

[00165] In certain embodiments, the host cell has a reduced level of activity of a alpha-1,6-mannosyltransferase compared to the level of activity in a wild-type host cell. Alpha-1,6-mannosyltransferase (EC 2.4.1.232) transfers an alpha-D-mannosyl residue from GDP-mannose into a protein-linked oligosaccharide, forming an elongation initiating alpha-(1->6)-D-mannosyl-D-mannose linkage in the Golgi apparathus.. Typically, the alpha-1,6-mannosyltransferase enzyme is encoded by an *och1* gene. In certain embodiments, the host cell has a reduced level of expression of an *och1* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *och1* gene is deleted from the host cell.

[00166] In certain embodiments, the host cell has a reduced level of protease activity. In certain embodiments, genes encoding various proteases are deleted from the host cell. These genes include, for example, genes encoding proteases such as pep1 (pepA in *Aspergillus*) and cellulolytic enzymes, such as cellobiohydrolase1 (cbh1).

[00167] In certain embodiments, the host cell may have a reduced level of activity of proteins involved in non-homologous end joining (NHEJ) in order to enhance the efficiency of homologous recombination. In certain embodiments, genes encoding these proteins are deleted from the host cell. The genes and their homologues include, but are not limited to, Ku70, Ku80, Lig4, Rad50, Xrs2, Sir4, Lifl, or Neil as described in, for example, Ninomiya et al. 2004, Ishibashi et al. 2006, Villalba et al. 2008, and Mizutani et al. 2008.

[00168] In certain embodiments of methods of producing a complex N-glycan, the host cell is a *Trichoderma* cell that has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase compared to the level of activity in a wild-type *Trichoderma* cell.

[00169] In other certain embodiments of methods of producing a complex N-glycan, the host cell is a yeast cell that has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase and a reduced level of activity of an alpha-1,6-mannosyltransferase compared to the levels of activity in a wild-type yeast cell and further comprises a polynucleotide encoding a α-1,2-mannosidase.

In Vitro Methods of Producing Complex N-Glycans

[00170] In another aspect, the invention provides a method of producing a complex N-glycan, including a step of incubating a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, an acceptor glycan, and an N-acetylglucosamine donor together in a buffer, where the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan. In certain embodiments the acceptor glycan is attached to an amino acid, a peptide, or a polypeptide. In certain embodiments the acceptor glycan is attached to a heterologous polypeptide. In certain embodiments, the acceptor glycan is Man_3. In certain embodiments the N-acetylglucosamine donor is a UDP-GlcNAc transporter. Typically the buffer contains a divalent cation such as Mn^{2+}, Ca^{2+}, or Mg^{2+} at concentrations of 1 μM to 100 mM, 100 μM to 50 mM, or 0.1 mM to 25 mM. The N-acetylglucosamine donor is typically used in molar excess, such as 1.1- 100 fold excess with regard to the reactive acceptor sites on the acceptor glycan. The concentration of the acceptor glycan is typically between 1 μM to 100 mM, 100 μM to 50 mM, or 1 to 25 mM. Where the acceptor glycan is attached to a polypeptide, the concentration ranges are typically at the lower end because of higher molecular weights. The concentrations of the components of the reaction may be adjusted based on their solubilities in the buffer. The amount of enzyme activity (units) may be adjusted to allow an effective reaction within a reasonable reaction time. A reasonable reaction time is typically from a few minutes to several days. In certain embodiments the reaction time will be from about 0.5 hours to one day or from 1 to 6 hours.

[00171] Useful buffers include buffers suitable for the fusion protein such as TRIS, HEPES, MOPS in pH ranges of about 5 to 8.5, 5.5. to 8.0, or 6.0 and 7.5. Typically concentrations of TRIS, HEPES, or MOPS buffers will be between 5 to 150 mM, between 10-100 mM, or 10-60 mM adjusted to maintain the pH. The reaction may be optimized by adding salt such as NaCl at 10-200 mM and/or an enzyme stabilizing but not glycosylatable protein (e.g., a pure non-glycosylated or non-acceptor glycan containing albumin. In a certain embodiment the *in vitro* reaction is adjusted to be performed in cell culture medium. Phosphate buffers may be used to reduce reaction speed.

Cells and Methods for Production of $Man_3GlcNAc_2$ Glycans

[00172] In another aspect, the present invention provides filamentous fungal cells containing a mutation of *alg3* and Man3GlcNAc2, where the Man3GlcNAc2 includes at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% (mol %) of neutral N-glycans secreted by the cells. The neutral N-glycans may be attached to an amino acid, a peptide, or a polypeptide. The *alg3* gene may be mutated by any means known in the art, such as point mutations or deletion of the entire *alg3* gene. Preferably, the function of the alg3 protein is reduced or eliminated by the mutation of *alg3*. The filamentous fungal cell may be an *Acremonium, Aspergillus, Aureobasidium, Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium, Fusarium, Gibberella, Humicola, Magnaporthe, Mucor, Myceliophthora, Myrothecium, Neocallimastix, Neurospora, Paecilomyces, Penicillium, Piromyces, Schizophyllum, Scytalidium, Talaromyces, Thermoascus, Thielavia, Tolypocladium*, or *Trichoderma* cell. In certain embodiments, the filamentous fungal cell is a *T. reesei* cell. In certain embodiments, the filamentous fungal cell further contains one or more polynucleotides encoding any of the recombinant proteins of the invention. For example, the filamentous fungal cell may further contain a first polynucleotide encoding an N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding an N-acetylglucosaminyltransferase II catalytic domain. Alternatively, the filamentous fungal cell may further contain a polynucleotide encoding a fusion protein including an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain.

[00173] In yet another aspect, the present invention provides methods of producing a $Man_3GlcNAc_2$ glycan in a host cell, including the steps of providing a host cell with a reduced level of activity of a mannosyltransferase compared to the level of activity in a wild-type host cell, and culturing the host cell to produce a $Man_3GlcNAc_2$ glycan, where the $Man_3GlcNAc_2$ glycan makes up at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or 100% (mol %) of the neutral N-glycans secreted by the host cell.

[00174] The $Man_3GlcNAc_2$ glycan may be attached to a molecule such as an amino acid, a peptide, or a polypeptide. In certain embodiments, the amino acid is an asparagine residue. The asparagine residue may be in aminoglycosidic linkage from the side-chain amide (a biologic mammalian protein N-glycan linkage structure) and may be part of a

peptide chain such as a dipeptide, an oligopeptide, or a polypeptide. The glycan may be a reducing end derivative such as an N-, O-, or C-linked, preferably glycosidic, derivative of the reducing GlcNAc or Man, such as a spacer or terminal organic residue with a certain glycan-linked structure selected from the group of an amino acid, alkyl, heteroalkyl, acyl, alkyloxy, aryl, arylalkyl, and heteroarylalkyl. The spacer may be further linked to a polyvalent carrier or a solid phase. In certain embodiments, alkyl-containing structures include methyl, ethyl, propyl, and C4-C26 alkyls, lipids such as glycerolipids, phospholipids, dolichol-phospholipids and ceramides and derivatives. The reducing end may also be derivatized by reductive amination to a secondary amine linkage or a derivative structure. Certain carriers include biopoly- or oligomers such as (poly)peptides, poly(saccharides) such as dextran, cellulose, amylose, or glycosaminoglycans, and other organic polymers or oligomers such as plastics including polyethylene, polypropylene, polyamides (e.g., nylon or polystyrene), polyacrylamide, and polylactic acids, dendrimers such as PAMAM, Starburst or Starfish dendrimers, or polylysine, and polyalkylglycols such as polyethylene glycol (PEG). Solid phases may include microtiter wells, silica particles, glass, metal including steel, gold and silver, polymer beads such as polystyrene or resin beads, polylactic acid beads, polysaccharide beads or organic spacers containing magnetic beads.

[00175] In certain embodiments, the $Man_3GlcNAc_2$ glycan is attached to a heterologous polypeptide. In certain embodiments, the heterologous polypeptide is a therapeutic protein. Therapeutic proteins may include monoclonal antibodies, erythropoietins, interferons, growth hormones, enzymes, or blood-clotting factors and may be useful in the treatment of humans or animals. For example, the $Man_3GlcNAc_2$ glycan may be attached to a therapeutic protein such as rituximab. Typically, the $Man_3GlcNAc_2$ glycan will be further modified to become a complex glycan. Such modification may take place *in vivo* in the host cell or by *in vitro* methods.

[00176] In certain embodiments, the mannosyltransferase is a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase. Typically, the dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase enzyme is encoded by an *alg3* gene. In certain embodiments, the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell. In certain embodiments, the *alg3* gene is deleted from the host cell. SEQ ID NOs: 97 and 98

provide the nucleic acid and amino acid sequences of the *alg3* gene in *T. reesei*, respectively.

[00177] In certain embodiments, the level of activity of alpha-1,6-mannosyltransferase in the host cell is not reduced compared to the level of activity in a wild-type host cell. Typically, the alpha-1,6-mannosyltransferase enzyme is encoded by an *och1* gene. In certain embodiments, the host cell contains an endogenous polynucleotide encoding an α-1,2-mannosidase.

[00178] In certain embodiments, the host cell is a *Trichoderma* cell, and in certain embodiments, the host cell is a *Trichoderma reesei* cell.

Filamentous Fungal Cells of the Invention

[00179] In a further aspect, the present invention provides filamentous fungal cells having a reduced level of expression of an *alg3* gene of the invention, compared to the level of expression of the *alg3* gene in a wild-type filamentous fungal cell, where the filamentous fungal cell also contains any of the recombinant proteins of the invention as described in the section entitled "Recombinant Proteins of the Invention.". For example, in certain embodiments the filamentous fungal cell further contains a polynucleotide encoding a fusion protein including an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain. The expression of the fusion protein may be controlled by a promoter that is operably linked to the polynucleotide. The promoter may be a constitutive promoter or an inducible promoter. In certain preferred embodiments, the promoter is an inducible promoter, such as the *cbh1* inducible promoter.

[00180] In another aspect, the present invention provides filamentous fungal cells having a reduced level of expression of an *alg3* gene of the invention, compared to the level of expression of the *alg3* gene in a wild-type filamentous fungal cell, where the filamentous fungal cell also contains a first polynucleotide encoding a recombinant N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding a recombinant N-acetylglucosaminyltransferase II catalytic domain. In such embodiments, the expression of the recombinant N-acetylglucosaminyltransferase I catalytic domain is controlled by a promoter that is operably linked to the first polynucleotide and the expression of the recombinant N-acetylglucosaminyltransferase II catalytic domain is

controlled by a promoter that is operably linked to the second polynucleotide. The promoter may be a constitutive promoter or an inducible promoter. In certain preferred embodiments, the promoter is an inducible promoter, such as the *cbh1* inducible promoter.

[00181] In other embodiments, a single polynucleotide may encode both the recombinant N-acetylglucosaminyltransferase I catalytic domain and the recombinant N-acetylglucosaminyltransferase II catalytic domain such that they are expressed as separate polypeptides. In such embodiments, the polynucleotide may contain an internal ribosome entry site that allows for the separate translation of each catalytic domain from the polynucleotide. In such embodiments, the expression of the recombinant N-acetylglucosaminyltransferase I catalytic domain is controlled by a promoter that is operably linked to the portion of the polynucleotide that encodes the N-acetylglucosaminyltransferase I catalytic domain and the expression of the recombinant N-acetylglucosaminyltransferase II catalytic domain is controlled by a promoter that is operably linked to the portion of the polynucleotide that encodes the N-acetylglucosaminyltransferase II catalytic domain. The promoter may be a constitutive promoter or an inducible promoter. In certain preferred embodiments, the promoter is an inducible promoter, such as the *cbh1* inducible promoter.

[00182] As disclosed herein, N-acetylglucosaminyltransferase I (GlcNAc-TI; GnTI; EC 2.4.1.101) catalyzes the reaction UDP-N-acetyl-D-glucosamine + 3-(alpha-D-mannosyl)-beta-D-mannosyl-R <=> UDP + 3-(2-(N-acetyl-beta-D-glucosaminyl)-alpha-D-mannosyl)-beta-D-mannosyl-R, where R represents the remainder of the N-linked oligosaccharide in the glycan acceptor. An N-acetylglucosaminyltransferase I catalytic domain is any portion of an N-acetylglucosaminyltransferase I enzyme that is capable of catalyzing this reaction. Amino acid sequences for N-acetylglucosaminyltransferase I enzymes from various organisms are listed in SEQ ID NOs: 1-19. Additional GnTI enzymes are listed in the CAZy database in the glycosyltransferase family 13 (cazy.org/GT13_all). Enzymatically characterized species includes *A. thaliana* AAR78757.1 (US6 653 459), *C. elegans* AAD03023.1 (Chen S. et al J. Biol.Chem 1999;274(1):288-97), *D. melanogaster* AAF57454.1 (Sarkar & Schachter Biol Chem. 2001 Feb;382(2):209-17); *C. griseus* AAC52872.1 (Puthalakath H. et al J. Biol.Chem 1996 271(44):27818-22); *H. sapiens* AAA52563.1 (Kumar R. et al Proc Natl Acad Sci U

S A. 1990 Dec;87(24):9948-52); *M. auratus* AAD04130.1 (Opat As et al Biochem J. 1998 Dec 15;336 (Pt 3):593-8), (including an example of deactivating mutant), Rabbit, *O. cuniculus* AAA31493.1 (Sarkar M et al. Proc Natl Acad Sci U S A. 1991 Jan 1;88(1):234-8). Additional examples of characterized active enzymes can be found at cazy.org/GT13_characterized. The 3D structure of the catalytic domain of rabbit GnTI was defined by X-ray crystallography in Unligil UM et al. EMBO J. 2000 Oct 16;19(20):5269-80. The Protein Data Bank (PDB) structures for GnTI are 1FO8, 1FO9, 1FOA, 2AM3, 2AM4, 2AM5, and 2APC. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain is from the human N-acetylglucosaminyltransferase I enzyme (SEQ ID NO: 1), or variants thereof. In certain embodiments, the N-acetylglucosaminyltransferase I catalytic domain contains a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues 84-445 of SEQ ID NO: 1. In some embodiments, a shorter sequence can be used as a catalytic domain (e.g. amino acid residues 105-445 of the human enzyme or amino acid residues 107-447 of the rabbit enzyme; Sarkar et al. (1998) Glycoconjugate J 15:193-197). Additional sequences that can be used as the GnTI catalytic domain include amino acid residues from about amino acid 30 to 445 of the human enzyme or any C-terminal stem domain starting between amino acid residue 30 to 105 and continuing to about amino acid 445 of the human enzyme, or corresponding homologous sequence of another GnTI or a catalytically active variant or mutant thereof. The catalytic domain may include N-terminal parts of the enzyme such as all or part of the stem domain, the transmembrane domain, or the cytoplasmic domain.

[00183] As disclosed herein, N-acetylglucosaminyltransferase II (GlcNAc-TII; GnTII; EC 2.4.1.143) catalyzes the reaction UDP-N-acetyl-D-glucosamine + 6-(alpha-D-mannosyl)-beta-D-mannosyl-R <=> UDP + 6-(2-(N-acetyl-beta-D-glucosaminyl)-alpha-D-mannosyl)-beta-D-mannosyl-R, where R represents the remainder of the N-linked oligosaccharide in the glycan acceptor. An N-acetylglucosaminyltransferase II catalytic domain is any portion of an N-acetylglucosaminyltransferase II enzyme that is capable of catalyzing this reaction. Amino acid sequences for N-acetylglucosaminyltransferase II enzymes from various organisms are listed in SEQ ID NOs: 20-33. In certain embodiments, the N-acetylglucosaminyltransferase II catalytic domain is from the human N-acetylglucosaminyltransferase II enzyme (SEQ ID NO: 20), or variants thereof.

Additional GnTII species are listed in the CAZy database in the glycosyltransferase family 16 (cazy.org/GT16_all). Enzymatically characterized species include GnTII of *C. elegans*, *D. melanogaster, Homo sapiens, Rattus norvegigus, Sus scrofa* (cazy.org/GT16_characterized). In certain embodiments, the N-acetylglucosaminyltransferase II catalytic domain contains a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues from about 30 to about 447 of SEQ ID NO: 21. The catalytic domain may include N-terminal parts of the enzyme such as all or part of the stem domain, the transmembrane domain, or the cytoplasmic domain.

[00184] In embodiments where the filamentous fungal cell contains a fusion protein of the invention, the fusion protein may further contain a spacer in between the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain. Any of the spacers of the invention as described in the section entitled "Spacers" may be used. In certain preferred embodiments, the spacer is an EGIV spacer, a 2xG4S spacer, a 3xG4S spacer, or a CBHI spacer. In other embodiments, the spacer contains a sequence from a stem domain.

[00185] For ER/Golgi expression the N-acetylglucosaminyltransferase I and/or N-acetylglucosaminyltransferase II catalytic domain is typically fused with a targeting peptide or a part of an ER or early Golgi protein, or expressed with an endogenous ER targeting structures of an animal or plant N-acetylglucosaminyltransferase enzyme. In certain preferred embodiments, the N-acetylglucosaminyltransferase I and/or N-acetylglucosaminyltransferase II catalytic domain contains any of the targeting peptides of the invention as described in the section entitled "Targeting peptides." Preferably, the targeting peptide is linked to the N-terminal end of the catalytic domain. In some embodiments, the targeting peptide contains any of the stem domains of the invention as described in the section entitled "Targeting peptides." In certain preferred embodiments, the targeting peptide is a Kre2 targeting peptide. In other embodiments, the targeting peptide further contains a transmembrane domain linked to the N-terminal end of the stem domain or a cytoplasmic domain linked to the N-terminal end of the stem domain. In embodiments where the targeting peptide further contains a transmembrane domain,

the targeting peptide may further contain a cytoplasmic domain linked to the N-terminal end of the transmembrane domain.

[00186] The level of expression of an *alg3* gene of the invention may be reduced by any suitable method known in the art, including, without limitation, mutating the *alg3* gene. The *alg3* may be mutated by, for example, point mutations or deletion of the entire *alg3* gene. Preferably, the function of the alg3 protein is reduced or eliminated by the mutation of *alg3*. The *alg3* gene encodes a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl alpha-1,3-mannosyltransferase.. As disclosed herein, a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase of the invention transfers an alpha-D-mannosyl residue from dolichyl-phosphate D-mannose into a membrane lipid-linked oligosaccharide.

[00187] In certain embodiments, the filamentous fungal cell may contain a polynucleotide encoding a UDP-GlcNAc transporter. The polynucleotide encoding the UDP-GlcNAc transporter may be endogenous (i.e., naturally present) in the filamentous fungal cell, or it may be heterologous to the filamentous fungal cell.

[00188] In other embodiments, the filamentous fungal cell may also contain a polynucleotide encoding a α-1,2-mannosidase of the invention as described in the section entitled "Host Cells." The polynucleotide encoding the α-1,2-mannosidase may be endogenous in the filamentous fungal cell, or it may be heterologous to the filamentous fungal cell. These polynucleotides are especially useful for a filamentous fungal cell expressing high-mannose glycans transferred from the Golgi to the ER without effective exo-α-2-mannosidase cleavage. For cytoplasmic expression the catalytic domain of the mannosidase is typically fused with a targeting peptide, such as HDEL, KDEL, or part of an ER or early Golgi protein, or expressed with an endogenous ER targeting structures of an animal or plant mannosidase I enzyme.

[00189] In further embodiments, the filamentous fungal cell may also contain a polynucleotide encoding a galactosyltransferase of the invention as described in the section entitled "Host Cells." Galactosyltransferases transfer β-linked galactosyl residues to terminal N-acetylglucosaminyl residue. In certain embodiments the galactosyltransferase is a β-4-galactosyltransferase. The galactosyltransferase may be expressed in the cytoplasm of the filamentous fungal. A heterologous targeting peptide,

such as a Kre2 peptide described in Schwientek J.Biol. Chem 1996 3398, may be used. Promoters that may be used for expression of the galactosyltransferase include constitutive promoters such as *gpd,* promoters of endogenous glycosylation enzymes and glycosyltransferases such as mannosyltransferases that synthesize N-glycans in the Golgi or ER, and inducible promoters of high-yield endogenous proteins such as the *cbh1* promoter. In embodiments of the invention where the host cell contains a polynucleotide encoding a galactosyltransferase, the host cell also contains a polynucleotide encoding a UDP-Gal and/or UDP-Gal transporter. In certain embodiments of the invention where the filamentous fungal cell contains a polynucleotide encoding a galactosyltransferase, lactose may be used as the carbon source instead of glucose when culturing the filamentous fungal cell. The culture medium may be between pH 4.5 and 7.0 or between 5.0 and 6.5. In certain embodiments of the invention where the filamentous fungal cell contains a polynucleotide encoding a galactosyltransferase and a polynucleotide encoding a UDP-Gal and/or UDP-Gal transporter, a divalent cation such as Mn2+, Ca2+ or Mg2+ may be added to the cell culture medium.

[00190] In other embodiments, the filamentous fungal cell may also contain a polynucleotide encoding a sialyltransferase of the invention as described in the section entitled "Host Cells.". A sialyltransferase transfers α3- or α6-linked sialic acid, such as Neu5Ac, to the terminal Gal of galactosylated complex glycans. The polynucleotide encoding the α3- or α6-linked sialyltransferase may be endogenous to the filamentous fungal cell, or it may be heterologous to the filamentous fungal cell. Sialylation in the filamentous fungal cell may require expression of enzymes synthesizing the donor CMP-sialic acid such as CMP-Neu5Ac, especially in fungal, plant, nematode/parasite, or insect cells.

[00191] Additionally, the filamentous fungal cell may have increased or reduced levels of activity of various additional endogenous enzymes. A reduced level of activity may be provided by inhibiting the activity of the endogenous enzyme with an inhibitor, an antibody, or the like. In certain embodiments, the filamentous fungal cell is genetically modified in ways to increase or reduce activity of one or more endogenous enzymes. Methods of genetically modifying a filamentous fungal cell to increase or reduce activity of one or more endogenous enzymes are well known in the art and include, without limitation, those described in the section entitled "Host Cells." In certain embodiments,

the filamentous fungal cell has a reduced level of activity of a alpha-1,6-mannosyltransferase compared to the level of activity in a wild-type filamentous fungal cell. Alpha-1,6-mannosyltransferase (EC 2.4.1.232) in the Golgi apparatus transfers an elongation initiating alpha-D-mannosyl residue from GDP-mannose into a protein-linked N-glycan oligosaccharide, forming an alpha-(1->6)-D-mannosyl-D-mannose linkage. Typically, the alpha-1,6-mannosyltransferase enzyme is encoded by an *och1* gene. In certain embodiments, the filamentous fungal cell has a reduced level of expression of an *och1* gene compared to the level of expression in a wild-type filamentous fungal cell. In certain embodiments, the *och1* gene is deleted from the filamentous fungal cell.

[00192] The filamentous fungal cell may be, for example, an *Acremonium, Aspergillus, Aureobasidium, Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium, Fusarium, Gibberella, Humicola, Magnaporthe, Mucor, Myceliophthora, Myrothecium, Neocallimastix, Neurospora, Paecilomyces, Penicillium, Piromyces, Schizophyllum, Scytalidium, Talaromyces, Thermoascus, Thielavia, Tolypocladium*, or *Trichoderma* cell. In certain embodiments, the filamentous fungal cell is a *T. reesei* cell.

Pharmaceutical Compositions Containing Complex N-Glycans Produced by the Methods of the Invention

[00193] In another aspect, the present invention provides a composition, e.g., a pharmaceutical composition, containing one or more complex N-glycans attached to a heterologous molecule produced by the methods of the invention, formulated together with a pharmaceutically acceptable carrier. Pharmaceutical compositions of the invention also can be administered in combination therapy, i.e., combined with other agents. For example, the combination therapy can include an complex N-glycans attached to a heterologous molecule according to the present invention combined with at least one other therapeutic agent.

[00194] As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible. Preferably, the carrier is suitable for intravenous, intramuscular, subcutaneous, parenteral, spinal or epidermal administration (e.g., by injection or infusion). Depending on the route of administration, the active compound, i.e., the complex N-glycan attached to a

heterologous molecule according to the invention, may be coated in a material to protect the compound from the action of acids and other natural conditions that may inactivate the compound.

[00195] The pharmaceutical compositions of the invention may include one or more pharmaceutically acceptable salts. A "pharmaceutically acceptable salt" refers to a salt that retains the desired biological activity of the parent compound and does not impart any undesired toxicological effects (see e.g., Berge, S.M., et al. (1977) J. Pharm. Sci. 66:1-19). Examples of such salts include acid addition salts and base addition salts. Acid addition salts include those derived from nontoxic inorganic acids, such as hydrochloric, nitric, phosphoric, sulfuric, hydrobromic, hydroiodic, phosphorous and the like, as well as from nontoxic organic acids such as aliphatic mono- and dicarboxylic acids, phenyl-substituted alkanoic acids, hydroxy alkanoic acids, aromatic acids, aliphatic and aromatic sulfonic acids and the like. Base addition salts include those derived from alkaline earth metals, such as sodium, potassium, magnesium, calcium and the like, as well as from nontoxic organic amines, such as N,N'-dibenzylethylenediamine, N-methylglucamine, chloroprocaine, choline, diethanolamine, ethylenediamine, procaine and the like.

[00196] A pharmaceutical composition of the invention also may include a pharmaceutically acceptable antioxidant. Examples of pharmaceutically acceptable antioxidants include: (1) water soluble antioxidants, such as ascorbic acid, cysteine hydrochloride, sodium bisulfate, sodium metabisulfite, sodium sulfite and the like; (2) oil-soluble antioxidants, such as ascorbyl palmitate, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), lecithin, propyl gallate, alpha-tocopherol, and the like; and (3) metal chelating agents, such as citric acid, ethylenediamine tetraacetic acid (EDTA), sorbitol, tartaric acid, phosphoric acid, and the like.

[00197] Examples of suitable aqueous and nonaqueous carriers that may be employed in the pharmaceutical compositions of the invention include water, ethanol, polyols (such as glycerol, propylene glycol, polyethylene glycol, and the like), and suitable mixtures thereof, vegetable oils, such as olive oil, and injectable organic esters, such as ethyl oleate. Proper fluidity can be maintained, for example, by the use of coating materials, such as lecithin, by the maintenance of the required particle size in the case of dispersions, and by the use of surfactants.

[00198] These compositions may also contain adjuvants such as preservatives, wetting agents, emulsifying agents and dispersing agents. Prevention of presence of microorganisms may be ensured both by sterilization procedures, and by the inclusion of various antibacterial and antifungal agents, for example, paraben, chlorobutanol, phenol sorbic acid, and the like. It may also be desirable to include isotonic agents, such as sugars, sodium chloride, and the like into the compositions. In addition, prolonged absorption of the injectable pharmaceutical form may be brought about by the inclusion of agents which delay absorption such as aluminum monostearate and gelatin.

[00199] Pharmaceutically acceptable carriers include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. The use of such media and agents for pharmaceutically active substances is known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the pharmaceutical compositions of the invention is contemplated. Supplementary active compounds can also be incorporated into the compositions.

[00200] Therapeutic compositions typically must be sterile and stable under the conditions of manufacture and storage. The composition can be formulated as a solution, microemulsion, liposome, or other ordered structure suitable to high drug concentration. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[00201] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by sterilization microfiltration. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from

those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the certain methods of preparation are vacuum drying and freeze-drying (lyophilization) that yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00202] The amount of active ingredient which can be combined with a carrier material to produce a single dosage form will vary depending upon the subject being treated, and the particular mode of administration. The amount of active ingredient which can be combined with a carrier material to produce a single dosage form will generally be that amount of the composition which produces a therapeutic effect. Generally, out of one hundred percent, this amount will range from about 0.01 percent to about ninety-nine percent of active ingredient, preferably from about 0.1 percent to about 70 percent, most preferably from about 1 percent to about 30 percent of active ingredient in combination with a pharmaceutically acceptable carrier.

[00203] Dosage regimens are adjusted to provide the optimum desired response (e.g., a therapeutic response). For example, a single bolus may be administered, several divided doses may be administered over time or the dose may be proportionally reduced or increased as indicated by the exigencies of the therapeutic situation. It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subjects to be treated; each unit contains a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active compound for the treatment of sensitivity in individuals.

[00204] For administration of the complex N-glycan attached to a heterologous molecule, in particular where the heterologous molecule is an antibody, the dosage ranges from about 0.0001 to 100 mg/kg, and more usually 0.01 to 5 mg/kg, of the host body weight. For example, dosages can be 0.3 mg/kg body weight, 1 mg/kg body weight, 3 mg/kg body weight, 5 mg/kg body weight or 10 mg/kg body weight or within the range of 1-10 mg/kg. An exemplary treatment regime entails administration once per week, once

every two weeks, once every three weeks, once every four weeks, once a month, once every 3 months or once every three to 6 months. Certain dosage regimens for a complex N-glycan attached to a heterologous antibody include 1 mg/kg body weight or 3 mg/kg body weight via intravenous administration, with the antibody being given using one of the following dosing schedules: (i) every four weeks for six dosages, then every three months; (ii) every three weeks; (iii) 3 mg/kg body weight once followed by 1 mg/kg body weight every three weeks.

[00205] Alternatively a complex N-glycan attached to a heterologous molecule according to the invention can be administered as a sustained release formulation, in which case less frequent administration is required. Dosage and frequency vary depending on the half-life of the administered substance in the patient. In general, human antibodies show the longest half life, followed by humanized antibodies, chimeric antibodies, and nonhuman antibodies. The dosage and frequency of administration can vary depending on whether the treatment is prophylactic or therapeutic. In prophylactic applications, a relatively low dosage is administered at relatively infrequent intervals over a long period of time. Some patients continue to receive treatment for the rest of their lives. In therapeutic applications, a relatively high dosage at relatively short intervals is sometimes required until progression of the disease is reduced or terminated, and preferably until the patient shows partial or complete amelioration of symptoms of disease. Thereafter, the patient can be administered a prophylactic regime.

[00206] Actual dosage levels of the active ingredients in the pharmaceutical compositions of the present invention may be varied so as to obtain an amount of the active ingredient which is effective to achieve the desired therapeutic response for a particular patient, composition, and mode of administration, without being toxic to the patient. The selected dosage level will depend upon a variety of pharmacokinetic factors including the activity of the particular compositions of the present invention employed, or the ester, salt or amide thereof, the route of administration, the time of administration, the rate of excretion of the particular compound being employed, the duration of the treatment, other drugs, compounds and/or materials used in combination with the particular compositions employed, the age, sex, weight, condition, general health and prior medical history of the patient being treated, and like factors well known in the medical arts.

[00207] A "therapeutically effective dosage" of immunoglobulin of the invention preferably results in a decrease in severity of disease symptoms, an increase in frequency and duration of disease symptom-free periods, or a prevention of impairment or disability due to the disease affliction. For example, for the treatment of tumors, a "therapeutically effective dosage" preferably inhibits cell growth or tumor growth by at least about 20%, more preferably by at least about 40%, even more preferably by at least about 60%, and still more preferably by at least about 80% relative to untreated subjects. The ability of a compound to inhibit tumor growth can be evaluated in an animal model system predictive of efficacy in human tumors. Alternatively, this property of a composition can be evaluated by examining the ability of the compound to inhibit, such inhibition *in vitro* by assays known to the skilled practitioner. A therapeutically effective amount of a therapeutic compound can decrease tumor size, or otherwise ameliorate symptoms in a subject. One of ordinary skill in the art would be able to determine such amounts based on such factors as the subject's size, the severity of the subject's symptoms, and the particular composition or route of administration selected.

[00208] A composition of the present invention can be administered via one or more routes of administration using one or more of a variety of methods known in the art. As will be appreciated by the skilled artisan, the route and/or mode of administration will vary depending upon the desired results. Certain routes of administration for binding moieties of the invention include intravenous, intramuscular, intradermal, intraperitoneal, subcutaneous, spinal or other parenteral routes of administration, for example by injection or infusion. The phrase "parenteral administration" as used herein means modes of administration other than enteral and topical administration, usually by injection, and includes, without limitation, intravenous, intramuscular, intraarterial, intrathecal, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural and intrasternal injection and infusion.

[00209] Alternatively, a complex N-glycan attached to a heterologous molecule according to the invention can be administered via a nonparenteral route, such as a topical, epidermal or mucosal route of administration, for example, intranasally, orally, vaginally, rectally, sublingually or topically.

[00210] The active compounds can be prepared with carriers that will protect the compound against rapid release, such as a controlled release formulation, including implants, transdermal patches, and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Many methods for the preparation of such formulations are patented or generally known to those skilled in the art. See, e.g., Sustained and Controlled Release Drug Delivery Systems, J.R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

[00211] Therapeutic compositions can be administered with medical devices known in the art. For example, in a certain embodiment, a therapeutic composition of the invention can be administered with a needleless hypodermic injection device, such as the devices disclosed in U.S. Patent Nos. 5,399,163; 5,383,851; 5,312,335; 5,064,413; 4,941,880; 4,790,824; or 4,596,556. Examples of well-known implants and modules useful in the present invention include: U.S. Patent No. 4,487,603, which discloses an implantable micro-infusion pump for dispensing medication at a controlled rate; U.S. Patent No. 4,486,194, which discloses a therapeutic device for administering medicants through the skin; U.S. Patent No. 4,447,233, which discloses a medication infusion pump for delivering medication at a precise infusion rate; U.S. Patent No. 4,447,224, which discloses a variable flow implantable infusion apparatus for continuous drug delivery; U.S. Patent No. 4,439,196, which discloses an osmotic drug delivery system having multi-chamber compartments; and U.S. Patent No. 4,475,196, which discloses an osmotic drug delivery system.

[00212] In certain embodiments, the use of the complex N-glycan attached to a heterologous molecule according to the invention is for the treatment of any disease that may be treated with therapeutic antibodies.

[00213] It is to be understood that, while the invention has been described in conjunction with the certain specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

[00214] The invention having been described, the following examples are offered to illustrate the subject invention by way of illustration, not by way of limitation.

EXAMPLES

Example 1 – Host Strain Selection for Glycoengineering

[00215] The aim of this example was to identify optimal *T. reesei* strains for glycoengineering. An optimal strain produces high amounts of Man5 N-glycans and low amounts of acidic glycans.

Samples

[00216] Different *T. reesei* strains including M44 (VTT-D-00775; Selinheimo et al., *FEBS J.* 2006, 273(18): 4322-35), M81, M84, M109, M110, M131, M132, M133, M134 and M124 (a mus53-deleted strain of M44) were analyzed. Each of the ten strains was grown in shake flask cultures. Samples were taken at three different time points: 3 days, 5 days, and 7 days. Both supernatants (secreted proteins) and cell pellets were collected and stored frozen at -20°C until glycan analysis was conducted.

[00217] N-glycans were isolated from secreted proteins from the indicated time points followed by matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) glycan profiling. Cell pellets from the 5 days time point were subjected to N-glycan profiling. A total of 80 samples (30 each of neutral- and acidic supernatant fractions, and 10 each of neutral- and acidic pellet fractions) were subjected to analysis.

[00218] Strain M44 was also subjected to batch and fed-batch fermentor cultivation in order to assess the difference on glycan profile between shake flask and fermentor culture. For glycan analysis, samples from three different time points were analyzed for a total of 12 samples (6 neutral and 6 acidic fractions). As a control, culture medium was analyzed.

Mass Spectrometry Methods

[00219] MALDI-TOF mass spectrometry was performed with a Bruker Ultraflex TOF/TOF instrument (Bruker Daltonics, Germany). Neutral N-glycans were detected in positive ion reflector mode as $[M + Na]^+$ ions, and acidic N-glycans were detected in

negative ion linear mode as $[M-H]^-$ ions. The relative molar abundance of neutral N-glycan components was assigned based on their relative signal intensities in the spectra. The resulting glycan signals in the presented glycan profiles were normalized to 100% to allow comparison between samples.

Protein-Specific Glycosylation Methods

[00220] Proteins from a fermentor-cultured sample were separated with SDS-PAGE and blotted to a PVDF membrane. The protein bands of interest were excised, and N-glycans were liberated by enzymatic release with PNGase F.

Neutral N-glycan Profile of T. reesei Strains

[00221] The desired Man5 structure can be observed as a $[M+Na]^+$ signal at m/z value of 1257.4 in the mass spectra presented in Figure 1. The neutral glycome of the analyzed *T. reesei* strains were found to have either Man5 or Man8 as the main neutral glycan species (H5N2 and H8N2 in Table 2).

Table 2: The percentage of different neutral N-glycan signals of analyzed *T. reesei* strains.

		Strain									
		M44	M81	M84	M109	M110	M131	M132	M133	M134	M124
Composition	m/z	%	%	%	%	%	%	%	%	%	%
H3N2	933	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
H4N2	1095	1.2	0.0	0.0	1.8	0.9	0.0	2.3	0.0	2.3	4.1
H5N2	1257	81.0	70.8	4.0	78.9	5.8	78.8	84.1	10.7	73.2	77.9
H6N2	1419	5.8	5.3	0.0	5.3	0.9	4.8	4.6	0.9	6.0	7.3
H7N2	1581	4.8	7.3	1.5	4.7	3.0	4.8	3.9	3.8	5.8	4.8
H8N2	1743	3.7	8.6	81.5	5.1	68.2	5.9	2.6	68.1	6.3	3.3
H9N2	1905	2.9	8.0	9.0	3.4	16.0	4.6	2.0	12.8	5.7	2.3
H10N2	2067	0.5	0.0	2.5	0.8	3.7	1.1	0.4	2.5	0.7	0.4
H11N2	2229	0.0	0.0	1.5	0.0	1.4	0.0	0.0	1.2	0.0	0.0
H12N2	2391	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[00222] Some acidic N-glycans were observed in neutral N-glycan fractions. This may have been due to specific properties of the phosphorylated glycans, *e.g.* presence of phosphodiester structures, or other properties of the phosphoglycans which could lead to leakage of acidic species to neutral fraction under the experimental conditions used in this study. To check the corresponding structure, the signal of interest was subjected to MS/MS analysis. Mass spectrometric fragmentation of glycans was performed using Bruker Ultraflex TOF/TOF in MS/MS analysis mode (Figure 2). Because the glycans

were not permethylated, definitive structural assignment based on the MS/MS data could not be obtained.

Acidic N-glycan Profiles of T. reesei Strains

[00223] For glycoengineering purposes it was useful to have strains with a minimum amount of acidic N-glycans. Therefore, acidic N-glycan profiles were analyzed from the strains used for screening. The acidic N-glycan spectra of analyzed strains are shown in Figure 3 and below in Table 3.

Table 3: The percentage of different acidic N-glycan signals of analyzed *T. reesei* strains.

		M44	M81	M84	M109	M110	M131	M132	M133	M134	M124
	m/z	%	%	%	%	%	%	%	%	%	%
Hex3HexNAc2SP	989	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex4HexNAc2SP	1151	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex5HexNAc2SP	1313	4.0	5.2	0.0	3.7	0.0	2.8	7.4	0.0	5.2	2.8
Hex5HexNAc2SP2	1393	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.9	0.0
Hex6HexNAc2SP	1475	23.7	27.3	2.2	18.1	2.1	22.4	21.0	3.9	24.9	26.3
Hex6HexNAc2SP2	1555	0.0	2.8	0.0	2.4	0.0	3.2	1.1	0.0	3.6	1.7
Hex7HexNAc2SP	1637	30.3	18.8	1.1	16.2	2.0	14.9	24.7	0.0	17.2	23.3
Hex7HexNAc2SP2	1717	0.0	7.7	0.0	8.6	0.0	10.7	2.5	0.0	10.4	7.0
Hex8HexNAc2SP	1799	18.4	11.8	17.9	12.8	9.7	9.1	19.7	14.5	8.8	11.2
Hex8HexNAc2SP2	1879	5.1	8.8	0.0	11.0	0.0	14.8	4.0	0.0	12.4	10.0
Hex9HexNAc2SP	1961	7.3	6.4	49.1	9.5	37.9	5.9	6.1	53.9	4.1	3.5
Hex9HexNAc2SP2	2041	4.2	5.0	0.0	5.7	0.0	7.3	5.1	0.0	5.9	7.2
Hex10HexNAc2SP	2123	2.8	2.9	19.7	4.5	28.1	2.6	2.3	19.3	2.1	1.6
Hex10HexNAc2SP2	2203	2.8	2.1	0.0	2.2	0.0	2.7	3.6	0.0	1.9	3.3
Hex11HexNAc2SP	2285	1.5	1.3	3.7	2.1	9.5	1.2	0.9	5.0	1.0	0.8
Hex11HexNAc2SP2	2365	0.0	0.0	0.0	0.9	0.0	1.3	1.5	0.0	0.8	1.3
Hex12HexNAc2SP	2447	0.0	0.0	1.3	1.0	1.6	1.0	0.0	0.0	0.5	0.0
Hex12HexNAc2SP2	2527	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.3	0.0
Hex13HexNAc2SP	2609	0.0	0.0	1.2	0.4	1.1	0.0	0.0	0.0	0.0	0.0
Hex14HexNAc2SP	2771	0.0	0.0	0.6	0.0	0.9	0.0	0.0	0.0	0.0	0.0

N-glycan Profile from Fermentor Cultured Strain M44

[00224] Strain M44 was cultivated in a fermentor in order to find out if different culture conditions can cause changes in its glycan profile. N-glycan analysis was performed for samples cultured in a fermentor (Batch; 41:10, 88:45 and 112:50 hours, and Fed batch; 45:50, 131:40 and 217:20 hours) and compared to that of shake flask culture. Neutral and acidic N-glycans of secreted proteins of *T. reesei* strain M44 cultured in fermentor are shown in Figure 4. Comparison between the N-glycan percentages from flask and fermentor cultures is presented below in Table 4.

Table 4: The percentage of N-glycan signals of *T. reesei* strain M44 cultured in flask and in fermentor.

Composition	m/z	flask %	fermentor %
H3N2	933	0.0	0.0
H4N2	1095	1.2	0.0
H5N2	1257	81.0	91.3
H6N2	1419	5.8	4.5
H7N2	1581	4.8	4.2
H8N2	1743	3.7	0.0
H9N2	1905	2.9	0.0
H10N2	2067	0.5	0.0
H11N2	2229	0.0	0.0
H12N2	2391	0.0	0.0

N-glycan Analysis of Shake Flask Culture Medium

[00225] As a control experiment, culture medium (without contact with fungus) of *T. reesei* was analyzed. Figure 5a shows neutral N-glycan analysis in which no N-glycans were observed. Only minor signals of hexose oligomers, most likely derived from the plant material used in the medium, were visible above the baseline. In Figure 5b (acidic glycans), no signals corresponding to N-glycans were observed.

N-glycosylation of Secreted Proteins

[00226] To check whether there is variation in glycosylation between individual secreted proteins, the samples from fermentation culture supernatants were separated with SDS-PAGE and blotted to PVDF membrane. The N-glycans of selected bands were then detached with on-membrane enzymatic release. Results are shown in Figures 6 and 7.

Conclusions: Neutral Glycans

[00227] The purpose of this study was to identify *T. reesei* strains for glycoengineering with the highest amount of Man5 N-glycans and the lowest amount of acidic glycans. Strains which have Man5 as a main peak in mass spectrometry analysis can have higher endogenous α-1,2-mannosidase activity. Based on the background information on *T. reesei* N-glycosylation, the likely structure for Man5 is Manα3[Manα3(Manα6)Manα6]Manβ4GlcNAcβ4GlcNAc (Salovuori *et al.* 1987; Stals *et al. Glycobiology* 14, 2004, page 725).

[00228] Some strains contained H8N2 as a major neutral glycoform. Based on the literature, this glycoform is most likely to be a Glcα3Manα2Manα2Man5 structure (Stals *et al. Glycobiology* 14, 2004, page 725). It is possible that glucosidase deficiency in these strains prevents the trimming of the glycans to the smaller glycoforms.

[00229] In some strains, acidic N-glycans were observed in neutral spectra. This situation may have been due to a higher proportion of acidic N-glycans or to leakage of specific structures into the neutral fraction during the separation of neutral glycans from acidic glycans.

[00230] The glycan profile of strains was a bit more favorable for glycoengineering when cultivated in a fermentor compared to in shake flasks. The glycosylation of individual proteins from fermentor-cultured samples didn't differ significantly from average glycosylation. All analyzed proteins contained Man5 as a main glycoform. This observation suggested that all secreted proteins go through similar glycan processing. Thus it appeared that the majority of secreted proteins were glycosylated similarly by the *T. reesei* host cells, which is not always the case with mammalian cells.

Acidic Glycans

[00231] The phosphorylation of N-glycan is not generally desired for glycoengineering because the terminal phosphate residue is not present in regular therapeutic proteins, including antibodies. Some exceptions to this rule are a few specialized proteins used for lysosomal glycosylation storage disorders. Phosphorylation of N-glycans may be protein-specific in fungi. In animals, mannose phosphorylation is a conserved lysosomal targeting signal.

[00232] To date there have been no reports of sulfation of *T. reesei* N-glycans. Therefore, the acidic structures referred to in this report were likely to be phosphorylated glycans.

[00233] Phosphorylation is more common when *T. reesei* is cultivated in low pH values, as is the case in flask cultures, which may be related to low pH stress and mycelia breakage (Stals *et al.*, 2004, *Glycobiology* 14:713-724). In this study a clear difference was observed between flask and fermentor cultured samples. Acidic N-glycans, all phosphorylated, were observed in shake flask culture samples. The amount of acidic N-

glycans in fermentor samples may have been below the detection limit, or, because of higher pH there may have been no significant phosphorylation of glycans. The proportion of acidic N-glycans to the total amount of N-glycans could not be verified with the method used in this study due to the different ionization efficiencies between neutral and acidic glycan species.

[00234] In order to determine phosphorylation levels, N-glycans were released by N-glycanase from 10 µg of *T. reesei* secreted protein cultured in batch and fed batch fermentor. Protein concentration was measured using a Bradford-based method with BSA as a standard. One pmol of standard molecule NeuAcHex4HexNAc2 was added to acidic N-glycans samples prior to MALDI-TOF analysis. Amounts of major glycoforms (Hex7HexNAc2P for fermentor and Hex6-8HexNAc2P for flask culture) were 0.9 pmol/10 µg of secreted protein of batch culture, 0.6 pmol/10 µg of secreted protein of fed batch culture, and 160 pmol/10 µg of secreted protein of flask culture when the pH of the culture was allowed to drop. The amount of neutral N-glycans was measured using 10 pmol of standard glycan Hex2HexNAc4 added to neutral N-glycan samples, prior to MALDI-TOF analysis. The amount of major glycoform Hex5HexNAc2 was 87 pmol/10 µg of secreted protein in batch and fed-batch cultures and 145 pmol/10 µg of secreted protein in flask culture. Thus, the proportion of acidic N-glycans to total amount of N-glycans was 1% in batch culture, 0.7% in fed-batch culture and 52% in flask culture. Quantitation was based only on signal intensity comparison using MALDI-TOF data.

[00235] N-glycans were also larger in acidic fraction. This may have been due to phospho-mannosylation reactions in which phosphorus with one hexose unit is attached to a glycan backbone. Some diphosphorylated structures were seen in acidic spectra. This explanation is in agreement with the previously published data on phosphorylated glycans found in *T. reesei* (Stals *et al.* 2004, Glycobiology 14:725-737). When cultured in a fermentor, the proportion of acidic N-glycans was very low, below the detection limit.

[00236] The N-glycan spectra of *T. reesei* culture media did not reveal contamination of the *T. reesei* N-glycome with glycans derived from plant material containing medium.

[00237] In conclusion, N-glycan analysis of different *T. reesei* strains revealed that the major glycoform in strains M44, M109, M131, M132 and M124 is Man5 or Manα3[Manα3(Manα6)Manα6]Manβ4GlcNAcβ4GlcNAc. The possible presence of

glucose, including H8N2 as a minor component in Man5-producing strains was considered. Two strains (M109 and M131) contained a larger amount of H8N2 than H7N2. The enrichment of H8N2 could have indicated partial glucosidase deficiency.

[00238] Strain M44 contained almost no phosphorylated glycans. Leaking acidic glycans observed in neutral glycan fraction as signals at m/z 1521 and m/z 1683 were observed in samples from strains M131, M109, M132 and M124, which indicated higher phosphorylation levels and the presence of potential phosphodiester structures.

[00239] The aim of this study was to find a strain with maximal production of Man5Gn2 structure and low-level production of acidic (phosphorylated) N-glycans. The best strains had over 80% of Man5 under pH-controlled shake flask culture conditions. The best strains also had reduced production of di-phosphorylated glycans and/or larger phosphorylated structures (see Table 3).

Example 2 – Generation of an Alg3-Deficient Trichoderma Strain

Vector Construction and Strain Generation

[00240] The gene encoding the ALG3 mannosyltransferase was identified in the *Trichoderma reesei* genome sequence. A disruption construct was designed to insert the acetamidase selection marker between 1000 bp 5' and 3' flanking region fragments of the *alg3* gene. The flanking region fragments were amplified by PCR, and the construct was made by homologous recombination cloning in *Saccharomyces cerevisiae*. The disruption cassette was released from its backbone vector by digestion and transformed into the *T. reesei* strain M124. Transformants were selected on acetamidase medium and screened by PCR with a forward primer outside the 5' flanking region fragment of the construct and the reverse primer inside the AmdS selection marker.

Screening of Transformants

[00241] Fifty-eight out of 62 screened transformants gave a PCR product of the size expected for integration of the construct to the *alg3* locus. Nine PCR-positive transformants were purified to uninuclear clones through single spore cultures, and spore suspensions were made from them. These nine clones were analyzed for the correct integration of the disruption cassette by Southern hybridization. EcoRI-digested genomic

DNA from the parental strain and from nine clones was hybridized with an *alg3* probe under standard hybridization conditions. The probe hybridized with DNA from the parental strain, but not with DNA from any of the clones, indicating successful deletion of *alg3* (Figure 8).

[00242] Further analysis was made by Southern hybridization with an AmdS probe. The AmdS gene was included in the deletion cassette and was predicted to be detectable in DNA from the transformants, but not in DNA from the parental strain. Genomic DNA of parental strain M124 and nine transformants was digested with EcoRI + PvuI (E + P) and KpnI + NheI (K + N). NotI digested plasmid carrying the *alg3*-AmdS deletion cassette was used as a positive control. The probe recognized the expected ~2.7 kb fragment (AmdS) from the positive control but did not hybridize with the parental strain. All transformants gave the expected signals (1.6 + 2.8 kb for E + P and 1.7 + 3.4 kb for K + N, shown with arrows in Figure 9B) indicating correct integration of the deletion cassette. Clones 11A and 15A also showed hybridization of some additional fragments suggesting unspecific integration of the deletion cassette to the genome (Figure 9B).

N-glycan Analytics

[00243] Shake-flask cultures of five different Alg3 knockout strains (4A, 5A, 6A, 10A and 16A) and parental strain M124 were analyzed for N-glycans. Samples were collected from time points of 3, 5, 7, and 9 days. All cultures were grown as duplicates.

[00244] The protein concentration of secreted proteins from a randomly selected knockout strain (4A) from all time points was measured using a Bradford-based assay against a BSA standard curve. The highest protein concentration was detected on day 5. Therefore, day 5 samples were used for N-glycan analysis for all five knockout strains. All samples, including the duplicate cultures, were analyzed as triplicates. Ten µg was used for N-glycan analysis. Both neutral and acidic N-glycans were analyzed by MALDI-TOF.

[00245] The major glycoform in parental strain M124 was Man5Gn2. In all Alg3 knockout strains the major glycoform was Man3 (Figure 10). No Man3 was found in the parental strain M124. In different Alg3 knockout strains the amount of Man3 ranged between 49.7% - 55.2% in the shake-flask cultures allowing pH drop. Hex6Gn2 was

increased in the parental strain. Signal intensities as percentages of observed neutral N-glycan signals are presented in Table 5 below.

Table 5: Neutral N-glycan content of Alg3 knockout strains.

		Strain											
		Parental M124		4A		5A		6A		10A		16A	
Composition	m\z	Average	STDEV	Average	STDEV	Average	STDEV	Average	STDEV	Average	STDEV	Average	STDEV
Hex3HexNAc2	933.31	0.0	0.0	53.6	0.2	55.2	4.2	49.7	0.5	53.3	0.9	53.4	0.9
Hex4HexNAc2	1095.37	1.6	0.1	2.7	0.0	2.9	0.7	3.4	0.1	3.2	0.4	3.4	0.4
Hex5HexNAc2	1257.42	70.2	3.3	8.5	0.2	7.3	1.1	10.4	0.5	8.6	0.9	9.7	0.9
Hex6HexNAc2	1419.48	7.9	1.1	35.0	0.3	34.4	1.9	36.1	0.6	34.9	0.5	33.2	0.7
Hex7HexNAc2	1581.53	7.8	0.6	0.3	0.4	0.3	0.4	0.3	0.4	0.0	0.0	0.3	0.4
Hex8HexNAc2	1743.58	5.9	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex9HexNAc2	1905.63	6.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex10HexNAc2	2067.69	0.7	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[00246] The presence of different isomers of each glycoform cannot be observed by MALDI MS analysis, so further tandem mass spectrometry studies were performed. First, the Man3 and Hex5Gn2 structures were investigated. For Man3 it was asked whether the Man3 structure is branched or linear. For this analysis, a sample containing both these structures was permethylated and analyzed with mass spectrometric fragmentation using the Bruker Ultraflex III TOF/TOF instrument according to the manufacturer's instructions (Figures 11 and 12).

[00247] Next, it was determined whether the hexose unit on the non-reducing end of the Hex6Gn2 structure is a mannose or a glucose. Alpha-mannosidase digestion was performed on all knockout strains and the parental strain (Figure 13). Jack bean mannosidase, which cleaves α-mannoses and leaves the β-mannose from backbone untouched, was used. The resulting structure was expected to be Man1Gn2.

[00248] Due to low molecular weight range effects in MALDI, the relative intensity of the Man1GlcNAc2 glycan may have been somewhat reduced, which explained a small increase in the relative amount of Hex6. After α-mannosidase digestion, Man3 and Man4 glycoforms disappeared. No Man2 structure was observed. However, Hex6 (m/z 1419) was not digested (Table 6) indicating that there was a glucose unit on the non-reducing end of the structure. Some non-digestible Hex5 was also present, likely produced by a weak reaction removing the sterically hindered Man6-branch of Hex6.

Table 6: Neutral N-glycans of Alg3 knockout strain 4A before (native) and after α-mannosidase digestion.

		4A	
		Native	a-Man'ase
Composition	m/z	Average	%
Hex1HexNAc2	609.21	0.0	53.2
Hex2HexNAc2	771.26	0.0	0.0
Hex3HexNAc2	933.31	47.5	0.0
Hex4HexNAc2	1095.37	3.8	0.0
Hex5HexNAc2	1257.42	11.7	5.0
Hex6HexNAc2	1419.48	36.8	41.0
Hex7HexNAc2	1581.53	0.2	0.8
Hex8HexNAc2	1743.58	0.0	0.0
Hex9HexNAc2	1905.63	0.0	0.0
Hex10HexNAc2	2067.69	0.0	0.0

[00249] For the final analysis of different structures found in the Alg3 knockout strains, a large-scale PNGase F digestion was performed to Alg3 knockout strain 4A. Two major glycans were purified with HPLC (Figure 14) and analyzed by NMR (Figure 15).

[00250] Based on the data presented in Figure 15A, the Hex3HexNAc2 species was unambiguously identified as Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc. The Manα3 and Manα6 H-1 units resonated at 5.105 and 4.914 ppm, respectively. The Manβ4 H-2 unit was observed at 4.245 ppm. This signal was very characteristic, due to the neighboring Manα 3-OH substitution. The N-acetyl group –CH3 signals of the core GlcNAc units were observed at 2.038 and 2.075. These values agreed well with those reported for this pentasaccharide in the Sugabase-database (www.boc.chem.uu.nl/sugabase/sugabase.html). Moreover, the proton-NMR spectrum was measured for a commercially produced Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc (Glycoseparations, Inc.) in identical experimental conditions, and nearly identical chemical shifts were obtained.

[00251] The NMR spectrum of the Hex6HexNAc2 component is shown in Figure 15B. The data implied that this component represents the octasaccharide Glcα1-3Manα1-2Manα1-2Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc. The presence of a glucose unit was evident from the 5.255 signal showing a typical αGlc 2.4 Hz coupling. All Man signals typically show < 1 Hz coupling due to the equatorial H-2 configuration. Small

differences were observed compared to the Sugabase data (Table 7), which may be ascribed to the different temperature used in the present NMR measurement (40°C vs. 26°C).

Table 7: Published NMR data of Glcα1-3Manα1-2Manα1-2Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc. Data was obtained from Sugabase (found at boc.chem.uu.nl/sugabase/sugabase).

```
                                    a-D-Manp-(1-6)+
                                                  |
                                 b-D-Manp-(1-4)-b-D-GlcpNAc-(1-4)-D-GlcNAc
                                                  |
a-D-Glcp-(1-3)-a-D-Manp-(1-2)-a-D-Manp-(1-2)-a-D-Manp-(1-3)+
```

Residue	Linkage	Proton	PPM	J	Hz
D-GlcNAc		H-1a	5.189		
		H-1b	4.694		
		H-2a	3.867		
		H-2b	3.692		
		NAc	2.038		
b-D-GlcpNAc	4	H-1	4.606		
		H-2	3.792		
		NAc	2.077		
b-D-Manp	4,4	H-1	4.773		
		H-2	4.237		
a-D-Manp	6,4,4	H-1	4.913		
		H-2	3.964		
a-D-Manp	3,4,4	H-1	5.346		
		H-2	4.080		
a-D-Manp	2,3,4,4	H-1	5.304		
		H-2	4.103		
a-D-Manp	2,2,3,4,4	H-1	5.038		
		H-2	4.224		
a-D-Glcp	3,2,2,3,4,4	H-1	5.247		
		H-2	3.544		

[00252] Finally, the N-glycan profiles of randomly selected knockout strain 4A were analyzed at different time points (days 3, 5, 7 and 9). The shake flask culture pH was 4.8 at the starting time point and 2.6 at the ending time point. Triplicate samples from every time point of duplicate cultures were analyzed. It was observed that in both duplicates, the relative amount of Man3Gn2 signal decreased as a function of growth time because of the reduction of pH. However, the amount of Hex6Gn2 signal increased as a function of growth time (Table 8).

Table 8: The percentages of signal intensities from observed neutral glycan signals of Alg3 4A knockout strain. Duplicate cultures (3A and 4A) from four different time points (days 3, 5, 7 and 9) were analyzed.

Alg3 knock out strain 4A (flask 3A)

		Day 3, 3A		Day 5, 3A		Day 7, 3A		Day 9, 3A	
Composition	m/z	average	stdev	average	stdev	average	stdev	average	stdev
Hex3HexNAc	730.24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex2HexNAc2	771.26	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hex3HexNAc2	933.31	61.7	3.7	61.3	0.8	61.1	1.9	52.7	7.7
Hex4HexNAc2	1095.37	2.6	0.2	2.5	0.1	2.1	0.4	3.7	1.0
Hex5HexNAc2	1257.42	4.3	0.6	6.5	0.4	5.7	0.6	6.4	1.0
Hex6HexNAc2	1419.48	31.4	3.5	29.8	0.4	31.1	1.6	37.2	5.7

Alg3 knock out strain 4A (flask 4A)

		Day 3, 4A		Day 5, 4A		Day 7, 4A		Day 9, 4A	
Composition	m/z	average	stdev	average	stdev	average	stdev	average	stdev
Hex3HexNAc	730.24	0.0	0.0	0.0	0.0	0.0	0.0	0.7	1.2
Hex2HexNAc2	771.26	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.5
Hex3HexNAc2	933.31	61.7	3.2	58.6	1.1	55.6	1.9	54.8	5.9
Hex4HexNAc2	1095.37	3.4	1.0	2.6	0.2	3.1	0.2	2.6	0.5
Hex5HexNAc2	1257.42	5.2	1.5	6.7	0.4	7.1	0.4	7.6	3.7
Hex6HexNAc2	1419.48	29.7	0.9	32.1	0.8	34.3	1.5	34.0	3.6

[00253] A difference between these two analyses *(*Tables 4 and 7*)* concerning the percentage of Man3 in clone 4A (Day 5) were noted. This difference may have been due to differences in the analyses procedures. Some lability of the heterogenous culture medium protein preparations was observed after freeze-thaw cycle(s), likely due to glycan and/or protein degradation, resulting in reduced amounts of larger glycans. Generation of the data in Table 5 included additional freeze thaw-cycles.

[00254] Acidic N-glycan fractions were also analyzed by MALDI (Figure 16). The abundance of different acidic compounds in parental strain M124 differed from all Alg3 knockout strains, among which the acidic fraction seemed to be very similar.

[00255] Three major glycans in the parental strain were H6N2P1, H7N2P1 and H8N2P1. In the Alg3 knockout the size shifted into smaller glycans: H5N2P1, H6N2P1 and H4N2P1. Additionally, diphosphorylated glycans were more abundant in the parental strain. This may have been due to a lack of a suitable substrate for the particular enzyme

that attaches phosphorylated mannose to a glycan. The phosphorylated mannose can be further elongated by other mannose residues. Phosphorylation was not substantially present in glycans of the parent M124 strain produced under fermentation conditions.

Comparison of Fermentor and Shake Flask Grown Samples

[00256] One Alg3 knockout strain (transformant 4A) was grown in batch fermentation on lactose and spent grain extract medium. The medium was 60 g/l lactose with 20 g/l spent grain extract with a volume of 7 liters (fermentor run bio01616) after inoculation. Other medium components were KH_2PO_4 and $(NH_4)_2SO_4$. Culture pH was controlled between 5.5 and 5.8. Biomass and culture supernatant samples were taken during the course of the run and stored at -20°C. Mycelial samples were also collected for possible RNA analysis and were frozen immediately in liquid nitrogen and transferred to -70°C. Samples collected from the whole course of these fermentations were analyzed for N-glycan composition. N-glycan analysis was carried out for fermentor run bio01616) and for the 5 days time point sample from the shake flask culture of transformant 4A (Figures 17 and 18). The main signal in the shake flask culture was Man3 (59%). In the fermentor culture, the main signal was Man3 (85%), and the proportion of Hex6 was decreased to 8%.

Conclusions

[00257] The Alg3 knockout was successful in producing 50% or more of the expected Man3 glycoform. The desired branched structure of Manα3(Manα6)Manβ- was verified by fragmentation mass spectrometry and NMR spectroscopy.

[00258] The other products of the Alg3 knockout included Man4 (mannose-containing minor product), Hex5 (a degradation product of Hex6 as indicated in Figure 13) and Hex6, which was the second largest component. The Hex6 component was characterized to contain terminal Glc by mannosidase resistance and specific NMR signals including Glcα3Man-terminal. It was considered that the glycan structure could be further optimized by methods for reducing the amount of the terminal Glc, which was likely causing suboptimal efficacy of glucosidase II with the glycan devoid of mannoses on the Manα6-arm of the molecule. Further optimization of fermentation conditions may reduce the amount of terminal Glc.

[00259] This data indicated better glycosylation results in the *T. reesei* Alg3 knockout compared to earlier data for Alg3 knockouts in *Aspergillus* (Kainz *et al.*, *Appl Environ Microb*. 2008 1076-86) and *P. pastoris* (Davidson *et al.*, *Glycobiology* 2004, 399-407). In the works of Kainz *et al.* and Davidson *et al.*, similar or higher Hex6 corresponding product levels were reported. Those studies also reported additional problems with α2-Mannose, OCH1 products and larger size, and cell type-specific glycans produced by *P. pastoris*. In conclusion, N-glycan analysis of *T. reesei* Alg3 knockouts revealed that the major glycoform in the knockout strains is Man3Gn2, a desired starting point for efficient generation of mammalian-type N-glycans.

Example 3 – Purification and Activity of Individual GnTI and GnTII Enzymes

[00260] Human GnTI and GnTII (N-acetylglucosaminyltransferase I and N-acetylglucosaminyltransferase II) were expressed as soluble, secreted proteins in *Pichia pastoris* in order to study their acceptor specificity and activity.

Generation of GnTI Construct for Production in P. Pastoris

[00261] Human GnTI (P26572) sequence was obtained as a full-length sequence and subcloned into *Trichoderma reesei* overexpression vectors. Protein coding sequences (CDS) encoding the soluble part of human GnTI were cloned to the pBLARG-SX expression vector in order to produce a secreted form of the protein in *Pichia pastoris* for enzymatic studies. During the cloning procedure, a His tag encoding sequence was added to 5′end of the frame to obtain a tag at the N-terminus of the truncated protein. The sequence was verified by sequencing analysis. Resulting vector pTTg5 was linearized and transformed by electroporation to *P. pastoris* GY190 cells to yield strain GY4. Arg^+ transformants were picked and screened by PCR. GY4 clones containing the integrated plasmid were tested for protein expression.

Expression and Purification of Soluble GnTI

[00262] *P. pastoris* strain GY4 expressing soluble GnTI was first grown overnight with shaking at +30°C in BMGY medium (1% yeast extract, 2% peptone, 100 mM potassium phosphate pH 6.0, 1.34% yeast nitrogen base, 4 x 10-5% biotin, 1% glycerol) to OD_{600} 2-6. The cells were then harvested by centrifugation and resuspended to OD_{600} of 1 in BMMY medium (like BMGY, but with 0.5% methanol instead of 1% glycerol).

The culture was placed in a baffled flask and returned to a shaking incubator at +16°C. 100% methanol was added to a final concentration of 0.5% every 24 h to maintain induction. 1 ml samples of the expression culture were taken 0, 24, 48, and 72 hours after induction, and both the cell pellets and the supernatants were stored for analysis. After 3 days of induction, the cells from the whole culture were harvested by centrifugation, and the supernatant was collected for further purification of GnTI.

Preparation of Crude GnTI Sample for Activity Assay

[00263] *Pichia pastoris* cell culture, which contained soluble His-tagged GnTI was processed for activity assay by concentration and buffer exchange. In brief, 40 ml of *P. pastoris* supernatant from shake flask culture was harvested at day 3 after induction with MeOH by pelleting the cells in 50 ml Falcon tube (Eppendorf 5810R, 3220 rcf, 5 min at +4°C) and collecting the supernatant. The supernatant was then concentrated to <2.5 ml by sequential centrifugations (Eppendorf 5810R or comparable, 3220 rcf, 10 min at +4°C) with Millipore Amicon Ultracel 30K concentrator. The volume of the concentrate was adjusted to 2.5 ml with 100 mM MES, pH 6.1. Concentrate was subjected to buffer exchange with a PD-10 gel filtration column (GE Healthcare 17-0851-01). The column was first equilibrated with 100 mM MES, pH 6.1 and then the sample (2.5 ml) was added, flow-through was discarded and elution with 2.25 ml of MES buffer was collected. Finally, 500 µl of the eluate was concentrated to 100 µl with Millipore Biomax 30K concentrator (Eppendorf 5417, 12 000 rcf, 5 min +4°C) and used directly in activity assays.

Activity Assay of GnTI Enzyme

[00264] Manα1-6(Manα1-3)Manβ1-4GlcNAc (Man_3Gn) was used as an acceptor for GnTI in the GnTI activity assay. The GnTI reaction was carried out by incubating the reaction mixture, which contained 0.1 mM acceptor $Man_3GlcNAc$, 20 mM UDP-GlcNAc, 50 mM GlcNAc, 100 mM $MnCl_2$, 0.5% BSA and 8 µl GnTI in 100 mM MES, pH 6.1, in a total volume of 10 µl at room temperature overnight. The reaction was stopped by incubating the reaction at 100°C for 5 min.

[00265] In parallel to the GnTI activity assay, the possible HexNAc'ase activity in the crude enzyme preparation was controlled. GlcNAcβ1-2Manα1-6(GlcNAcβ1-

2Manα1−3)Manβ1−4GlcNAcβ1-4GlcNAc-Asn (= $Gn_2Man_3Gn_2$-Asn) was used as a substrate for HexNAc'ase. The reaction was carried out in a similar way as for GnTI, except 100 pmol of $Gn_2Man_3Gn_2$-Asn was added instead of Man_3Gn and UDP-GlcNAc. No HexNAc'ase activity was detected.

[00266] The reaction mixture was purified for MALDI analysis by sequential Hypersep C_{18} (100 mg, Thermo Scientific, cat no: 60300-428) and Hypercarb (10 mg/96 well plate/1 PKG, cat no 60302-606) chromatography on HyperSep 96 –well Vacuum Manifold, Thermo Scientific. Hypersep C_{18} was prepared with 300 µl EtOH and 300 µl MQ water, the collection plate was then put under, and samples were loaded and eluted with 150 µl MQ water. Hypercarb was prepared with 300 µl MeOH and 300 µl MQ water. Eluates from Hypersep C_{18} were loaded, salts were removed with 150 µl 0.5 M NH_4Ac, and wells were washed with 2 x 300 µl MQ water. GnTI reaction products were eluted with 150 µl 25% ACN, and HexNAc'ase reaction products were eluted with 25% ACN and 0.05% TFA. Samples were dried in a Speedvac.

[00267] Matrix-assisted laser desorption-ionization time-of-light (MALDI-TOF) mass spectrometry (MS) was performed with a Bruker Ultraflex TOF/TOF instrument (Bruker Daltonics, Germany). Acceptor saccharide and product were detected in positive ion reflector mode as [M+Na]+ ions. Calculated m/z values for [M+Na]+ -signals of $Hex_3HexNAc_1$ and Hex_3 $HexNAc_2$ were 733.238 and 933.318, respectively. The percent ratio of the acceptor and the product was calculated from the signals corresponding to $Hex_3HexNAc_1$ and Hex_3 $HexNAc_2$ (Figure 19).

Generation of GnTII Construct for Production in P. pastoris

[00268] The nucleotide sequence encoding human GnTII was PCR-amplified with primers GP3 and GP13, which contained KpnI and EcoRI restriction sites, respectively. The EcoRI/KpnI-digested PCR fragment was ligated to a similarly digested pBLARG-SX cloning vector. After verifying the sequence, the final construct was transformed to *P. pastoris* strain GS190 to yield strain GY22. Positive yeast transformants were screened by PCR. Two clones (only one of which is shown in Figure 20) were studied for expression of GnTII under the control of the methanol-inducible AOX1 promoter at +16°C and at +30°C.

Expression of Soluble GnTII

[00269] According to Western blot analysis (Figure 20), *P. pastoris* strain GY22 produced soluble recombinant GnTII enzyme. GnTII has a calculated molecular mass of 49049.0 Da and two predicted N-glycosylation sites. The recombinant GnTII was secreted into the culture medium at +16°C (lane 9). When grown at +30°C, the recombinant GnTII was arrested inside the cells (lane 4).

Activity Assays of Soluble GnTII

[00270] *P. pastoris* cell culture containing soluble His-tagged GnTII was processed for an activity assay as described for GnTI above. Cell culture was centrifuged, supernatant was harvested and concentrated, buffer exchange to 100 mM MES, pH 6.1 was conducted, and the resulting sample was further concentrated prior to activity testing.

[00271] The activity assay was carried out similarly as for GnTI. GnMan3Gn was used as a GnTII acceptor.

[00272] The GnTII reaction was carried out in the presence of 0.1 mM acceptor GnMan3Gn, 20 mM UDP-GlcNAc, 50 mM GlcNAc, 100 mM MnCl2, 0.5% BSA, and GnTII in 100 mM MES, pH 6.1. Purification of the reaction mixture for MALDI-TOF MS analysis was performed by sequential Hypersep C18 and Hypercarb chromatography on a 96-well plate on vacuum manifold as described for GnTI above.

[00273] MALDI-TOF MS was performed with a Bruker Ultraflex TOF/TOF instrument (Bruker Daltonics, Germany). Acceptor saccharide and product were detected in positive ion reflector mode as [M + Na]+ ions. Ratio of the product and acceptor at the end of the reaction was calculated from their signal intensities (calculated m/z values for [M+Na]+ signals of GnMan3Gn acceptor and product with one GlcNAc addition are 933.318 and 1136.397, respectively).

[00274] Cultivation of *P. pastoris* producing GnTII was repeated, and GnTII concentrate (60x) from supernatant was prepared and its activity measured according to the methods described above. MALDI spectrum of time point samples at 2.5h, 5h, and overnight showed that 80%, 83%, and 82% of the acceptor was converted to product, respectively. The close-to-maximum reaction was reached in 2.5 hours.

[00275] In addition, a crude GnTII sample was prepared, and the activity assay was carried out as described above for the crude GnTI sample. The reaction mixture was incubated overnight, purified, and subjected to MALDI analysis. MALDI spectra revealed GnTII activity (Figure 21). HexNAc'ase activity was not detected in the crude GnTII sample.

[00276] The methods used to synthesize a GnTII acceptor for use in the above-described GnTII activity assays were as follows. A GnTI sample was prepared from a *P. pastoris* cultivation medium as described above. This GnTI sample showed high GnTI activity and, therefore, it could be used in conversion of about 40 nmol of Man3Gn to GnMan3Gn. The reaction was carried out in the presence of 0.5 mM Man3Gn, 20 mM UDP-GlcNAc, 50 mM GlcNAc, 100 mM MnCl2, 0.5% BSA, and GnTI sample. The reaction mixture was incubated three days at room temperature. A sample of ~1% was subjected to purification by Hypercarb chromatography and MALDI analysis. The GnTI reaction converted almost all of Man3Gn acceptor to GnMan3Gn product according to MALDI spectrum. Only 2.8% of the acceptor was not converted.

Example 4 – GnTI/GnTII Fusion Protein

Generation of GnTI/GnTII Expression Construct

[00277] A recombinant GnTI/II fusion protein was constructed by amplifying a 1313 bp GnTII fragment with a 65-mer fusion primer at the 5'-end, which contained an in-frame fusion site (a short sequence from GnTI containing a naturally occurring AleI restriction site with the stop-codon removed and overlapped with GnTII sequence) and 3'-end primers homologous to GntII containing either SpeI or NdeI restriction sites. This fusion site allowed the cloning of a fusion fragment directly to a *T. reesei* overexpression vector with wild type GnTI under the control of the *cbh1* promoter (cloning with AleI/NdeI) or with wild type GnTI under the control of the *gpd* promoter (cloning with AleI/SpeI). High-fidelity Phusion polymerase (Finnzymes) and standard amplification and cloning procedures were used. The sequence was verified by sequencing directly from expression vectors. The resulting vector was used to express the fusion as a transmembrane protein in *T. reesei*.

[00278] To gain more information on the functionality of the fusion proteins, fusion GnTI/II proteins were also expressed as soluble proteins in *P. pastoris*. CDS of the

GnTI/II fusion encoding the soluble part of the protein was cloned to the pBLARG-SX expression vector in order to produce protein for enzymatic studies. During the cloning procedure, His tag encoding sequence was added to the 5′-end of the frame to obtain a tag at the N-terminus of the truncated protein. The sequence was verified by sequencing analysis. The resulting vector was linearized and transformed by electroporation to *P. pastoris* strain GS190 to yield strain GY6. Arg^+ transformants were picked and screened by PCR. *P. pastoris* clones containing the integrated plasmid were tested for protein expression.

Purification of Soluble GnTI/II Produced in P. pastoris

[00279] Expression in *P. pastoris* and purification procedures were carried out as described above with recombinant GnTI protein.

Enzyme Activity Tests of GnTI/II Fusion Protein

[00280] Activity assays were carried out as described above for GnTI assays using Man3Gn oligosaccharide as an acceptor and UDP-GlcNAc donor. The products of the reaction were analyzed by MALDI- TOF mass spectrometry. Only GnTI activity was observed for the GnTI/GnTII fusion protein (Figure 22).

Transformation of T. reesei with GnTI/GnTII Construct by Random Integration

[00281] A chimeric human GnTI/GnTII plasmid with a *gpdA* promoter was co-transformed into the *T. reesei* M124 strain with random integration. Selection was obtained by co-transformation of a plasmid containing an acetamidase marker gene. Twenty PCR positive transformants were purified to uninuclear clones and grown in shake flask cultures for glycan analysis. All transformants and the parental strain M124 were cultivated in *Trichoderma* minimal medium (TrMM), pH 4.8, supplemented with 4% lactose and 2% spent grain extract. Supernatant and mycelia samples were collected on days 3, 5, and 7, and were stored frozen until analysis. In addition, as a control, *T. reesei* was transformed with a GnTI construct by random integration.

Glycan Analysis of T. reesei GnTI/GnTII Strains Obtained by Random Integration

[00282] Samples from 20 different clones at three different time points (days 3, 5 and 7) from *T. reesei* strain M124 GnTI/GnTII transformants were analyzed. Samples from two parental M124 strains were analyzed for controls. N-glycanase reactions without SDS denaturation were performed in 96-well plates in triplicate for 5µg of supernatant protein. The protein concentration of the supernatants was measured by Bradford-based assay (Bio-Rad Quick Start Bradford Protein Assay) using BSA as a standard. Both neutral and acidic N-glycans were analyzed by MALDI-TOF MS. No Go product was detected using the GnTI/GnTII construct in any of the clones at any time point as well as in clones of GnTI transformants with *gpdA* promoter.

Transformation of T. reesei with GnTI/GnTII Construct by Targeted Integration

[00283] A chimeric GnTI/GnTII sequence was subcloned into a pTTv38 backbone, a vector that contains an acetamidase marker gene and 5'- and 3'-flanking sequence sites for *alg3* locus integration. The vector was transformed into *T. reesei* M124 strain as a digested fragment. From this transformation, 18 PCR positive transformants, yielding PCR fragments indicating correct integration to the *alg3* locus, were detected. These transformants were cultured in shake flasks after a single spore purification step and were analyzed as described below.

Glycan Analysis of T. reesei GnTI/GnTII Strains Obtained by Targeting to alg3 Locus

[00284] Supernatant samples of 10 different clones at three different time points (days 3, 5 and 7) of Δ*alg3 T. reesei* GnTI/GnTII transformants were obtained. Clones had been cultivated in shake flasks with two different media compositions. TrMM, pH 5.5, with 2% spent grain extract, 4% lactose, and K-phthalate buffering was used for all clones and, in parallel, TrMM, pH 5.5, with 2% spent grain extract, 4% lactose, 1% casamino acids, and K-phthalate buffering was used for five of the clones. Cultivation was continued for 7 days: 5 days at +28°C and days 6 and 7 at +24°C.

[00285] N-glycan analyses were made in triplicate in 96-well plates for 5 μg of supernatant protein. Samples were analyzed from days 3, 5, and 7. The protein concentration of the supernatants was measured by Bradford-based assay (Bio-Rad Quick Start Bradford Protein Assay) using BSA as a standard. Both neutral and acidic N-glycans were analyzed by MALDI-TOF MS.

[00286] Detectable amounts of glycoform G0 were found in every clone. Clone 201A contained the most with 1.2% of Gn2Man3 (Figure 23 and Table 9). In addition, the amount of Hex6 was lowest in this particular clone. The second medium with 1% casamino acids did not give any extra production of G0/GlcNAcβ2Manα3(GlcNAcβ2Manα6)Manβ4GlcNAcβ4GlcNAcβ. The results of the days 3 and 7 samples were essentially the same as for the day 5 sample.

Table 9: The signal intensity percentages of observed N-glycans from secreted proteins of *T. reesei* GnTI/II transformants (GnTI/II integrated into the *alg3* locus). Clones with letter A in their name were cultivated in medium A) and clones with B in medium B), which had an extra 1% casamino acids compared to medium A).

		clone 201A, day 5					clone 202A, day 5					clone 208A, day 5					clone 210A, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Man2	771.3	0.6	0.5	86.8	0.0	1.0	0.4	0.7	173.2	0.0	1.1	0.6	0.6	92.3	0.0	1.1	0.1	0.2	173.2	0.0	0.4
Man3	933.3	47.9	14.5	30.2	39.0	64.6	41.3	0.2	0.4	41.1	41.5	38.2	1.1	2.8	37.0	38.9	38.2	1.1	3.0	37.5	39.5
Man4	1095.4	7.9	2.9	36.5	5.9	11.3	6.4	0.6	8.7	6.0	7.0	5.3	0.2	4.0	5.0	5.5	6.0	0.4	6.6	5.5	6.2
GnMan3	1136.4	1.4	0.7	46.9	1.0	2.2	1.1	0.3	23.5	0.8	1.3	1.0	0.2	17.0	0.9	1.2	1.1	0.1	8.9	1.0	1.2
Man5	1257.4	10.5	2.5	23.5	8.7	13.3	8.6	0.8	9.7	7.7	9.4	8.2	0.3	4.0	7.8	8.5	8.9	0.3	3.7	8.6	9.3
Gn2Man3	1339.5	1.2	0.8	69.1	0.6	2.2	0.6	0.1	21.0	0.5	0.8	0.6	0.1	21.5	0.5	0.7	0.6	0.1	17.5	0.6	0.8
Hex6	1419.5	27.3	23.7	86.7	0.0	42.0	40.5	0.6	1.5	39.9	41.1	44.7	0.7	1.6	43.9	45.2	43.2	0.7	1.6	42.7	44.0
Hex7	1581.5	2.9	3.0	103.3	1.1	6.4	1.0	0.1	11.0	1.0	1.2	1.1	0.1	11.7	1.0	1.2	1.2	0.0	3.7	1.2	1.2
Hex8	1743.6	0.1	0.2	173.2	0.0	0.4	0.2	0.3	173.2	0.0	0.5	0.3	0.2	87.0	0.0	0.4	0.6	0.3	57.0	0.3	1.0

		clone 212A, day 5					clone 213A, day 5					clone 215A, day 5					clone 216A, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Man2	771.3	0.4	0.4	86.8	0.0	0.7	0.6	0.6	94.4	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Man3	933.3	45.6	1.3	2.8	44.2	46.8	40.0	2.8	7.0	37.3	42.9	43.4	1.9	4.4	41.3	45.1	42.6	2.0	4.6	40.5	44.4
Man4	1095.4	5.6	0.3	5.1	5.4	5.9	6.5	0.6	8.8	6.0	7.1	6.3	0.5	8.5	5.7	6.8	6.1	0.6	10.3	5.4	6.7
GnMan3	1136.4	0.9	0.2	22.4	0.7	1.1	0.9	0.1	8.5	0.8	1.0	1.1	0.1	6.9	1.0	1.2	1.1	0.2	14.1	0.9	1.2
Man5	1257.4	7.2	0.5	7.0	6.8	7.7	9.5	0.4	3.8	9.1	9.8	8.5	0.4	4.2	8.2	8.9	7.7	0.6	8.4	7.0	8.3
Gn2Man3	1339.5	0.5	0.1	11.9	0.5	0.6	0.6	0.1	18.3	0.5	0.7	0.7	0.2	29.3	0.6	1.0	0.7	0.2	26.4	0.5	0.9
Hex6	1419.5	38.6	1.2	3.0	37.4	39.7	40.7	2.5	6.1	38.2	43.2	38.5	1.8	4.6	37.4	40.5	40.5	1.7	4.2	39.0	42.4
Hex7	1581.5	0.8	0.0	4.1	0.8	0.8	1.0	0.1	10.8	0.9	1.2	1.1	0.1	4.5	1.1	1.2	1.0	0.1	6.4	0.9	1.0
Hex8	1743.6	0.4	0.1	34.8	0.3	0.5	0.1	0.2	173.2	0.0	0.3	0.4	0.3	88.5	0.0	0.6	0.4	0.3	87.6	0.0	0.6

		clone 217A, day 5					clone 219A, day 5					clone 201B, day 5					clone 202B, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Man2	771.3	0.9	0.0	5.0	0.9	1.0	0.5	0.4	96.7	0.0	0.9	0.4	0.7	173.2	0.0	1.1	0.6	1.1	173.2	0.0	1.8
Man3	933.3	54.1	1.1	1.9	53.0	55.0	44.0	1.8	4.1	42.4	45.9	46.9	0.2	0.5	46.6	47.1	40.6	1.7	4.3	38.6	41.8
Man4	1095.4	5.2	0.3	6.5	4.9	5.5	5.7	0.1	1.5	5.6	5.8	6.9	0.9	12.7	6.0	7.8	8.5	0.9	10.0	7.7	9.4
GnMan3	1136.4	0.9	0.2	17.4	0.7	1.0	1.0	0.2	16.6	0.9	1.2	1.2	0.4	32.1	0.9	1.6	1.3	0.4	0.0	0.9	1.8
Man5	1257.4	5.8	0.1	2.6	5.6	5.9	8.0	1.2	15.6	6.7	9.2	8.1	0.5	5.7	7.8	8.6	10.0	0.6	6.2	9.5	10.6
Gn2Man3	1339.5	0.7	0.1	14.7	0.6	0.7	0.9	0.1	14.2	0.8	1.0	0.8	0.1	7.1	0.8	0.9	0.7	0.5	70.8	0.3	1.3
Hex6	1419.5	31.5	1.5	4.7	30.5	33.3	38.5	1.1	2.8	37.3	39.2	34.2	0.7	2.1	33.8	35.1	37.5	1.1	2.8	36.7	38.7
Hex7	1581.5	0.9	0.1	12.9	0.8	1.0	1.0	0.2	15.4	0.8	1.1	1.1	0.1	5.2	1.0	1.2	0.8	0.7	86.9	0.0	1.2
Hex8	1743.6	0.0	0.0	0.0	0.0	0.0	0.4	0.1	17.9	0.3	0.5	0.4	0.3	90.7	0.0	0.7	0.0	0.0	0.0	0.0	0.0

		clone 208B, day 5					clone 210B, day 5					clone 219B, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Man2	771.3	0.9	0.8	87.1	0.0	1.5	0.8	0.7	86.7	0.0	1.2	1.2	0.1	10.3	1.0	1.3
Man3	933.3	48.4	1.2	2.4	47.3	49.6	39.6	1.1	2.7	38.6	40.8	34.9	1.8	5.2	33.2	36.8
Man4	1095.4	7.2	0.2	2.2	7.0	7.3	7.9	0.6	8.0	7.3	8.5	8.1	0.3	4.1	7.8	8.4
GnMan3	1136.4	0.6	0.6	92.1	0.0	1.1	1.0	0.1	12.7	0.9	1.1	1.1	0.1	12.1	1.0	1.2
Man5	1257.4	8.7	0.7	7.6	7.9	9.1	9.6	0.2	2.0	9.4	9.8	11.3	0.8	7.5	10.7	12.3
Gn2Man3	1339.5	0.4	0.2	44.3	0.2	0.6	0.6	0.2	32.4	0.4	0.8	0.6	0.1	13.9	0.5	0.6
Hex6	1419.5	32.4	0.4	1.4	32.1	32.9	38.5	0.3	0.8	38.3	38.9	40.6	0.7	1.8	39.8	41.1
Hex7	1581.5	1.0	0.2	15.5	0.8	1.1	1.5	0.1	8.2	1.4	1.6	1.4	0.2	13.5	1.2	1.5
Hex8	1743.6	0.4	0.4	87.7	0.0	0.7	0.5	0.5	92.4	0.0	0.9	0.8	0.1	16.3	0.7	0.9

Example 5 - GnTII/GnTI Fusion Protein

Generation of GnTII/GnTI Expression Construct

[00287] A GnTII/GnTI fusion expression construct was generated by applying PCR overlap techniques. Fusion fragments were amplified from GnTII and GnTI templates separately with primers containing 50 bp in-frame overlaps at the fusion site. Fragments were purified from an agarose gel and used as PCR template for amplification of the fusion construct according to standard procedures. The fusion construct was cloned into a

vector with ApaI/SpeI restriction sites. The resulting construct was verified by sequencing analysis. A vector was generated for expressing the soluble form of GnTII/GnTI in *P. pastoris* with His tagging at the N-terminus of the target protein. This vector was generated in a similar manner as described above for the GnTI/II fusion construct.

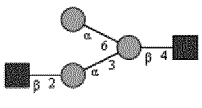
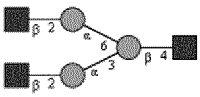
Purification of Soluble GnTII/GnTI Produced in P. pastoris

[00288] Expression in *P. pastoris* and purification procedures were carried out as described above for recombinant GnTI protein.

Enzyme Activity Tests of GnTII/GnTI Fusion Protein

[00289] Activity assays were carried out as described above for GnTI using Man3Gn oligosaccharide as an acceptor. A MALDI spectrum of the purified reaction mixture from the GnTII/GnTI reaction showed that two GlcNAcβ-residues were transferred to the acceptor (Figure 24).

Table 10: Summary of GnTII/GnTI fusion protein activities.

		Products formed	
GnTII/GnTI transformant	**Acceptor concentration**		
Transformant 1	0.5 mM	47%	5%
Transformant 1	0.1 mM	-	11%
Transformant 2	0.5 mM	3%	2.4%

Characterization by β-N-acetylglucosaminidase

[00290] The mixture formed in the GnTII/GnTI activity reaction was treated with β1-2,3,4,6-N-acetylglucosaminidase from *Streptococcus pneumoniae*. MALDI MS analysis was used to determine that both transferred β-linked GlcNAc residues were cleaved (Figure 25).

Galactosylation by β1-4GalT

[00291] The mixture formed in the GnTII/GnTI activity reaction was treated with β1-4GalT from bovine milk. β1-4GalT was expected to galactosylate the terminal GlcNAc residues in the product mixture. According to MALDI spectrum of the β1-4GalT reaction mixture, both products were galactosylated. Two galactoses were transferred to the Gn2Man3Gn product, which indicated that the GlcNAc residues were linked to separate mannose branches (Figure 26).

Transformation of T. reesei with GnTII/GnTI Construct by Random Integration

[00292] A chimeric GnTII/GnTI sequence was designed and cloned into a vector containing the *gpdA* promoter. After verification of the plasmid sequence, it was co-transformed into the *T. reesei* M124 strain with the hygromycin marker gene. Thirteen PCR positive transformants were identified. All positive transformants and the parental strain M124 were cultivated in TrMM, pH 4.8, supplemented with 4% lactose and 2% spent grain extract. In addition, seven transformants and the parental strain were cultivated in TrMM, pH 5.5, with 4% lactose, 2% spent grain extract, and 1% casamino acids, buffered with 100 mM PIPPS (piperazine1,4bis2propanesulfonic acid). pH measurements were used to monitor the growth rate of the strains. Supernatant and mycelia samples were collected on days 3, 5, and 7, stored frozen, and analyzed for glycan structures. The GnTII/GnTI sequence was also cloned into a plasmid containing the *cbh1* promoter. In addition, as a control, *T. reesei* was transformed with a GnTI construct by random integration.

Glycan Analysis of T. reesei GnTII/GnTI Strains Obtained by Random Integration

[00293] 156 supernatant samples of *T. reesei* strain M124 GnTII/GnTI transformants and parental M124 strain cultivated in two different media were analyzed. The first medium was TrMM, pH 4.8, supplemented with 2% spent grain extract and 4% lactose, and the second medium was TrMM, pH 5.5, supplemented with 2% spent grain extract, 4% lactose, 100 mM PIPPS, and 1% casamino acids. Cells were grown in both types of media for 3, 5 and 7 days.

[00294] N-glycanase reactions without SDS denaturation were carried out in 96-well plates in triplicate for 5µg of supernatant protein for samples from time points of 3 and 5 days. The protein concentration of the supernatants was measured by Bradford-based assay (Bio Rad Quick Start Bradford Protein Assay) using BSA as a standard. Both neutral and acidic N-glycans were analyzed by MALDI-TOF MS.

[00295] No sign of the expected GnTII/GnTI product was visible in any of the clones from time points of 3 and 5 days. In addition, no product was observed from GnTI and GnTI/II transformants with *gpdA* promoters that were generated by random integration.

Transformation of T. reesei with GnTII/GnTI Construct by Targeted Integration

[00296] A vector having the chimeric GnTII/GnTI sequence under the control of the *cbh1* promoter was constructed with a *pyr4* gene loopout marker and subcloned into a backbone vector between *alg3* flanking region fragments for targeted integration. A PmeI-digested expression cassette was transformed into *T. reesei* strain M127 (*pyr4*⁻ strain of M124). After plate selection, the clones were PCR-screened and purified through single spores. To obtain material for glycan analyses, shake flask cultivations were performed as described. Five PCR positive transformants indicating correct integration to the *alg3* locus in the M127 transformation were cultivated in a 300 ml volume for seven days at +28°C in a media containing TrMM, pH 5.5, supplemented with 40 g/l lactose, 20 g/l spent grain extract, and 100 mM PIPPS. To avoid bacterial contamination, 100 mg/l ampicillin was added into the flasks at the time of inoculation. Samples for glycan analyses were collected on days 3, 5 and 7.

Glycan Analysis of T. reesei GnTII/GnTI Strains Obtained by Targeting to alg3 Locus

[00297] Supernatant samples of *T. reesei* strain M124 (control), five different clones of M127 GnTII/GnTI transformants, and control medium samples were prepared in triplicate on 96-well plates for 5 µg of supernatant protein. The protein concentrations of the supernatants were measured by Bradford-based assay (Bio-Rad Quick Start Bradford Protein Assay) using BSA as a standard. PNGase F reactions were performed as described, but without SDS denaturation. The released N-glycans were first purified with Hypersep C-18 and then with Hypersep Hypercarb (both from Thermo Scientific) where

neutral and acidic glycans were separated. Both purifications were performed in 96-well format. Neutral N-glycans were analyzed by MALDI-TOF MS.

[00298] The proportions of neutral N-glycans from *T. reesei* M127 GnTII/GnTI transformants were compared to proportions from strain M124, which was otherwise the same as strain M127 but *pyr4* positive. Four of the five GnTII/GnTI transformants produced G0 as a main glycoform at all time points (3, 5 and 7 days). Only clone 46A was G0 negative (Figure 27). The proportion of Man3Gn was small in every clone at all time points, but the proportion of Hex6 was still quite large. On day 7, clone 17A produced the most G0 and the least Hex6 in comparison to other clones (Figure 27). Four clones of the GnTII/GnTI transformants produced around 40% of glycoform G0 on day 5 in shake flask conditions (Figure 27). Fermentation conditions with controlled pH can increase the amount of G0 product and reduce the amount of Hex6 in *alg3* knock-outs.

[00299] In the medium sample, a series of plant-type N-glycans were observed, but no signals corresponding to G0 were observed.

Transformation of Rituximab-Producing T. reesei with GnTII/GnTI Construct by Targeted Integration

[00300] The expression cassette described in the section entitled "Transformation of *T. reesei* with GnTII/GnTI Construct by Targeted Integration" was transformed into *T. reesei* strain M279 (*pyr4*⁻ strain of the strain M202). M202 was obtained by deleting pep1 protease in M124 and introducing rituximab heavy and light chain (with Kex2 cleavage site). After plate selection, the clones were PCR-screened and purified through single spores. To obtain material for glycan analyses, shake flask cultivations were performed as described in the section entitled "Transformation of *T. reesei* with GnTII/GnTI Construct by Targeted Integration" and, in addition, some culture media were supplemented with 0.3 mg/ml soybean trypsin inhibitor (SBTI) and 1% casamino acids. SBTI was added first at inoculation and then daily on days 3-6. PMSF and Pepstatin A were added to all samples before freezing.

Glycan Analysis of Rituximab-Producing T. reesei GnTII/GnTI Strains Obtained by Targeting to alg3 Locus

[00301] Rituximab was purified with Protein G affinity chromatography from day 5 supernatant samples with SBTI and from day 5 and 7 samples without SBTI. PNGase F reactions were performed for ~10 µg of denatured protein. The released N-glycans were first purified with Hypersep C-18 and then with Hypersep Hypercarb (both from Thermo Scientific) where neutral and acidic glycans were separated. The purification steps were performed in 96-well format. Neutral and acidic N-glycans were analyzed by MALDI-TOF MS. Two of the GnTII/GnTI transformant clones, 9A-1 and 31A-1, produced G0 glycoform at ~30% and ~24%, respectively. However, reasonable quantities of Hex6 and GnMan3 were still observed (Figure 28). Rituximab from the other clones contained little or no G0.

Optimization of Spacers

[00302] A series of spacer modifications for GnTII/GnTI fusion proteins were constructed. These variants were produced in *Pichia* and studied *in vitro* for enzyme stability and activity.

[00303] The materials and methods for cloning the GnTII/GnTI fusion proteins are described here. T45 sequence was amplified in two parts by using PCR overlapping strategy. First, a fragment was amplified with GP13 5' primer and GP93 3' primer, and a second fragment was amplified with GP92 5' primer and GP2 3' primer. Amplification was carried out with Phusion high-fidelity PCR polymerase (Finnzymes) under the standard conditions provided by the supplier. Cycling conditions were as follows: initial denaturation at 98°C for 30 seconds, denaturation at 98°C for 5 seconds, annealing at 65°C for 30 seconds, extension at 72°C for 45 seconds, repeat 20 times, and final extension at 72°C for 20 minutes. The resulting PCR products were purified from the agarose gel with a Fermentas GeneJET gel extraction kit. These fragments with overlapping, modified sequences were combined in the same reaction mixture with standard conditions without primers. Ten annealing/extension cycles were carried out as follows: initial denaturation at 98°C for 30 seconds, denaturation at 98°C for 5 seconds, annealing at 65°C for 30 seconds, extension at 72°C for 45 seconds, repeat 10 times, and final extension at 72°C for 20 minutes. Primers GP13 (5') and GP2 (3') were added, and

cycling was continued as described above for 20 amplification cycles. The amplified T45 fragment was purified with a Fermentas GeneJET PCR purification kit, digested with EcoRI/KpnI (New England Biolabs) according to standard protocols, and cloned into EcoRI/KpnI digested yeast expression vector pBLARG-SX. The resulting vector was sequenced with primers 3'AOX, 5'AOX, GP9, GP37, GP38 and GP122. The sequence was found to be correct.

[00304] This resulting plasmid was used as a template for the 3xG4S spacer modification. Cloning of the T46 sequence was done as described above with T45. GP13 5'-primer and GP95 3'-primer were used for first fragment synthesis, and GP94 5'-primer and GP2 3'-primer were used for second fragment synthesis. Fragments were combined, and primers GP13 (5') and GP2 (3') were added for amplification. Amplified fragment T46 was then digested with EcoRI/KpnI and cloned into yeast expression vector pBLARG-SX. The resulting vector was sequenced with the primers described above, and the sequence was found to be correct.

[00305] Cellulase-related natural spacers were constructed with a similar PCR overlap method. With the CBHI-related spacer, the first fragment was amplified with GP13 5'-primer and GP107 3'-primer. The second fragment was amplified with GP108 5'-primer and GP2 3'-primer (Table 11). With the EGIV-related spacer, the first fragment was amplified with GP13 5'primer and GP109 3'primer. The second fragment was amplified with GP110 5'-primer and GP2 3'-primer (Table 11). In both cases, PCR products were purified from agarose gel, combined, and used as a template for the next PCR reaction to amplify the sequences T50 and T51. T50 and T51 PCR products were then digested with EcoRI/KpnI and cloned into yeast expression vector pBLARG-SX.

[00306] All PCR amplifications were made with high-fidelity Phusion polymerase (Finnzymes). Primers (Table 11) were ordered from MWG Operon. Sequencing was performed by the DNA Sequencing Laboratory of the Institute of Biotechnology, University of Helsinki, as a commercial service.

Table 11: Primer sequences.

Primer	Sequence 5'- 3'
3'AOX	GCAAATGGCATTCTGACATCC (SEQ ID NO: 99)
5'AOX	GACTGGTTCCAATTGACAAGC (SEQ ID NO: 100)
GP2	CAGTGGTACCCTAATTCCAGCTAGGATCATAGCCCTCCCACG (SEQ ID NO: 101)
GP9	CGGACCACCGCAAGTTCC (SEQ ID NO: 102)
GP13	ATGCGGAATTCTGCATCATCATCATCATCATCGCCAGCGTAAGAACGAGGCCCT (6 × HIS) (SEQ ID NO: 103)
GP37	CCTTTCTCTATCCAACTCTACC (SEQ ID NO: 104)
GP38	GGAACTTGCGGTGGTCCG (SEQ ID NO: 105)
GP92	CCGCCGGCTCCAGGGAGGTGGGGGCAGTGGAGGTGGCGGCAGTGGGAGGGTGCCCACC GCCGCCCC (SEQ ID NO: 106)
GP93	GCGGTGGGCACCCTCCCACTGCCGCCACCTCCACTGCCCCCACCTCCCTGGAGCCGGCGG TAAGAC (SEQ ID NO: 107)
GP94	AGGTGGGGGCAGTGGAGGTGGCGGCAGTGGCGGCGGTGGAAGTGGGAGGGTGCCCACC GCCGCCC (SEQ ID NO: 108)
GP95	CGGTGGGCACCCTCCCACTTCCACCGCCGCCACTGCCGCCACCTCCACTGCCCCCACCTC CCTG (SEQ ID NO: 109)
GP107	GTTTCCGCCGGGAGGGTTGCCGCCGCTAGGGTTGCCGGTGCTCTGGAGCCGGCGGTAAG ACTTGC (SEQ ID NO: 110)
GP108	GCAACCCTCCCGGCGGAAACCCGCCTGGCAGCACCGGGAGGGTGCCCACCGCCGCCCCT CCCGCCC (SEQ ID NO: 111)
GP109	CCGCCTCCAGGAACAGTGGCGCTGGCGGTGGCCGTCGCGGCGGAGCTCTGGAGCCGGCG GTAAGACTTGC (SEQ ID NO: 112)
GP110	CGCCACTGTTCCTGGAGGCGGTAGCGGCCCCACCAGCGGGAGGGTGCCCACCGCCGCCC CTCCCGCCCAGC (SEQ ID NO: 113)
GP122	CATTAGCGAGAAGTTTACGG (SEQ ID NO: 114)

[00307] Spacer modified (3xG4S and 2xG4S) GnTII/GnTI fusion enzymes were processed for an activity assay by concentration and buffer exchange in a similar way as described for GnTI in Example 3. Activity assays were carried out with Man3Gn acceptor, and reaction mixtures were purified as described in the GnTI activity assay. MALDI analysis was also performed as described with the GnTI reaction mixture, but, in addition, formation of the GnTII product, Hex3HexNAc3, was followed. The calculated m/z values for the [M+Na]+ signal of Hex3HexNAc3 was 1136.318 (Figure 29).

Spacer variants

[00308] GnTII/I spacer variants were modified from the wild type spacer sequence of the GnTII/I fusion protein. The modified spacers are listed in Table 12. All four spacer variant strains (GY32, GY33, GY49, and GY50), wild-type GnTII/I fusion strain (GY7-2), and mock strain (GY3) were expressed at +16°C with protease inhibitors. Strains were

inoculated in 60 ml of BMGY-medium at +30°C, 220 rpm, over-night (o/n). Over-night cultures were pelleted and cells were resuspended in 60 ml of BMMY-medium. Protease inhibitors, 1 mM EDTA, 1.5 µM Pepstatin A (Sigma) and 1 Complete EDTA free protease inhibitor cocktail tablet (Roche) were added in cultures at the same time when MeOH induction was started and after that once in a day. 25 ml samples were taken from cultures on day 3 and day 4, and supernatant samples were concentrated using concentration tubes (Millipore), buffer was exchanged in PD-10 columns into 100 mM MES pH 6.1 and concentrated into final 50x. Cell pellets were resuspended in 500 µl of 1xPBS, except cell pellet of wild type (3rd), which was resuspended in 500 µl of 100 mM MES pH 6.1 and complete (EDTA free) inhibitor cocktail.

[00309] The amino acid sequence of the GnTII/GnTI fusion protein containing the 3xG4S spacer is set forth in SEQ ID NO: 119. The nucleotide sequence of the GnTII/GnTI fusion protein containing the 3xG4S spacer is set forth in SEQ ID NO: 141. The amino acid sequence of the GnTII/GnTI fusion protein containing the 2xG4S spacer is set forth in SEQ ID NO: 121. The nucleotide sequence of the GnTII/GnTI fusion protein containing the 2xG4S spacer is set forth in SEQ ID NO: 139. The amino acid sequence of the GnTII/GnTI fusion protein containing the CBHI spacer is set forth in SEQ ID NO: 123. The nucleotide sequence of the GnTII/GnTI fusion protein containing the CBHI spacer is set forth in SEQ ID NO: 143. The amino acid sequence of the GnTII/GnTI fusion protein containing the EGIV spacer is set forth in SEQ ID NO: 125. The nucleotide sequence of the GnTII/GnTI fusion protein containing the EGIV spacer is set forth in SEQ ID NO: 145.

[00310] A 200 µl sample of cell suspension was washed by repeating centrifuging and resuspending cells in 100 mM MES pH 6.1 with complete (EDTA free) inhibitor cocktail. A cell lysate was prepared by taking 200 µl of washed cell sample, adding 50 µl glass beads and 2 µl Triton X-100 and putting in bead beater for 6 min. GnTI activity assays of 50x concentrated *P. pastoris* culture supernantants, cell sample and cell lysate were performed as above.

Table 12: Description of yeast strains.

Yeast Strains	Description	Sequence of spacer variant
GY3	Mock strain	
GY7-2	Wild-type GnTII/I fusion	
GY32-5 GY32-9	GnTII/I fusion 3xG4S spacer variant	SEQ ID NO: 118
GY33-7 GY33-8	GnTII/I fusion 2xG4S spacer variant	SEQ ID NO: 120
GY49-3	GnTII/I fusion CBHI spacer variant	SEQ ID NO: 122
GY50-7 GY50-10	GnTII/I fusion EGIV spacer variant	SEQ ID NO: 124

[00311] Western blots analysis of cell pellets and 50x concentrated culture supernatants from day 3 are shown in Figure 30. The CBHI spacer variant (GY49) gave a strong signal from the cell pellet sample but not from the supernatant. The EGIV spacer variant (GY50) was detected from the supernatant, but only faint signal was obtained. Faint signals from supernatant samples were also obtained with the wild-type GnTII/I fusion strain (GY7-2) and the 2xG4S spacer variant strains GY33-7 and GY33-8 (Figure 30).

[00312] The activities of the GnTII/I fusion protein containing the spacer variants were then compared to the activity of the GnTII/I fusion protein containing the wild-type spacer.

[00313] *Fusion GnTII/I activity in supernatants.* The GnTI substrate Man3Gn was provided and the reaction product, GnMan3Gn, acted as the acceptor for the GnTII activity of the fusion protein. Samples for activity assays were taken after day 3 and day 4 expression phases. Figure 31 shows activity assay results of cultures of GnTII/I fusion proteins containing either the wild type spacer or the spacer variants. Sample cultivations were done in the presence of inhibitors (1,5 µM pepstatin A, 1 mM EDTA, 1 tablet/50 ml of complete EDTA free protease inhibitor cocktail tablet). For simplicity, the GnTI and GnTII reaction products were added together. All activity assay samples contained only minor amounts (< 5%) of GnTI product GnMan3Gn, indicating that GnTII actively transformed the GnMan3Gn to Gn2Man3Gn.

[00314] All four spacer variants showed GnT activities, although there was some variability between clones and cultivation days. The GnTII/I fusion proteins containing

the 2xG4S (clone_1), 3xG4S (clone_1 and clone_2), or EGIV spacer variants showed higher activy than the enzyme with the wild-type spacer (Figure 31). The GnTII/I fusion protein containing the CBHI spacer variant showed comparable activity with the enzyme with the wild-type spacer (Figure 31). The GnTII/I fusion protein contining the 2xG4S variant (clone_2) had lower activity than the enzyme with the wild-type spacer (Figure 31). Day 4 samples had higher activities than day 3 samples, with the exception of the GnTII/I fusion protein contining the 3xG4S spacer variants (clone_1 and clone_2), which showed higher activity on day 3 (Figure 31). The GnTII/I fusion protein contining the EGIV spacer variant had the highest activity on day 4 (Figure 31).

[00315] *Fusion GnTII/I activity in cells and cell lysates.* Activity assays of cell, cell lysate, and supernatant samples from cells containing the GnTII/I fusion protein having the wild-type spacer indicated that lysate samples contained the highest activity (Figure 32). The second highest activity was on the cell surface, and lowest activity was seen in the supernatant samples (Figure 32). Accordingly, it appears that most of the GnTII/I fusion protein was localized in cells or on the cell surface, with only a small amount being secreted.

[00316] GnT activities of cells containing GnTII/I fusion proteins having either the wild-type spacer or the spacer variants are shown in Figure 33. The cells were resuspended in 500 µl of 100 mM MES, pH 6,1 with complete EDTA free inhibitor cocktail and spacer variants in 500 µl PBS and cells and lysates for activity testing were prepared as above.

[00317] As shown in Figure 33, GnTII/I fusion proteins containing the spacer variants had much higher GnTII/I activity in cells than in supernatants. In lysates, the enzymes appeared to be inactive. It is believed that this lack of activity is due to the action of released proteases. The GnTII/I fusion protein containing the CBHI spacer variant showed a high activity in cells and lysates (Figure 33), which correlates with Western blot analysis showing higher signal in the cell pellet sample (Figure 30).

[00318] *Discussion.* In supernatants, the GnTII/I fusion proteins containing the 2xG4S and 3xG4S spacer variants had higher activity that the GnTII/I fusion protein containing the wild-type spacer, while the CBHI spacer variant had comparable activity to the GnTII/I fusion protein containing the wild-type spacer. Moreover, the GnTII/I fusion

protein containing the EGIV spacer variant showed the highest GnT activity. Western blot analysis of day 3 samples had some correlation with the results of day 4 activities. Western blot analysis showed faint bands with supernatant samples of wild-type, both clones of 2xG4S and EGIV. The activities were detected in the following order: EGIV > 2xG4S (clone_1) > 3xG4S (clone_2) > 3xG4S (clone_1) ≥ CBHI = wild-type = 2xG4S (clone_2).

[00319] Determination of GnTII/I fusion protein activity in supernatant, cell, and cell lysate samples of the GnTII/I fusion protein containing the wild-type spacer showed that most of the activity is associated within the cells and lower amount is secreted. It is believed that this explains why much better signals of His-tagged GnTII/I were seen in cell fractions rather than in supernatant fractions in Western blot analysis.

[00320] The inhibition of serine and cysteine proteases by complete EDTA free inhibitor tablet, metalloproteinases by EDTA and aspartic proteases by pepstatin A, improved the yield of GnTII/I fusion protein. This observation on the use of serine protease inhibitor is in accordance with the work of Salamin et al. (*Appl. Environ. Microbiol.*, 76 (2010) 4269-4276), which showed that serine type protease activity in the media of *P. pastoris* was completed inhibited with PMSF. In addition, Vad et al. (*J. Biotechnol.* 116 (2005) 251-260) reported high production, over 300 mg/l, of intact human parathyroid hormone in *P. pastoris* in the presence of 10 mM EDTA combined with co-expression of *Saccharomyces cerevisiae* protein disulphide isomerase.

[00321] All GnTII/I fusion proteins containing each of the four spacer variants possessed GnTII/I activity, and the activity of the enzymes having the 2xG4S and EGIV spacer variants had higher activities that the GnTII/I fusion protein containing the wild-type spacer.

Example 6 – Use of Fusion Proteins with Man5 as the Acceptor Glycan

Construction of Rituximab-Expressing T. reesei Strain with Man5 Type N-glycosylation

[00322] The native rituximab sequence is codon harmonized. Original plasmids containing the synthesized rituximab light chain and heavy chain are generated. The antibody chains and CBHI fusion protein are designed with 40-nucleotide overlapping

sequences as are the expression vectors pHHO1 (acetamidase selection marker, *cbh1* flanks for integration into the *cbh1* locus) for the heavy chain or pHHO2 (hygromycin selection marker, *egl1* flanks for integration into the *egl1* locus) for the light chain, to enable cloning using yeast homologous recombination.

[00323] The obtained gene plasmids are transformed into *E. coli*. DNA is prepared, and the synthetic genes are digested and isolated from the plasmid backbones. The expression vectors are constructed by yeast homologous recombination on the *T. reesei* expression vectors with the CBHI fusion protein and either heavy or light chain. The recombined plasmids are rescued from yeast and transformed into *E. coli*. After PCR screening, correct clones are isolated and sequenced. The expression cassette fragments are digested and isolated from the plasmid backbone resulting in around 10.2 kb fragments for the heavy chain constructs and 10.8 kb fragments for the light chain constructs. The heavy and light chain fragments are cotransformed into the *T. reesei* strain M124. Transformants are selected for hygromycin resistance and ability to grow on acetamide as a sole nitrogen source. Transformants are streaked on the double selective medium for two successive rounds and tested by PCR for integration of the expression constructs into the genome.

Introduction of GnTII/I Tandem Enzyme and Mannosidase II to T. reesei Strain Expressing Rituximab Antibody

[00324] In addition to introducing a recombinant GnTII/I into a Man5-producing strain such as M124, a mannosidase II activity is further needed to remove two mannoses from the GlcNAcMan5 glycan structure so that GnTII/I can use GlcNAcMan3 as an acceptor molecule.

[00325] The GnTII/I expression cassette described in previous examples can be targeted to, for example, the *cbh2* locus of *T. reesei*, using methods essentially as described above. To generate a GlcNAcMan3 acceptor molecule for GnTII/I fusion protein, mannosidase II activity is then introduced to the strain using transformation methods described above.

[00326] Mannosidase II activity is introduced to the rituximab antibody-expressing M124 strain by designing a desired mannosidase-containing expression cassette with a promoter for driving the mannosidase expression. Useful promoters are those from *gpdA*

or *cbh1*. Mannosidase II activity can be transformed by random integration followed by screening of strains with most suitable expression level. The expression cassette is linked with a proprietary selection marker gene, or a selection marker is co-transformed as a separate expression cassette. Transformation is performed according methods described above.

[00327] A mannosidase II fusion construct can be derived from a *T. reesei* cytoplasmic, transmembrane and stem domain, or targeting peptide, of KRE2 and ligated in-frame to an N-terminal amino acid deletion of a human mannosidase II. The encoded fusion protein localizes in the ER/Golgi by means of the KRE2 targeting peptide sequence while retaining its mannosidase catalytic domain activity and is capable of hydrolyzing GlcNAcMan5GlcNAc2 into GlcNAcMan3GlcNAc2. In certain embodiments, a full-length human mannosidase II can be expressed in an M124 strain.

[00328] The KRE2 targeting peptide comprises the amino acids from about 1 to about 106 or from about 1 to about 83 of KRE2.

Kre2 aa 1-106
MASTNARYVRYLLIAFFTILVFYFVSNSKYEGVDLNKGTFTAPDSTKTTPKPPATG DAKDFPLALTPNDPGFNDLVGIAPGPRMNATFVTLARNSDVWDIARSIRQ (SEQ ID NO: 115)

Kre2 aa 1-83
MASTNARYVRYLLIAFFTILVFYFVSNSKYEGVDLNKGTFTAPDSTKTTPKPPATG DAKDFPLALTPNDPGFNDLVGIAPGPR (SEQ ID NO: 116)

[00329] After transformation of *Trichoderma* with the mannosidase II construct described above, *Trichoderma* strains are selected, streaked on selective medium for two successive rounds, and tested by PCR for integration of the expression constructs into the genome. Selected transformants of *Trichoderma* strains producing Man5 and expressing the GnTII/I fusion protein, mannosidase II, and rituximab antibody are then cultured in shake flasks or fermentor conditions and analyzed for glycan content as described above.

Example 7 – Expression of GnTI and GnTII in T. reesei

Transformation of T. reesei M124 with GnTI Construct by Random Integration

[00330] Codon optimized human *GntI* was transformed into the *T. reesei* M124 strain. The *GntI* gene was cloned into a vector under the control of two different promoters: (1) the inducible promoter of the *cbh1* gene; and (2) the constitutively expressed promoter of the *gpdA* gene. The vectors containing *GntI* under either of the two promoters were each co-transformed into the *T. reesei* M124 strain with a plasmid containing either an acetamidase or a hygromycin resistance marker gene.

[00331] Thirty-four transformants with *GntI* under the *gpdA* promoter and under acetamide selection were screened by PCR, and all were positive for *GntI*. For transformants with *GntI* under the *cbh1* promoter and under acetamide selection, 19 of 26 were PCR-positive for the *GntI* construct. In addition, initial DNA extraction was performed for five strains with *GntI* under the *cbh1* promoter and under hygromycin selection. All of these strains were PCR-positive. Twenty-five *gpdA* promoter transformants and all of the *cbh1* promoter transformants (14+5) were purified to uninuclear clones and spore suspensions were prepared.

[00332] For initial analysis purposes, 23 *gpdA* promoter transformants and 19 *cbh1* promoter transformants (14 grown from acetamide and five from hygromycin selection), as well as the parental strain M124 were cultivated in 250 ml shake flasks with 50 ml of *Trichoderma* minimal medium supplied with 2% spent grain extract and 4% lactose. Growth of the strains was monitored by pH measurements. Samples (supernatants and mycelia) were collected on days 3, 5, and 7, stored frozen until used for glycan structure analysis.

Glycan analysis of T. reesei GnTI Strains Obtained by Random Integration

[00333] The protein concentration of all supernatant samples was measured by Bradford-based assay (BioRad Quickstart Bradford Protein Assay) using BSA as a standard. Secreted protein content of samples subjected to N-glycan analysis was adjusted to 5 µg or 10 µg. N-glycan analysis was performed either on 96-well plates for 5 µg of supernatant protein, or in 1.5 ml tubes for 10 µg of supernatant protein. All N-glycan

analyses were performed in triplicate. Both neutral and acidic N-glycans were analyzed with MALDI-TOF MS.

[00334] To get more exact measurements of the amount of the GnT1 product Gn1Hex5 produced in four of GnT1 transformants (from days 3 and 5) and also of the amount of produced acidic N-glycans, the MALDI spectra was spiked with a known glycan. For neutral and acidic
N-glycans, an internal calibrant of 2 pmol/spectrum Hex2HexNAc4 at the mass value of 1177 Da and 0.5 pmol of monosialylated Hex4HexNAc2 at the mass value of 1362 Da were used, respectively. Analyses were performed in triplicate.

[00335] No GnT1 product was observed in any of the *gpdA* promoter transformants. However, eight *cbhI* promoter transformants produced the GnT1 product Gn1Man5 (Figures 34 and 35, and Table 13); five with hygromycin selection, three with acetamide selection.

Table 13: The percentages of signal intensities of Man5 and Gn1Man5 compared to internal calibrant Hex2HexNAc4 in four positive GnT1 transformants and parental M124 strain on days 3 and 5. Man5 is the main glycoform in parental M124 strain.

		M124 1., day 3					M124 1., day 5					M124/GNT1, clone HM1, day 3					M124/GNT1, clone HM1, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Hex2HexNAc4	1177.42	97.7	0.5	0.5	97.1	98.0	36.5	0.8	2.3	35.9	37.1	78.5	14.5	18.4	68.2	88.7	50.1	10.6	21.2	42.6	57.6
Hex5HexNAc2	1257.42	2.3	0.5	22.5	2.0	2.9	63.5	0.8	1.3	62.9	64.1	14.5	9.9	68.0	7.5	21.5	44.0	9.6	21.9	37.2	50.8
Hex5HexNAc3	1460.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	4.6	65.6	3.8	10.3	5.9	1.0	16.7	5.2	6.6
		M124/GNT1, clone 8, day 3					M124/GNT1, clone 8, day 5					M124/GNT1, clone 39, day 3					M124/GNT1, clone 39, day 5				
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX
Hex2HexNAc4	1177.42	77.3	7.6	9.8	72.0	82.7	67.3	10.0	14.9	56.5	76.3	83.7	1.5	1.8	82.7	84.8	40.0	1.9	4.6	37.9	41.1
Hex5HexNAc2	1257.42	15.0	5.2	34.4	11.4	18.7	18.9	6.2	32.5	12.8	25.1	8.3	1.0	11.7	7.6	8.9	46.9	1.8	3.8	45.6	49.0
Hex5HexNAc3	1460.5	7.6	2.4	31.6	5.9	9.3	13.8	4.0	29.1	10.8	18.3	8.0	0.6	6.9	7.6	8.4	13.1	0.3	2.1	12.7	13.3
		M124/GNT1, clone 90, day 3					M124/GNT1, clone 90, day 5														
Composition	m\z	Average	SD	RSD	MIN	MAX	Average	SD	RSD	MIN	MAX										
Hex2HexNAc4	1177.42	93.8	1.6	1.7	92.4	95.6	92.6	2.7	2.9	89.8	95.3										
Hex5HexNAc2	1257.42	3.7	1.0	25.9	2.6	4.5	4.7	1.4	30.9	3.2	6.0										
Hex5HexNAc3	1460.5	2.5	0.7	26.2	1.8	3.1	2.7	1.3	47.8	1.5	4.1										

[00336] The GnT1 products Gn1Man6P1, Gn1Man7P1, and Gn1Man8P1 were also found in phosphorylated N-glycans of all positive transformants. The amount of phosphorylated N-glycans had increased in GnT1 transformants, and the profile was biased toward larger N-glycans, with Man7P1 or Man8P1 having the strongest signal (Man6P1 in parental M124) (Figure 36).

[00337] Eight GnTI transformants produced the Gn1Man5 structure. Gn1Man5 was most abundant in clone 39. However, the best clone appeared to be clone 8, which produced the second highest level of Gn1Man5, but had a high proportion of Man5 and Gn1Man5 (Figure 35). Clone 8, which contains *GnTI* under the control of the *cbhI* promoter, was named strain M198, and was selected for continued analysis.

Transformation of T. reesei M198 Strain with GnTII Construct by Targeted Integration

[00338] Five *GnTII*-harboring vectors were created (Table 14). Two of the vectors contained the native mammalian Golgi targeting peptide in GNTII. In the three other vectors, the mammalian targeting peptide was replaced by a *T. reesei* MNT1 (α-1,2-mannosyltransferase) targeting peptide. All five vectors contained either a *cbh1* promoter or a *gpdA* promoter, and a *pyr4* loop-out marker. Additionally, all five vectors were targeted to integrate into the *alg3* locus, thus deleting the *alg3* gene. In the MNT1/GnTII constructs under the *cbh1* promoter, two different sized GnTII sequence deletions were tested.

Table 14: Constructed GNT2 vectors.

Plasmid name	Promoter	Targeting peptide	N-terminal deletion (GnTII)
pTTv140	*cbh1*	mammalian	N/A
pTTv141	*gpdA*	mammalian	N/A
pTTv142	*cbh1*	*Trichoderma* MNT1	74 amino acids
pTTv143	*cbh1*	*Trichoderma* MNT1	104 amino acid
pTTv144	*gpdA*	*Trichoderma* MNT1	74 amino acids

[00339] These vectors, except for the pTTv144 vector, were transformed into the best *py4*- negative *GnTI* producing strain M198 (M319) as *Pme*I fragments. Transformants were purified to uninuclear clones and PCR screened. Clones showing the correct integration at both ends were then selected for continued analysis.

[00340] To study the growth characteristics of the generated GNTII-expressing strains, large shake flask cultures were prepared. Shake flask culture were prepared in two separate batches. The first batch contained pTTv140, pTTv142, and pTTv143. The second batch contained pTTv141. The parental strain M198 was used as a control strain. The cells were grown in TrMM medium supplemented with 40 g/l lactose, 20 g/l spent

grain extract, and 100 mM PIPPS, pH 5.5. Five transformants per construct were cultured. The pTTv140, pTTv142, and pTTv143 cultures were sampled on days 3, 5, 7, and 9. The pTTv141 cultures were sampled on days 3, 5, 7, and 10. The pH and cell dry weight of each sample were measured and culture supernatant samples were used for glycan structure analysis.

Glycan Analysis of T. reesei Strains Obtained by Targeting GnTII to alg3 Locus of T. reesei M198 Strain

[00341] Five different clones containing the pTTv140 vector (containing the native targeting peptide and the *cbhI* promoter), the pTTv142 vector (containing the MNT1 targeting peptide, the GNTII 74 aa N-terminal deletion, and the *cbhI* promoter), the pTTv143 vector (containing the MNT1 targeting peptide, the GNTII 110 aa N-terminal deletion, and the *cbhI* promoter), and the pTTv141 vector (containing the targeting peptide and the *gpdA* promoter) were analyzed.

[00342] N-glycan analyses were prepared in triplicate for day 5 samples, and in duplicate for day 3 and 7 samples on 96-well plates for 5 μg of supernatant protein. The protein concentrations of the supernatants were measured by Bradford-based assay (BioRad Quickstart Bradford Protein Assay) using BSA as a standard. PNGase F reactions were performed as described. The released N-glycans were first purified with Hypersep C-18 100 mg and then with Hypersep Hypercarb 10 mg (both from Thermo Scientific) where neutral and acidic glycans were separated. Both purifications were performed in 96-well format. Neutral N-glycans were analyzed by MALDI-TOF MS.

[00343] N-glycans of four different strains transformed with GnTII were analyzed. Clone
1-117A, which was transformed with the pTTv140 vector, and thus contained the native targeting peptide and the *cbhI* promoter, produced about 40% of G0 and about 13% of Hex6 (Figure 37A). Clones transformed with the pTTv143 vector, thus containing the MNT1 targeting peptide, the GnTII 110 aa N-terminal deletion, and the *cbhI* promoter, produced about 10% of G0 (Figure 37C). Clone 3B, which contained the *gbdA* promoter produced about 28% of G0 and about 19% of Hex6 (Figure 37D).

[00344] The glycosylation patterns of representative clones containing the pTTv140, pTTv141, and pTTv142 vectors were also shown to be stable as function of time (Figure 38).

Protein specific glycosylation

[00345] To analyze protein specific changes in glycosylation, samples from the pTTv142 vector-containing clone 3-17A and from the parental strain M198 were separated with SDS-PAGE and blotted to a PVDF membrane. The protein bands of interest (four bands of M198 and four of the 3-17A clone) were excised, and the N-glycans were liberated with on-membrane enzymatic release with PNGase F (Figure 39).

[00346] Detached and purified neutral N-glycans were analyzed using MALDI-TOF MS. The glycosylation pattern of total secreted proteins was similar to a separated 50 kDa protein of the M198 parental strain (Figure 40). The smallest size protein band was unglycosylated.

[00347] In the GnTII clone 3-17A, most of the untypical signals had disappeared, confirming their origin from the medium. Additionally, the glycosylation pattern of clone 3-17A differed from the glycan patterns of total secreted proteins (Figure 40B). The amount of G0 from clone 3-17A was about 35 to 36 % (Figure 40B).

Fermenter cultivation of GnTII strain

[00348] Fermenter cultivation of the GnTII strain 1-117A M329 (which contains the pTTv140vector) was fermented in TrMM pH 5.5 + 2% Spent grain extract + 6% lactose + 0.5% KH_2PO_4 + 0.5% $(NH_4)_2SO_4$ at +28°C (pH 5.5). N-glycan analysis was performed in triplicate to 5 µg of the secreted proteins described in the "Protein specific glycosylation" section above on samples taken on day 3. The amount of G0 was about 48% and the amount of Hex6 was about 19% on day 3 (Figure 41).

Example 8 –T. reesei ALG3 Homologs

Transformation of T. reesei M124 with GnTI Construct by Random Integration

[00349] *T. reesei* ALG3 homologs were identified from other organisms. These homologs can be used to design ALG3 deletion constructs for filamentous fungal cells

other than *T. reesei*. The ALG3 homologs are listed in Table 15. A multiple amino acid sequence alignment of *T. reesei* ALG3 and ALG3 homologs are shown in Figure 42.

Table 15: ALG3 Homologs.

Reference sequence	Organism	SEQ ID NO:
Trire2|104121|fgenesh5_pg.C_scaffold_3000076	*Trichoderma reesei*	126
Triat2|270085|fgenesh1_pg.contig_14_#_149	*Trichoderma atroviride*	127
TriviGv29_8_2|194462|fgenesh1_pm.87_#_115	*Trichoderma virens*	128
EGU81920.1	*Fusarium oxysporum Fo5176*	129
XP_389829.1	*Gibberella zeae PH-1*	130
AEO60805.1	*Myceliophthora thermophila*	131
XP_962259.1	*Neurospora crassa OR74A*	132
XP_001824044.1	*Aspergillus oryzae RIB40*	133
XP_001259497.1	*Neosartorya fischeri NRRL 181*	134
XP_001398696.2	*Aspergillus niger CBS 513.88*	135
XP_362427.2	*Magnaporthe oryzae 70-15*	136
NP_593853.1	*Schizosaccharomyces pombe 972h*	137

Example 9 –GnTII/GnTI Fusion Protein Variants

Generation of GnTII/GnTI Expression Construct

[00350] A recombinant GnTI/II fusion protein under the control of the inducible promoter *cbh1* and containing 1 of 4 spacer variants is constructed as described in Examples 4 and 5. The four spacer variants are the 2xG4S spacer, the 3xG4S spacer, the CBHI spacer, and the EGIV spacer.

[00351] Briefly, the fusion fragments are amplified from GnTII and GnTI templates separately with primers containing 50 bp in-frame overlaps at the fusion site. Fragments are purified from an agarose gel and used as PCR template for amplification of the fusion construct according to standard procedures. The fusion construct is cloned into a vector with ApaI/SpeI restriction sites, under the control of the inducible promoter *cbh1*. Additionally, the native mammalian Golgi targeting peptide in the GNTII domain was replaced by a *T. reesei* MNT1 (α-1,2-mannosyltransferase) targeting peptide.

[00352] To introduce the 2xG4S spacer variants into the fusion protein, T45 sequence is amplified in two parts by using PCR overlapping strategy. First, a fragment is amplified with AKT1-6-1 5' primer (GGTACCGGGCCCACTGCGCATCATGCGCTTCCGAATCTACAAGCG (SEQ ID

NO: 146)) and GP93 3' primer, and a second fragment is amplified with GP92 5' primer and AKT1-6-4 3' primer (GGCGCGCCACTAGTCTAATTCCAGCTGGGATCATAGCC (SEQ ID NO: 147)). Amplification is carried out with Phusion high-fidelity PCR polymerase (Finnzymes) under the standard conditions provided by the supplier. Cycling conditions are as described in Example 5. The resulting PCR product is purified from the agarose gel, and the fragments with overlapping, modified sequences are combined in the same reaction mixture with standard conditions without primers. Ten annealing/extension cycles are carried out as described in Example 5. Primers AKT1-6-1 (5') and AKT1-6-4 (3') are added, and cycling is continued as described in Example 5 for 20 amplification cycles. The amplified T45 fragment is then purified, digested with ApaI/SpeI (New England Biolabs) according to standard protocols, and cloned into the *Trichoderma reesei* expression vector. The cloned fragment is then verified by sequencing with appropriate set of primers and the generated sequence is used for construction of *T. reesei* expression vector with 2xG4S promoter and *alg3* targeting.

[00353] The resulting plasmid is used as a template for the 3xG4S spacer modification. Cloning of the T46 sequence is done as described above with T45. AKT1-6-1 5'-primer and GP95 3'-primer are used for first fragment synthesis, and GP94 5'-primer and AKT1-6-4 3'-primer are used for second fragment synthesis. Fragments are combined, and primers AKT1-6-1 (5') and AKT1-6-4 (3') are added for amplification. Amplified fragment T46 is then digested with ApaI/SpeI and cloned into the *Trichoderma reesei* expression vector. The cloned fragment is then verified by sequencing with an appropriate set of primers and the generated sequence is used for construction of *T. reesei* expression vector with 3xG4S promoter and *alg3* targeting.

[00354] The CBHI and EGIV spacers are constructed with a similar PCR overlap method. For the CBHI spacer, the first fragment is amplified with AKT1-6-1 5'-primer and GP107 3'-primer. The second fragment is amplified with GP108 5'-primer and AKT1-6-4 3'-primer (Table 11). For the EGIV spacer, the first fragment is amplified with AKT1-6-1 5'primer and GP109 3'primer. The second fragment is amplified with GP110 5'-primer and AKT1-6-4 3'-primer (Table 11). In both cases, the PCR products are purified from agarose gel, combined, and used as a template for the next PCR reaction to amplify the sequences T50 and T51. T50 and T51 PCR products are then digested

with ApaI/SpeI and cloned into the *Trichoderma reesei* expression vector. The cloned fragments are then verified by sequencing with appropriate sets of primers and the generated sequences are used for construction of *T. reesei* expression vectors with either CBHI or EGIV promoter and *alg3* targeting.

[00355] All PCR amplifications are made with high-fidelity Phusion polymerase (Finnzymes). Primers (Table 11) are ordered from MWG Operon. Sequencing is performed by the DNA Sequencing Laboratory of the Institute of Biotechnology, University of Helsinki, as a commercial service.

[00356] The *Trichoderma reesei* expression vectors with the described chimeric GnTII/GnTI sequences with spacer variations (2xG4S, 3xG4S, CBHI, and EGIV) are subcloned under the control of the *cbh1* promoter, with a *pyr4* gene loopout marker and *alg3* flanking region fragments for targeted integration in the backbone are then constructed. Expression cassettes are transformed into *T. reesei* strain M279 (*pyr4*⁻ strain of M202). After plate selection, the clones are PCR-screened and purified through single spores. To obtain material for glycan analyses, shake flask cultivations are performed as described.

Introduction of GnTII/I Fusion Protein Variants to T. reesei Strain Expressing Rituximab Antibody

[00357] The recombinant GnTII/I fusion protein variants are introduced into the rituximab-expressing *T. reesei* strain M279 described in Example 5.

[00358] Briefly, the vectors having the GnTII/GnTI fusion protein under the control of the *cbh1* promoter, the MNTI targeting peptide, the *pyr4* loop-out marker, and each of the 4 spacer variants are each subcloned into a backbone vector between *alg3* flanking region fragments for targeted integration, thus deleting the *alg3* gene. A PmeI-digested expression cassette is transformed into *T. reesei* strain M279 (a *pyr4*⁻ strain). After plate selection, the clones are PCR-screened and purified through single spores.

Glycan Analysis of Rituximab-Producing T. reesei GnTII/GnTI Variant Strains Obtained by Targeting to alg3 Locus

[00359] To obtain material for glycan analysis, shake flask cultivations are performed as described in Example 5 and, in addition, some culture media are supplemented with 0.3 mg/ml soybean trypsin inhibitor (SBTI) and 1% casamino acids. SBTI is added first at inoculation and then daily on days 3-6. PMSF and Pepstatin A is added to all samples before freezing.

[00360] Rituximab is purified with Protein G affinity chromatography from day 5 supernatant samples with SBTI and from day 5 and 7 samples without SBTI. PNGase F reactions are performed for ~10 µg of denatured protein. The released N-glycans are first purified with Hypersep C-18 and then with Hypersep Hypercarb (both from Thermo Scientific) where neutral and acidic glycans are separated. The purification steps are performed in 96-well format. Neutral and acidic N-glycans are analyzed by MALDI-TOF MS to test for the presence of the G0 glycoform on the rituximab antibody.

We claim:

1. A recombinant protein having N-acetylglucosaminyltransferase activity, wherein the recombinant protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and catalyzes the transfer of N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan, and wherein the recombinant protein comprises a catalytic domain from at least two different enzymes.

2. The recombinant protein of claim 1, wherein the recombinant protein is a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain.

3. The recombinant protein of claim 2, wherein the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain are from human enzymes.

4. The recombinant protein of claim 3, wherein the N-acetylglucosaminyltransferase I catalytic domain comprises a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to amino acid residues 105-445 of SEQ ID NO: 1.

5. The recombinant protein of claim 3 or claim 4, wherein the N-acetylglucosaminyltransferase II catalytic domain comprises a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical amino acid residues 30-447 of SEQ ID NO: 21.

6. The recombinant protein of any one of claims 2-5, wherein the N-acetylglucosaminyltransferase II catalytic domain is N-terminal to the N-acetylglucosaminyltransferase I catalytic domain.

7. The recombinant protein of any one of claims 2-6, further comprising a spacer in between the N-acetylglucosaminyltransferase I catalytic domain and the N-acetylglucosaminyltransferase II catalytic domain.

8. The recombinant protein of claim 7, wherein the spacer comprises sequence from a stem domain.

9. The recombinant protein of claim 7 or claim 8, wherein the spacer comprises a sequence selected from the group consisting of SEQ ID NO: 118, SEQ ID NO: 120, SEQ ID NO: 122, and SEQ ID NO: 124.

10. The recombinant protein of claim 9, wherein the spacer comprises a sequence selected from the group consisting of SEQ ID NO: 118, SEQ ID NO: 120, and SEQ ID NO: 124.

11. The recombinant protein of claim 9, wherein the spacer comprises the sequence of SEQ ID NO: 124 or SEQ ID NO: 120.

12. The recombinant protein of claim 9, wherein the spacer comprises the sequence of SEQ ID NO: 124.

13. The recombinant protein of any one of claims 2-12, further comprising a targeting peptide linked to the N-terminal end of the catalytic domains.

14. The recombinant protein of claim 13, wherein the targeting peptide comprises a stem domain.

15. The recombinant protein of claim 8 or claim 14, wherein the stem domain is from a protein selected from the group consisting of a mannosidase, a mannosyltransferase, a glycosyltransferase, a Type 2 Golgi protein, MNN2, MNN4, MNN6, MNN9, MNN10, MNS1, KRE2, VAN1, and OCH1.

16. The recombinant protein of claim 15, wherein the protein is from an organism selected from the group consisting of *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, and *Trichoderma*.

17. The recombinant protein of claim 13 or claim 14, wherein the targeting peptide is a Kre2 targeting peptide.

18. The recombinant protein of any one of claims 14-17, wherein the targeting peptide further comprises a transmembrane domain linked to the N-terminal end of the stem domain.

19. The recombinant protein of any one of claims 14-17, wherein the targeting peptide further comprises a cytoplasmic domain linked to the N-terminal end of the stem domain.

20. The recombinant protein of claim 18 wherein the targeting peptide further comprises a cytoplasmic domain linked to the N-terminal end of the transmembrane domain.

21. A recombinant protein comprising a human N-acetylglucosaminyltransferase II catalytic domain and a human N-acetylglucosaminyltransferase I catalytic domain wherein the N-acetylglucosaminyltransferase II catalytic domain is located N-terminal to the N-acetylglucosaminyltransferase I catalytic domain, a spacer sequence comprising sequence from a human N-acetylglucosaminyltransferase I stem domain located in between the catalytic domains, and a targeting peptide located N-terminal to the N-acetylglucosaminyltransferase II catalytic domain wherein the targeting peptide comprises a cytoplasmic domain, a transmembrane domain, and a stem domain from human N-acetylglucosaminyltransferase II.

22. A recombinant protein comprising a sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 95.

23. A recombinant protein comprising a human N-acetylglucosaminyltransferase II catalytic domain and a N-acetylglucosaminyltransferase I catalytic domain, wherein the N-acetylglucosaminyltransferase II catalytic domain is located N-terminal to the N-acetylglucosaminyltransferase I catalytic domain; a spacer located in between the catalytic domains, wherein the spacer comprises sequence selected from the group consisting of SEQ ID NO: 118, SEQ ID NO: 120, SEQ ID NO: 122, and SEQ ID NO: 124; and a targeting peptide located N-terminal to the N-acetylglucosaminyltransferase II catalytic domain wherein the targeting peptide comprises a cytoplasmic domain, a transmembrane domain, and a stem domain from human N-acetylglucosaminyltransferase II.

24. The recombinant protein of claim 23, wherein the spacer comprises a sequence selected from the group consisting of SEQ ID NO: 118, SEQ ID NO: 120, and SEQ ID NO: 124.

25. The recombinant protein of claim 23, wherein the spacer comprises the sequence of SEQ ID NO: 124 or SEQ ID NO: 120.

26. The recombinant protein of claim 23, wherein the spacer comprises the sequence of SEQ ID NO: 124.

27. An isolated polynucleotide encoding the recombinant protein of any of the above claims.

28. An expression vector comprising the isolated polynucleotide of claim 27 operably linked to a promoter.

29. A host cell comprising the expression vector of claim 28.

30. A method of producing a complex N-glycan, comprising :

(1) providing a host cell, wherein the host cell comprises a polynucleotide encoding a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain; and
(2) culturing the host cell such that the fusion protein is expressed, wherein the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan.

31. The method of claim 30, wherein the acceptor glycan is attached to a heterologous polypeptide.

32. The method of claim 30 or claim 31, wherein the complex N-glycan is GlcNAcβ2Manβ3(GlcNAcβ2Manβ6)Manβ4GlcNAcβ4GlcNAc.

33. The method of any one of claims 30-32, wherein the host cell is a filamentous fungal cell selected from the group consisting of *Trichoderma* sp., *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*,

Neurospora, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, and *Tolypocladium*.

34. The method of any one of claims 30-33, wherein the host cell further comprises a polynucleotide encoding a UDP-GlcNAc transporter.

35. The method of any one of claims 30-34, wherein the host cell has a reduced level of activity of a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase compared to the level of activity in a wild-type host cell.

36. The method of claim 35, wherein the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell.

37. The method of claim 36, wherein the *alg3* gene is deleted from the host cell.

38. The method of any one of claims 30-37, wherein the host cell has a reduced level of activity of an α-1,6-mannosyltransferase compared to the level of activity in a wild-type host cell.

39. The method of claim 38, wherein the host cell has a reduced level of expression of an och1 gene compared to the level of expression in a wild-type host cell.

40. The method of any one of claims 30-39, wherein the host cell further comprises a polynucleotide encoding an α-1,2-mannosidase.

41. The method of any one of claims 30-40, wherein the host cell further comprises a polynucleotide encoding a β-1,4-galactosyltransferase.

42. The method of any one of claims 30-41, wherein the host cell further comprises a polynucleotide encoding a sialyltransferase.

43. A method of producing a complex N-glycan, comprising :

(1) providing a *Trichoderma* host cell, wherein the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell and comprises a first polynucleotide encoding an N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding an N-acetylglucosaminyltransferase II catalytic domain; and
(2) culturing the host cell to produce a complex N-glycan.

44. A method of producing a complex N-glycan, comprising incubating a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain, an acceptor glycan, and an N-acetylglucosamine donor together in a buffer, wherein the fusion protein catalyzes the transfer of N-acetylglucosamine to a terminal Manα3 residue and N-acetylglucosamine to a terminal Manα6 residue of an acceptor glycan to produce a complex N-glycan.

45. A filamentous fungal cell comprising a mutation of *alg3* and Man3GlcNAc2, wherein the Man3GlcNAc2 comprises at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100% of neutral N-glycans in the cell.

46. The filamentous fungal cell of claim 45, wherein the mutation of *alg3* is a deletion of *alg3*.

47. The filamentous fungal cell of claim 45 or claim 46, wherein the cell is a Trichoderma reesei cell.

48. The filamentous fungal cell of any one of claims 45-47, further comprising a first polynucleotide encoding an N-acetylglucosaminyltransferase I catalytic domain and a second polynucleotide encoding an N-acetylglucosaminyltransferase II catalytic domain.

49. The filamentous fungal cell of any one of claims 45-47, further comprising a polynucleotide encoding a fusion protein comprising an N-acetylglucosaminyltransferase I catalytic domain and an N-acetylglucosaminyltransferase II catalytic domain.

50. A method of producing a Man3GlcNAc2 glycan in a host cell, comprising:

(1) providing a host cell with a reduced level of activity of a mannosyltransferase compared to the level of activity in a wild-type host cell; and
(2) culturing the host cell to produce a Man3GlcNAc2 glycan, wherein the Man3GlcNAc2 glycan comprises at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100% of the neutral N-glycans in the host cell.

51. The method of claim 50, wherein the Man3GlcNAc2 glycan is attached to a heterologous polypeptide

52. The method of claim 50 or claim 51, wherein the mannosyltransferase is a dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase.

53. The method of any one of claims 50-52, wherein the host cell has a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type host cell.

54. The method of any one of claims 50-53, wherein the host cell is a *Trichoderma* cell.

55. The method of any one of claims 50-54, wherein the level of activity of alpha-1,6-mannosyltransferase in the host cell is not reduced compared to the level of activity in a wild-type host cell.

56. The method of any one of claims 50-55, wherein the host cell comprises an endogenous polynucleotide encoding an α-1,2-mannosidase.

57. A filamentous fungal cell having a reduced level of expression of an *alg3* gene compared to the level of expression in a wild-type filamentous fungal cell, wherein the filamentous fungal cell comprises a recombinant protein according to any one of claims 2-26.

58. The filamentous fungal cell of claim 57, wherein the *alg3* gene comprises a mutation.

59. The filamentous fungal cell of claim 58, wherein the mutation of the *alg3* gene is a deletion of the *alg3* gene.

60. The filamentous fungal cell of any one of claims 57-59, wherein the fusion protein is encoded by a polynucleotide operably linked to a promoter.

61. The filamentous fungal cell of claim 60, wherein the promoter is an inducible promoter.

62. The filamentous fungal cell of claim 61, wherein the inducible promoter is the *cbh1* promoter.

63. The filamentous fungal cell of any one of claims 57-62, wherein the filamentous fungal cell further comprises a polynucleotide encoding a UDP-GlcNAc transporter.

64. The filamentous fungal cell of any one of claims 57-63, wherein the filamentous fungal cell has a reduced level of activity of an α-1,6-mannosyltransferase compared to the level of activity in a wild-type filamentous fungal cell.

65. The filamentous fungal cell of claim 64, wherein the filamentous fungal cell has a reduced level of expression of an och1 gene compared to the level of expression in a wild-type filamentous fungal cell.

66. The filamentous fungal cell of any one of claims 57-65, wherein the filamentous fungal cell further comprises a polynucleotide encoding an α-1,2-mannosidase.

67. The filamentous fungal cell of any one of claims 57-66, wherein the filamentous fungal cell further comprises a polynucleotide encoding a β-1,4-galactosyltransferase.

68. The filamentous fungal cell of any one of claims 57-67, wherein the filamentous fungal cell further comprises a polynucleotide encoding a sialyltransferase.

69. The filamentous fungal cell of any one of claims 57-68, wherein the filamentous fungal cell is selected from the group consisting of *Trichoderma* sp., *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Chrysosporium*, *Chrysosporium lucknowense*, *Filibasidium*, *Fusarium*, *Gibberella*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Myrothecium*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, and *Tolypocladium*.